ppe pesquisa e planejamento econômico

volume 49
número 1
abr. 2019

TESTANDO A ESTABILIDADE DO EFEITO DOS DETERMINANTES MICROECONÔMICOS DA TAXA DE POUPANÇA DAS FAMÍLIAS
Ajax Bello Moreira
Marcos Antonio Coutinho da Silveira

A EFICIÊNCIA TÉCNICA DAS DISTRIBUIDORAS DE ELETRICIDADE DO BRASIL DURANTE 2003-2012: UMA APLICAÇÃO DE FUNÇÕES DE DISTÂNCIA ESTOCÁSTICAS
Mariana De Santis
Fernando Damonte
Elías Curi

EFEITOS DO PROGRAMA DE QUALIFICAÇÃO PROFISSIONAL BOLSA FUTURO NA EMPREGABILIDADE E SALÁRIO DOS TRABALHADORES
Guilherme Resende Oliveira
Felipe Resende Oliveira
Rafael Terra
Ana Carolina Zoghbi

UMA ABORDAGEM DE DECOMPOSIÇÃO QUANTÍLICA PARA DIFERENCIAIS REGIONAIS DE SALÁRIOS NO BRASIL METROPOLITANO
Matheus Menezes dos Santos
Ana Maria Hermeto

TAL PAI, TAL FILHO? UMA ANÁLISE DOS EFEITOS FATORES DE CIRCUNSTÂNCIAS SOBRE O DESEMPENHO DOS ALUNOS NA AVALIAÇÃO DO PISA 2012
Lauro Nogueira
Erik Figueiredo

INSERÇÃO PRODUTIVA INTERNACIONAL E INTER-REGIONAL: UMA ANÁLISE A PARTIR DAS CADEIAS GLOBAIS DE VALOR
Fernando Salgueiro Perobelli
Inácio Fernandes de Araújo
Vinícius de Almeida Vale
Mônica de Moura Pires

ipea

volume 49 | número 1 | abr. 2019

pesquisa e planejamento econômico ▪ ppe

ipea

Governo Federal
Ministério da Economia
Ministro Paulo Guedes

ipea Instituto de Pesquisa Econômica Aplicada

Fundação pública vinculada ao Ministério da Economia, o Ipea fornece suporte técnico e institucional às ações governamentais – possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiros – e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

Presidente
Carlos von Doellinger

Diretor de Desenvolvimento Institucional, Substituto
Manoel Rodrigues dos Santos Junior

Diretor de Estudos e Políticas do Estado, das Instituições e da Democracia
Alexandre de Ávila Gomide

Diretor de Estudos e Políticas Macroeconômicas
José Ronaldo de Castro Souza Júnior

Diretor de Estudos e Políticas Regionais, Urbanas e Ambientais
Aristides Monteiro Neto

Diretor de Estudos e Políticas Setoriais de Inovação e Infraestrutura, Substituto
Bruno César Pino Oliveira de Araújo

Diretora de Estudos e Políticas Sociais
Lenita Maria Turchi

Diretor de Estudos e Relações Econômicas e Políticas Internacionais
Ivan Tiago Machado Oliveira

Assessora-chefe de Imprensa e Comunicação
Mylena Pinheiro Fiori

Ouvidoria: http://www.ipea.gov.br/ouvidoria
URL: http://www.ipea.gov.br

Pesquisa e Planejamento Econômico (PPE)
Publicação quadrimestral com análises teóricas e empíricas sobre uma ampla gama de temas relacionados à economia brasileira. Estabelecida em 1971 sob o título Pesquisa e Planejamento, PPE é publicada em abril, agosto e dezembro.

Corpo Editorial

Editor
Mauricio Cortez Reis

Coeditores
Marco A.F.H. Cavalcanti
José Gustavo Feres
Danilo Santa Cruz Coelho

Membros
Lauro Ramos (Ipea)
Eduardo Fiuza (Ipea)
Alexandre Xavier Ywata de Carvalho (Ipea)
Daniel da Mata (Ipea)
Carlos Viana de Carvalho (PUC-RJ)
Eduardo Rios Neto (CEDEPLAR-UFMG)
José Raimundo Carvalho (CAEN-UFC)
Marcelo Portugal (UFRGS)
Marco Bonomo (Insper)
Mônica Viegas de Andrade (CEDEPLAR-UFMG)
Rafael Coutinho Costa Lima (UFPE)
Renata Narita (FEA-USP)
Ricardo Paes de Barros (Insper)
Roberto G. Ellery (UnB)
Sergio Firpo (Insper)
Vladimir Ponczek (EESP-FGV/SP)

Secretária Executiva
Thamires Fernandes da Silva

© Instituto de Pesquisa Econômica Aplicada – **ipea** 2019

Pesquisa e Planejamento Econômico v. 1 – n.1 – jun. 1971.

Rio de Janeiro: Instituto de Pesquisa Econômica Aplicada, 1990 – v. – quadrimestral.

Título anterior: Pesquisa e Planejamento v. 1, n. 1 e 2, 1971

Periodicidade anterior: semestral de 1971–1975.

1. Economia – Pesquisa – Periódicos. 2. Planejamento Econômico – Brasil. I. Brasil. Instituto de Pesquisa Econômica Aplicada.

ISSN - 0 100-0551 CDD 330.05
33(81) (05)

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou o do Ministério da Economia.

É permitida a reprodução dos textos deste volume e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

SUMÁRIO

TESTANDO A ESTABILIDADE DO EFEITO DOS DETERMINANTES MICROECONÔMICOS DA TAXA DE POUPANÇA DAS FAMÍLIAS 9
Ajax Bello Moreira
Marcos Antonio Coutinho da Silveira

A EFICIÊNCIA TÉCNICA DAS DISTRIBUIDORAS DE ELETRICIDADE DO BRASIL DURANTE 2003-2012: UMA APLICAÇÃO DE FUNÇÕES DE DISTÂNCIA ESTOCÁSTICAS 35
Mariana De Santis
Fernando Damonte
Elías Curi

EFEITOS DO PROGRAMA DE QUALIFICAÇÃO PROFISSIONAL BOLSA FUTURO NA EMPREGABILIDADE E SALÁRIO DOS TRABALHADORES 69
Guilherme Resende Oliveira
Felipe Resende Oliveira
Rafael Terra
Ana Carolina Zoghbi

UMA ABORDAGEM DE DECOMPOSIÇÃO QUANTÍLICA PARA DIFERENCIAIS REGIONAIS DE SALÁRIOS NO BRASIL METROPOLITANO 99
Matheus Menezes dos Santos
Ana Maria Hermeto

TAL PAI, TAL FILHO? UMA ANÁLISE DOS EFEITOS FATORES DE CIRCUNSTÂNCIAS SOBRE O DESEMPENHO DOS ALUNOS NA AVALIAÇÃO DO PISA 2012 133
Lauro Nogueira
Erik Figueiredo

INSERÇÃO PRODUTIVA INTERNACIONAL E INTER-REGIONAL: UMA ANÁLISE A PARTIR DAS CADEIAS GLOBAIS DE VALOR 163
Fernando Salgueiro Perobelli
Inácio Fernandes de Araújo
Vinícius de Almeida Vale
Mônica de Moura Pires

CONTENTS

TESTING THE STABILITY OF THE EFFECT OF THE MICROECONOMIC DETERMINANTS OF THE FAMILIES' SAVING RATE 9
Ajax Bello Moreira
Marcos Antonio Coutinho da Silveira

TECHNICAL EFFICIENCY OF ELECTRICITY DISTRIBUIDORS IN BRAZIL DURING 2003-2012: AN APPLICATION OF STOCHASTIC DISTANCE FUNCTIONS 35
Mariana De Santis
Fernando Damonte
Elías Curi

THE EFFECTS OF THE JOB TRAINING PROGRAM BOLSA FUTURO IN EMPLOYMENT AND WAGE OF WORKERS 69
Guilherme Resende Oliveira
Felipe Resende Oliveira
Rafael Terra
Ana Carolina Zoghbi

A QUANTILE DECOMPOSITION APPROACH FOR REGIONAL WAGE DIFFERENTIALS IN METROPOLITAN BRAZIL 99
Matheus Menezes dos Santos
Ana Maria Hermeto

LIKE FATHER LIKE SON? AN ANALYSIS OF FACTORING EFFECTS OF CIRCUMSTANCES ON THE PERFORMANCE OF STUDENTS IN THE PISA EVALUATION 2012 133
Lauro Nogueira
Erik Figueiredo

INTERNATIONAL AND INTER-REGIONAL PRODUCTIVE INSERTION: AN ANALYSIS FROM THE GLOBAL VALUE CHAINS 163
Fernando Salgueiro Perobelli
Inácio Fernandes de Araújo
Vinícius de Almeida Vale
Mônica de Moura Pires

Prêmio Roberto Campos de Excelência em Pesquisa

O IPEA concedeu em 2018 o Prêmio Roberto Campos de Excelência em Pesquisa para os dois artigos que foram considerados os melhores publicados na Revista PPE em 2017 (Números 1, 2 e 3 do volume 47). Os artigos declarados como os vencedores pela comissão julgadora, formada pelos integrantes do Corpo editorial de PPE, foram:

1º Lugar: A relação entre o ensino superior público e privado e a renda e emprego nos municípios brasileiros (PPE, volume 47, número 3, p. 39-69).

Autores: *Roberto Hsu Rocha, Naercio Menezes Filho, Alison Pablo Oliveira, Bruno Kawaoka Komatsu.*

2º Lugar: Substituição de combustíveis fósseis por etanol e biodiesel no Brasil e seus impactos econômicos: uma avaliação do Plano Nacional de Energia 2030 (PPE, volume 47, número 3, p. 185-216).

Autores: *Jeronimo Alves dos Santos e Joaquim Bento de Souza Ferreira Filho.*

TESTANDO A ESTABILIDADE DO EFEITO DOS DETERMINANTES MICROECONÔMICOS DA TAXA DE POUPANÇA DAS FAMÍLIAS[1]

Ajax Bello Moreira[2]
Marcos Antonio Coutinho da Silveira[3]

O objetivo deste trabalho é testar a estabilidade dos parâmetros dos modelos de projeção da taxa de poupança agregada das famílias brasileiras com os dados da Pesquisa de Orçamentos Familiares (POF) do Instituto Brasileiro de Geografia e Estatística (IBGE). A diferença entre as taxas de poupança calculadas com as POFs de 2008 e 2002 foi decomposta segundo a metodologia proposta por Blinder-Oaxaca e por Neuman-Oaxaca na soma de componentes relacionados à variação das variáveis explicativas e à variação dos parâmetros dos modelos de projeção. Utilizando cinco definições da taxa de poupança e quatro modelos de projeção, verifica-se que a maior parte daquela diferença deve-se à variação dos parâmetros, restando um papel secundário para o efeito das mudanças nas médias amostrais das variáveis explicativas. Este resultado compromete o uso de modelos estimados com dados da POF para a projeção da taxa de poupança agregada das famílias brasileiras.

Palavras-chave: taxa de poupança agregada das famílias; decomposição de Blinder-Oaxaca; decomposição de Neuman-Oaxaca.

TESTING THE STABILITY OF THE EFFECT OF THE MICROECONOMIC DETERMINANTS OF THE FAMILIES' SAVING RATE

The purpose of this work is to test the stability of the parameters of the projection models of the aggregate saving rate of the Brazilian families with the data of the Survey of Household Budgets (POF) of IBGE. The difference between the savings rates calculated with the POFs of 2008 and 2002 was decomposed according to the methodology proposed by Blinder-Oaxaca and Neuman-Oaxaca in the sum of components related to the variation of the explanatory variables and the variation of the parameters of the projection models. Using five definitions of the savings rate and four projection models, most of this difference is due to the variation of the parameters, remaining a secondary role for the effect of the changes in the sample means of the explanatory variables. This result compromises the use of models estimated with POF data for the projection of the aggregate savings rate of Brazilian families.

Keywords: households' aggregate saving rate; Blinder-Oaxaca decomposition; Neuman-Oaxaca decomposition.

JEL: C01; E21.

1. Os autores agradecem as sugestões de um parecerista em uma versão anterior deste texto.

2. Coordenador de estudos de conjuntura e flutuações cíclicas na Diretoria de Estudos e Políticas Macroeconômicas (Dimac) do Ipea.

3. Técnico de planejamento e pesquisa na Dimac do Ipea.

1 INTRODUÇÃO

Entender e mensurar a formação da poupança doméstica é pesquisa de fundamental importância na atualidade econômica brasileira. Produz subsídio valioso para o desenho de políticas públicas orientadas para o aumento da taxa de investimento agregado com menor dependência de recursos externos. Entre todos os componentes da poupança doméstica, a poupança das famílias é a mais difícil de entender, estimar e prever, provavelmente devido à precariedade das bases de dados existentes. Embora dados agregados sejam disponíveis em alta frequência, estes são obtidos indiretamente como um resíduo de outros componentes das contas nacionais e, portanto, podem estar sujeitos a sérios erros de mensuração. Além disso, do ponto de vista analítico, a maior parte da literatura teórica busca explicar a poupança das famílias na esfera das decisões individuais. Consequentemente, a disponibilidade de dados microeconômicos possibilita estudar e calcular a poupança das famílias a partir de informação diretamente reportada pelos agentes tomadores de decisão.

A Pesquisa de Orçamentos Familiares (POF) do Instituto Brasileiro de Geografia e Estatística (IBGE) é a única fonte de dados microeconômicos com base na qual é possível explicar e estimar a taxa de poupança agregada das famílias brasileiras. Esta pesquisa coleta informação não apenas sobre a renda e o consumo das famílias, como também sobre suas principais características demográficas, de forma que pode ser usada não apenas para calcular diretamente a taxa de poupança de diferentes grupos socioeconômicos, como também para estimar o efeito sobre a poupança de um conjunto de potenciais variáveis explicativas sugeridas pela literatura empírica e teórica.

No entanto, como a POF é realizada com um intervalo aproximado de cinco anos, a projeção da taxa de poupança das famílias nos anos intermediários precisa ser feita a partir de modelos estimados com os dados de uma POF realizada em ano anterior ao da projeção. Se os parâmetros do modelo são suficientemente estáveis no tempo, o próximo passo é utilizar outras fontes de dados para projetar as variáveis explicativas incluídas no modelo. Caso contrário, as projeções não são confiáveis, uma vez que os parâmetros do modelo de projeção podem ter sido substancialmente alterados. É importante ressaltar que, mesmo se a POF fosse realizada todo ano, a avaliação da estabilidade dos modelos de projeção ainda é útil para a realização de exercícios de previsão e simulação da taxa de poupança.

Antes de tudo, cabe analisar se existem argumentos que justificariam uma eventual instabilidade dos parâmetros dos modelos de projeção. A princípio, isto poderia ser o resultado de mudanças no contexto cultural e institucional que balizam as decisões financeiras das famílias, incluindo também o *status* da política econômica vigente. Por exemplo, o efeito do ciclo da vida sobre a poupança vai depender da predisposição das famílias em deixar alguma herança para seus descendentes, e sabe-se

que isto é uma característica cultural que tem evoluído bastante no passado recente. Da mesma forma, o efeito do regime de previdência pública sobre a taxa de poupança privada não pode ser dissociado da expectativa dos agentes econômicos em relação à sustentabilidade fiscal do governo. Não menos importante, a falta de credibilidade dos agentes econômicos na condução da política econômica pode levá-los a antecipar ou postergar seus gastos com as diversas categorias de consumo.

Dados os argumentos teóricos colocados anteriormente, o objetivo deste trabalho é avaliar empiricamente a estabilidade temporal de modelos de projeção da taxa de poupança agregada das famílias brasileiras. Para tanto, o estudo explora a oportunidade oferecida pela existência de duas pesquisas conceitualmente similares e distantes no tempo – a POF 2002 e a POF 2008 – para decompor, por meio dos métodos de Blinder-Oaxaca e de Neuman-Oaxaca, a variação da taxa de poupança agregada das famílias entre 2002 e 2008 em uma componente relacionada à variação das variáveis explicativas e outra componente relacionada à variação dos parâmetros do modelo. A significância e o tamanho relativo entre estes dois componentes fornecem uma medida para avaliar se a estabilidade dos modelos é suficientemente elevada para gerar projeções confiáveis. Se a parcela atribuída à variação dos parâmetros não é significativa, ou explica uma parcela relativamente muito pequena da variação da taxa de poupança, infere-se que o modelo produza projeções confiáveis desta variável.

As decomposições foram aplicadas utilizando quatro modelos de projeção alternativos, que fazem hipóteses diferentes sobre o comportamento das famílias. O primeiro modelo baseia-se no procedimento usado pelo IBGE para estimar a taxa de poupança agregada das famílias em anos sem disponibilidade de POF. Primeiro, divide-se a amostra de famílias da POF em segmentos caracterizados por renda e região e, em seguida, calcula-se a taxa de poupança agregada como uma média ponderada das taxas dos segmentos. O segundo é uma extensão do anterior, no sentido em que os segmentos são caracterizados por um maior número de variáveis preditivas. Já o terceiro modelo incorpora um leque maior de variáveis explicativas sugeridas pela literatura sobre poupança. Uma importante característica dos dados é a quantidade proporcionalmente grande de famílias com poupança nula. Para lidar com este fato, o modelo de Heckman é o quarto modelo alternativo para projetar a taxa de poupança. Enquanto a estrutura linear dos três primeiros modelos permite a aplicação direta da decomposição de Blinder-Oaxaca na sua avaliação, o quarto modelo possui uma estrutura não linear, que torna necessária uma adaptação conhecida como decomposição de Neuman-Oaxaca.

A taxa de poupança agregada é definida como a razão entre a poupança total das famílias e a renda total, o que é conceitualmente diferente da média das taxas de poupança das famílias. A fim de calcular a taxa de poupança agregada, é necessário

ponderar a taxa de poupança de cada família com a sua renda. Esta ponderação permite calcular com dados microeconômicos uma variável macroeconômica. Como o trabalho está focado na projeção da taxa de poupança agregada, todos os estimadores foram definidos ponderando-se cada observação com a renda da família associada.

A taxa de poupança pode ser definida segundo diferentes horizontes de postergação do consumo. Pode englobar apenas a poupança financeira ou pode ainda incluir a compra de imóveis, bens duráveis e formação de capital humano (educação e saúde). Logo, o trabalho analisa cinco medidas alternativas de poupança, as quais correspondem a diferentes graus de abrangência do conceito. Para cada uma delas, a variação da taxa de poupança foi estimada e decomposta segundo os quatro modelos anteriormente definidos, em um total de vinte decomposições.

Os resultados encontrados não sustentam a hipótese de estabilidade da relação entre a taxa de poupança agregada das famílias e as variáveis explicativas dos modelos de projeção. Para os quatro modelos propostos, a parcela da variação da taxa de poupança agregada explicada pela mudança dos parâmetros do modelo é significativamente maior que a parcela explicada pela variação das características observáveis das famílias. Este resultado coloca em evidência as limitações dos procedimentos de interpolação/extrapolação usados pelo IBGE para estimar a taxa de poupança agregada das famílias em anos sem disponibilidade de POF.

Este trabalho possui mais seis seções, além desta introdução. A segunda e a terceira discutem a metodologia e a base de dados, nesta ordem. A quarta e a quinta descrevem, respectivamente, as cinco medidas de poupanças e as variáveis explicativas dos modelos de projeção. A sexta apresenta os resultados. Por fim, a sétima conclui.

2 METODOLOGIA

Esta seção começa com uma descrição sumária do procedimento utilizado pelo IBGE para estimar a taxa de poupança agregada das famílias em anos sem disponibilidade de POF. Em seguida, descreve os principais passos da metodologia empregada no trabalho para avaliar a viabilidade deste procedimento.

Seja um modelo $S^t = F(X^t;\beta^t)$ que explica a taxa de poupança agregada das famílias no ano t em função de um conjunto de variáveis explicativas X^t. Consequentemente, a variação $S^{t+1}-S^t$ desta taxa entre t+1-t pode ser causada tanto pela variação $X^{t+1}-X^t$ das variáveis explicativas quanto pela variação $\beta^{t+1}-\beta^t$ dos parâmetros do modelo. Suponha que a POF foi realizada no ano t, de forma que β^t pode ser estimado com os dados para S^t e X^t. Caso a POF não seja realizada no ano t+1, a estimativa de β^t é usada pelo IBGE para projetar a taxa de poupança

deste ano como $F(X^{t+1};\beta^t)$, onde observações para X^{t+1} são provenientes de fontes de dado apuradas anualmente. Consequentemente, a confiabilidade desta projeção está condicionada à estabilidade dos parâmetros β entre t e t+1.

O trabalho tem o objetivo limitado de avaliar esta estabilidade, uma vez que este é o passo inicial para validar o uso do procedimento descrito anteriormente. A metodologia empregada para tanto pode ser divida em três partes, as quais são detalhadas nas subseções a seguir. A primeira descreve a especificação dos modelos de projeção $S^t = F(X^t;\beta^t)$, com especial atenção para a escolha das explicativas X^t. A segunda consiste na estimação dos modelos ponderando cada família pela sua participação na renda agregada, de forma que a poupança agregada possa ser calculada a partir de dados microeconômicos. A terceira consiste no uso dos procedimentos de Blinder-Oaxaca e Neuman-Oaxaca para decompor a diferença entre as taxas de poupanças agregadas, calculadas com as POFs de 2002 e 2008, em uma componente explicada pela variação das características observáveis X, e outra explicada pela variação dos parâmetros β. A estabilidade dos parâmetros β é avaliada com base na significância e no tamanho relativo entre estes dois componentes.

2.1 Modelos de projeção

Os modelos de projeção devem apresentar capacidade explicativa e descrever padrões de comportamento estáveis temporalmente. Quatro modelos foram especificados com base em duas vertentes: *i)* a metodologia utilizada pelo IBGE para interpolação/extrapolação da taxa de poupança das famílias nas Contas Nacionais; e *ii)* a literatura teórica sobre os fatores determinantes da poupança das famílias.

O primeiro modelo, denominado renda-estado, é inspirado no procedimento adotado pelo Sistema de Contas Nacionais do IBGE (2008) para projetar o perfil de consumo das famílias nos anos em que a POF não é realizada. Este procedimento pressupõe a estabilidade dos padrões de poupança segundo o nível de renda relativo e a localização geográfica. A amostra de famílias da POF realizada no ano j é repartida por 27 regiões, cada qual correspondendo a um estado da Federação, e por dez faixas de renda delimitadas pelos decis da renda nacional, totalizando 270 segmentos amostrais. A taxa de poupança S_{ij} da família i é então explicada pela equação de regressão:

$$S_{ij} = \alpha_j + \sum_{rx} \theta_{rx}^{j} D_{ij}^{rx} + \varepsilon_{ij} \,, \tag{1}$$

onde ε_{ij} é um termo aleatório, enquanto (D_{ij}^{rx}) são variáveis indicadoras do pertencimento da família i ao segmento (r,x), com x=2, ..., 27 e r=2, ..., 10, tal que os segmentos (r,x) são definidos em subconjunto de pares que garanta a ausência de colinearidade.

O segundo modelo, denominado renda-perfil, é uma modificação do anterior, que reparte a amostra de famílias da POF em 320 segmentos, caracterizados pela combinação de 64 perfis socioeconômicos e cinco faixas de renda delimitadas pelo vintis da renda nacional. Cada perfil é, por sua vez, uma combinação de cinco variáveis: gênero, idade e escolaridade do chefe da família e a presença de um membro aposentado ou funcionário público na família.[4] Comparado ao primeiro, este modelo permite capturar o efeito de um leque maior de variáveis preditivas.

Nos modelos citados anteriormente, as classes de renda foram definidas com base nos decis ou vintis da distribuição da renda observada em 2002, os quais foram atualizados monetariamente para definir as correspondentes classes em 2008. Desta forma, o valor médio do indicador em 2002, ponderado com a renda, é a proporção da renda recebida por cada decil ou vintil em 2002, e o valor recebido por famílias equivalentes em 2008. A variação entre anos destas médias ponderadas é uma medida da alteração na distribuição da renda entre os anos.

O terceiro modelo segue a literatura teórica para selecionar as variáveis (X) que explicam a taxa de poupança das famílias por meio da regressão linear:

$$S_{ij} = \alpha_j + \sum_k \beta_k^j x_{ij}^k + \varepsilon_{ij} \,. \tag{2}$$

Um aspecto importante na especificação deste modelo, e também do modelo seguinte, é a seleção das variáveis explicativas, o que é discutido na seção 5, à luz da literatura teórica e empírica sobre os determinantes da taxa de poupança.

O quarto modelo, denominado teórico estendido, é uma adaptação do anterior, para lidar com o fato de que uma porção não desprezível das famílias na amostra tem poupança nula. Admitindo que esta condição reflete uma característica não observada das famílias, busca-se diferenciar as famílias com e sem poupança nula por meio de um modelo não linear, originalmente proposto por Heckman (1979), para lidar com o problema de viés de seleção, o qual é descrito pelas equações (3) e (4):

$$\begin{aligned} &Pr\left[S_{ij} \neq 0 | Z_{ij}\right] = \Phi\left(\theta_j Z_{ij}\right), \\ &E\left[S_{ij} | S_{ij} \neq 0, X_{ij}, \lambda_{ij}\right] = \beta_j X_{ij} + \gamma_j \lambda_{ij}, \end{aligned} \tag{3}$$

onde:

$$\lambda_{ij} = \phi\left(\theta_j Z_{ij}\right) / \Phi\left(\theta_j Z_{ij}\right) \tag{4}$$

é a inversa de Mills, na qual $\phi(\cdot)$ e $\Phi(\cdot)$ são, respectivamente, a função densidade e a função de distribuição acumulada da normal padrão. A primeira equação é um modelo Probit que explica a probabilidade de a família i na POF do ano j ter

4. Neste modelo, cada *dummy* indica uma combinação destas seis características, e a idade e a escolaridade do chefe da família são dicotomizadas em dois grupos separados pelas medianas das distribuições destas variáveis.

poupança não nula, onde o vetor Z_{ij} inclui, além das variáveis do terceiro modelo, uma variável *dummy* adicional para indicar se a família reportou a realização de pelo menos uma operação financeira.[5] A segunda equação é um modelo de regressão linear que explica a taxa de poupança condicionada ao fato de não ser nula, onde o vetor X_{ij} inclui apenas as variáveis incluídas no terceiro modelo, enquanto λ_{ij} é a inversa de Mills para lidar com o viés de seleção. O modelo pode ser estimado em dois estágios: inicialmente, o modelo Probit é estimado com a amostra completa, rendendo o estimador $\hat{\theta}_j$; em seguida, o modelo linear é estimado com a subamostra das famílias com poupança não nula, usando a estimativa da inversa de Mills $\hat{\lambda}_{ij} = \phi(\hat{\theta}_j Z_{ij})/\Phi(\hat{\theta}_j Z_{ij})$ como sua observação, rendendo os estimadores de mínimos quadrados ordinários $\hat{\beta}_j$ e $\hat{\gamma}_j$.

2.2 Estimador ponderado

Os três primeiros modelos podem ser escritos na forma mais geral de um modelo de regressão linear $E[S_{ij}|X_{ij}] = \beta_j X_{ij}$, onde S_{ij} e X_{ij} são, respectivamente, a taxa de poupança e o vetor de características observáveis da família i na POF do ano j. Cabe salientar que o vetor de coeficientes β_j pode variar entre as populações de pesquisas distintas. Como o objetivo do trabalho é projetar a taxa de poupança agregada das famílias, e não a taxa de poupança média delas, as estimativas $\hat{\beta}_j$ de mínimos quadrados dos coeficientes são obtidas ponderando cada família por sua participação na renda agregada da amostra. Para melhor esclarecer este argumento, é preciso observar que, embora ambas a taxa de poupança agregada e a taxa de poupança média de uma população sejam calculadas como uma média ponderada das taxas de poupança das unidades familiares, na primeira o fator de ponderação é igual a 1 para cada família (média aritmética), enquanto na segunda o fator de ponderação de cada família é a sua participação na renda agregada. Logo, a taxa de poupança agregada depende não apenas das taxas de poupança das unidades familiares, mas também da distribuição de renda na população. Para provar este argumento, sejam Y_i e R_i os valores da poupança e da renda da família i em uma população de tamanho N. A taxa de poupança média $\tilde{S}$ é simplesmente a média aritmética das taxas de poupança individuais, ou seja, $\tilde{S} = \widetilde{\left(\frac{Y}{R}\right)} = \frac{1}{N}\sum_i \frac{Y_i}{R_i}$. Contudo, denotando a participação da família na renda agregada como $w_i = \frac{R_i}{\sum_i R_i}$, segue que a taxa de poupança agregada $\bar{S}$ é dada pela expressão:

$$\bar{S} = \frac{\tilde{Y}}{\tilde{R}} = \frac{\sum_i Y_i}{\sum_i R_i} = \sum_i \left(\frac{R_i}{\sum_i R_i}\right)\frac{Y_i}{R_i} = \sum_i w_i \frac{Y_i}{R_i},$$

onde $\tilde{Y} = \frac{1}{N}\sum_i Y_i$ e $\tilde{R} = \frac{1}{N}\sum_i R_i$ são, respectivamente, as médias aritméticas da poupança e da renda.

5. Esta variável, supostamente exógena, é incluída para lidar com o problema de identificação do modelo.

A estimação por mínimos quadrados *ordinários* do modelo de regressão linear implica que a estimativa $\tilde{S}_j$ da taxa de poupança *média* do ano *j* provida pela POF_j satisfaz a equação $\tilde{S}_j = \tilde{\beta}_j \tilde{X}_j$ onde $\tilde{\beta}_j$ é o estimador de mínimos quadrados ordinários, enquanto $\tilde{S}_j = \frac{1}{N}\sum_{i \in POF_j} S_{ij}$ e $\tilde{X}_j = \frac{1}{N}\sum_{i \in POF_j} X_{ij}$. Entretanto, a estimação por mínimos quadrados *ponderados pela renda* do modelo de regressão linear implica que a estimativa $\bar{S}_j$ da taxa de poupança *agregada* do ano *j* provida pela POF_j satisfaz a equação $\bar{S}_j = \hat{\beta}_j \bar{X}_j$ tal que $\hat{\beta}_j$ é o estimador de mínimos quadrados ponderado pela renda, enquanto $\bar{S}_j = \sum_{i \in POF_j} w_{ij} S_{ij}$ e $\bar{X}_j = \sum_{i \in POF_j} w_{ij} X_{ij}$ onde w_{ij} é a estimativa da participação da família *i* na renda agregada no ano da POF_j.

O quarto modelo apresenta uma estrutura linear somente quando restrito à subamostra de famílias com poupança não nula. Logo, para este subconjunto, a estimativa da taxa de poupança agregada $\bar{S}_j$ satisfaz a equação $\bar{S}_j = \hat{\beta}_j \bar{X}_j + \hat{\gamma}_j \bar{\lambda}_j$, tal que $\bar{\lambda}_j = \sum_{i \in POF_j} w_{ij} \hat{\lambda}_{ij}$ é a média ponderada da inversa de Mills no ano *j*. Para obter a taxa de poupança agregada irrestrita, basta multiplicar $\bar{S}_j$ pela participação total das famílias com renda não nula na renda agregada.

2.3 Decomposição de Blinder-Oaxaca

O grau de confiabilidade do modelo de projeção depende da estabilidade de seus parâmetros no intervalo de tempo entre o ano da coleta de dados e o ano da projeção. Eventualmente, mudanças comportamentais, rupturas institucionais ou reviravoltas na política econômica poderiam impactar as decisões financeiras das famílias e, portanto, alterar os parâmetros dos modelos de projeção. Uma forma de avaliar a estabilidade dos parâmetros é utilizar a decomposição de Blinder-Oaxaca,[6] segundo a qual a diferença entre as taxas de poupança agregada estimadas com as versões da POF de 2002 e 2008, dada por $R = \bar{S}_8 - \bar{S}_2$ onde $\bar{S}_8 = \hat{\beta}_8 \bar{X}_8$ e $\bar{S}_2 = \hat{\beta}_2 \bar{X}_2$, é fatorada como:

$$R = \bar{S}_8 - \bar{S}_2 = \hat{\beta}_8 \bar{X}_8 - \hat{\beta}_2 \bar{X}_2 \; = Q + P + I, \tag{5}$$

onde:

$$Q = \hat{\beta}_2 (\bar{X}_8 - \bar{X}_2),$$

$$P = \left(\hat{\beta}_8 - \hat{\beta}_2\right)\bar{X}_2,$$

$$I = \left(\hat{\beta}_8 - \hat{\beta}_2\right)(\bar{X}_8 - \bar{X}_2).$$

A parcela *Q*, conhecida como efeito-explicativa, mede a contribuição da variação das médias das características observáveis das famílias (*X*) sobre a variação da taxa

6. Esta decomposição foi proposta em uma série de artigos com aplicações em economia do trabalho, entre os quais se destacam Blinder (1973), Oaxaca (1973) e Oaxaca e Ransom (1998). Um resumo é encontrado em Jann (2005; 2008).

de poupança, enquanto a segunda parcela *P*, conhecida como efeito-coeficiente, mede a contribuição da variação dos coeficientes estimados sobre a variação da taxa de poupança. A terceira parcela *I*, conhecida como efeito-interação, decorre da variação conjunta dos coeficientes e das explicativas. O erro-padrão de cada termo da decomposição pode ser calculado para testar sua significância. As duas últimas parcelas podem ser consideradas como o efeito de componentes não observáveis, ou seja, a soma destas parcela é o erro de projeção $\epsilon = P+I$. A rejeição da hipótese nula de que o efeito-coeficiente *P* (ou o efeito-interação *I*) é igual a zero depõe contra a estabilidade dos coeficientes. Entretanto, mesmo neste caso, agora é possível comparar as magnitudes do efeito-explicativa *Q* e do efeito-coeficiente *P* para verificar se o primeiro é suficientemente maior que o segundo para prover toda a informação relevante sobre a direção e a magnitude da variação da taxa de poupança agregada das famílias.

Em relação ao quarto modelo (teórico estendido), o método de Oaxaca-Blinder não decompõe corretamente a variação da taxa de poupança agregada em um efeito-observável e um erro de projeção não observável. No procedimento de Heckit, a correção do viés de seleção dá-se por meio da inclusão das inversas de Mills como regressores na equação linear estimada com a amostra das famílias de poupança não nula. Logo, a variação média das inversas de Mills entre as POFs, denotada por $\bar{\lambda}_8 - \bar{\lambda}_2$, faz parte do efeito-explicativa *Q* na decomposição de Blinder-Oaxaca. No entanto, isto é conceitualmente errado, uma vez que, como fica evidente na equação (4), esta variação depende não apenas da variação $\bar{Z}_8 - \bar{Z}_2$ das características observáveis *Z*, mas também da variação $(\hat{\theta}_8 - \hat{\theta}_2)$ dos coeficientes estimados θ. Para contornar este problema, Neuman e Oaxaca (2004) propõem uma decomposição alternativa para a diferença entre as taxas de poupança agregadas entre as POFs de 2002 e 2008, denominada decomposição de Neuman-Oaxaca, a qual é dada por:

$$R = \bar{S}_8 - \bar{S}_2 = Q + \epsilon,$$

onde

$$\begin{aligned} Q &= \hat{\beta}_2(\bar{X}_8 - \bar{X}_2) + \hat{\gamma}_2(\bar{\bar{\lambda}}_8 - \bar{\lambda}_2), \\ \epsilon &= (\hat{\beta}_8 - \hat{\beta}_2)\bar{X}_8 + (\hat{\gamma}_8 - \hat{\gamma}_2)\bar{\lambda}_8 + \hat{\gamma}_2(\bar{\bar{\lambda}}_8 - \bar{\lambda}_8). \end{aligned} \tag{6}$$

O ponto mais importante desta decomposição é que $\bar{\bar{\lambda}}_8$ é a média das inversas de Mills de 2008 calculadas por meio da expressão (4), com coeficientes θ_j estimados com a POF de 2002, enquanto $\bar{\lambda}_2$ *e* $\bar{\lambda}_8$ são as médias calculadas com coeficientes θ_j estimados das POFs de 2002 e 2008, respectivamente. Logo, *Q* é o efeito-explicativa da decomposição, uma vez que $\bar{\bar{\lambda}}_8$ não aparece neste termo e, portanto, não depende da variação dos coeficientes $(\hat{\theta}_8 - \hat{\theta}_2)$. O erro de projeção ϵ

corresponde ao efeito não observável, cujas duas primeiras parcelas equivale à soma do efeito-coeficiente com o efeito-interação na decomposição de Blinder-Oaxaca. Já o último termo do erro de projeção é exclusivo do quarto modelo e decorre da estimação das inversas de Mills de 2008 a partir dos coeficientes estimados com a POF de 2002. Finalmente, como a decomposição anteriormente citada está circunscrita à subamostra de famílias com poupança não nula, a decomposição análoga referente à amostra total requer que seus termos sejam ajustados pela participação na renda agregada das famílias com poupança não nula.

3 BASE DE DADOS

A fonte de dados microeconômicos usada neste trabalho foi produzida pelas POFs de 2002-2003 e de 2008-2009 do IBGE.[7] Cada pesquisa é composta por questionários com perguntas bastante detalhadas sobre a estrutura de consumo, gastos e rendimentos de uma amostra de famílias brasileiras. A partir desta informação, é possível estimar a taxa de poupança das famílias. As pesquisas têm sido realizadas com uma defasagem de pelo menos cinco anos, o que possibilita a divergência do conteúdo de variáveis que tenham o mesmo nome. Esta é uma questão relevante que não foi considerada no trabalho, o qual parte da hipótese de que as pesquisas são conceitualmente compatíveis. A única questão tratada explicitamente refere-se à desvalorização do valor da moeda, e, para isto, utilizam-se os corretores propostos pelo IBGE e disponíveis nas bases de dados das pesquisas. O tempo da pesquisa da POF é de doze meses, de forma a capturar o efeito de flutuações sazonais nos orçamentos familiares. Todos os valores reportados são anualizados e corrigidos pelo efeito da inflação.

A unidade de observação usada na parte empírica do trabalho corresponde ao conceito de unidade de consumo da POF, a qual é definida como um morador ou grupo de moradores de um domicílio particular permanente que compartilham a mesma fonte de alimentação. O tamanho amostral da POF 2002-2003 e da POF 2008-2009 consiste em 48.000 e 56.091 unidades de consumo, respectivamente, envolvendo todo o território nacional, inclusive áreas rurais. Cabe comentar que a POF faz uma distinção entre família e unidade de consumo, de forma que esta pode conter várias famílias.[8] No entanto, este caso engloba menos que 1% da amostra e, além disso, a definição de unidade de consumo é mais próxima do conceito de

7. Duas outras pesquisas foram realizadas anteriormente: 1987-1988 e 1995-1996.

8. A POF define família como um conjunto de pessoas ligadas por laços de parentesco, dependência doméstica ou normas de convivência, que morem na mesma unidade domiciliar; ou como uma pessoa que more só em uma unidade domiciliar.

família relevante para o objetivo do trabalho. A partir deste ponto do estudo, as duas versões da POF serão referidas como POF 2002 e POF 2008.[9]

4 TAXA DE POUPANÇA AGREGADA

O foco do trabalho é a projeção da taxa de poupança agregada das famílias brasileiras, definida como a razão entre a poupança agregada e a renda agregada. É preciso, então, que a definição destas variáveis seja coerente com os objetivos do trabalho. Na prática, o conceito de poupança é ambíguo, uma vez que depende dos ativos reais e financeiros, que são usualmente comprados com a finalidade de reserva de valor e acumulação de capital. Consequentemente, o estudo considera cinco medidas alternativas de poupança: S1, a medida mais restrita, definida como o valor da compra líquida de ativos estritamente financeiros; S2, uma medida usual de poupança, definida como a soma da poupança S1 com a compra líquida de bens imóveis; S3, definida como a soma da poupança S2 com a compra líquida de veículos; S4, definida como a soma da poupança S3 com a compra líquida dos demais bens duráveis, tais como eletrodomésticos; e S5, definida como a poupança S4 acrescida dos gastos com capital humano (educação e saúde).[10]

As duas primeiras medidas de poupança estão relacionadas com a taxa de poupança das Contas Nacionais e dispensam comentário quanto à sua relevância. Contudo, do ponto de vista da economia familiar, outros tipos de gastos também podem ser entendidos como uma forma de poupança. Neste sentido, as medidas S3 e S4 trazem implícita a ideia de que os bens duráveis não apenas proporcionam um fluxo de serviços de consumo durante sua vida útil, mas também podem ser usados como reserva de valor, descontada sua depreciação, o que é particularmente importante no caso das famílias mais pobres, as quais têm acesso restrito aos instrumentos convencionais de poupança disponibilizados pelos mercados financeiros. O tratamento em separado dos imóveis e veículos, em relação aos demais bens duráveis, deve-se aos seus valores unitários relativamente mais elevados, o que implica um distinto grau de restrição de crédito para suas aquisições. Além disso, a taxa de depreciação temporal dos bens imóveis é bem menor que a dos outros bens duráveis, como veículos e eletrodomésticos, de forma que aqueles funcionam melhor como reserva de valor.[11] A medida S5 embute a ideia de que a acumulação de capital humano é uma importante forma de

9. A POF 2002 foi realizada entre julho de 2002 e julho de 2003, enquanto a POF 2008 foi realizada entre julho de 2008 e julho de 2009.

10. A medida S1 inclui caderneta de poupança, compra de títulos de renda fixa ou variável, fundos de investimento, previdência privada.

11. Outra característica distinta dos bens imóveis é que podem ser usados como instrumento de garantia para a obtenção de crédito.

investimento, que proporciona maior fluxo de renda futuro. A medida de renda usada no cálculo da taxa de poupança de uma família, segundo qualquer critério de medida, é a soma dos rendimentos correntes, monetários e não monetários, de todos os seus membros, livres de impostos e contribuições previdenciárias públicas compulsórias.[12]

A especificação do modelo de projeção da taxa de poupança das famílias precisa refletir as principais características dos dados. Uma importante característica, ilustrada pela tabela 1, é a relevância empírica das famílias com poupança nula e negativa. As primeiras colunas da tabela apresentam as médias amostrais da renda e da poupança dos segmentos de famílias com poupança positiva, nula e negativa, enquanto as últimas colunas da tabela apresentam as distribuições amostrais da renda e da população entre estes três segmentos. É notória a relevância do segmento das famílias com poupança nula, tanto em relação à sua participação numérica na população quanto em relação à sua participação na renda agregada. Considerando a amostra conjunta das duas pesquisas POF, a participação numérica deste segmento é muito elevada no caso das três medidas de poupança menos abrangentes (S1, S2 e S3), alcançando 74%, 59% e 46%, respectivamente. Esta participação cai acentuadamente para 11% e 3% no caso das medidas S4 e S5, nesta ordem. Analogamente, a fatia da renda agregada abocanhada pelas famílias com poupança nula, embora menos proeminente que a participação numérica, também é bastante elevada, alcançando 58%, 42% e 30%, no caso das medidas S1, S2 e S3, respectivamente, e novamente sofrendo uma queda significativa para 5% e 1%, no caso das medidas S1 e S2. No que tange ao segmento de famílias com poupança negativa, sua participação numérica na população e na renda agregada é, em geral, relativamente pequena, embora ainda relevante, especialmente no caso das medidas de poupança menos abrangentes. A tabela 1 também revela que, no caso de todas as medidas de poupança, a renda média é relativamente maior no segmento de famílias com poupança positiva e menor no segmento com poupança nula.[13]

12. Duas questões são relevantes neste cálculo. Primeiro, conforme procedimento usual, algumas despesas compulsórias, como pagamento de fiança e multas, são tratadas como choques negativos na renda, e, portanto, deduzidas da renda total. Segundo, as receitas monetárias englobam rendimentos do trabalho e do capital (juros, dividendos, aluguéis e lucros), bem como aposentadorias privadas, pensões, transferências governamentais (renda mínima, bolsa-escola etc.) e receitas esporádicas. A razão pela qual as receitas esporádicas foram incluídas no cálculo da renda corrente é que um importante argumento teórico para a determinação da taxa de poupança é a hipótese de suavização do consumo no tempo, a qual implica que choques transitórios na renda têm efeito desprezível sobre o consumo corrente.

13. Todos os resultados foram calculados descartando as famílias com taxa de poupança fora do intervalo [-1,1] e com valores espúrios das variáveis explicativas. Estes critérios descartam menos do que 1% das famílias.

TABELA 1
Média e proporção das famílias na população e na renda agregada por sinal da poupança e POF

	Média (R$)					Proporção (%) na renda agregada e na população					
	Poupança negativa		Poupança positiva		Poupança nula	Poupança negativa		Poupança positiva		Poupança nula	
	Renda	Poupança	Renda	Poupança	Renda	Renda	População	Renda	População	Renda	População
POF			S1					S1			
2002	1.756	-286	2.345	391	1.089	12	9	19	10	69	81
2008	1.898	-410	2.954	438	1.341	10	10	39	23	51	67
Total	**1.833**	**-356**	**2.784**	**425**	**1.212**	**11**	**9**	**31**	**17**	**58**	**74**
POF			S2					S2			
2002	1.440	-442	1.804	307	997	9	8	43	30	48	61
2008	1.573	-511	2.649	557	1.225	8	9	54	36	38	55
Total	**1.509**	**-480**	**2.293**	**451**	**1.113**	**8**	**8**	**50**	**33**	**42**	**59**
POF			S3					S3			
2002	1.204	-427	1.758	362	842	8	8	61	44	31	47
2008	1.438	-522	2.418	689	1.138	6	8	65	48	29	45
Total	**1.322**	**-475**	**2.113**	**543**	**997**	**7**	**8**	**63**	**46**	**30**	**46**
POF			S4					S4			
2002	1.108	-534	1.375	260	583	5	6	90	84	5	10
2008	1.223	-670	1.924	492	801	3	5	92	84	5	11
Total	**1.165**	**-602**	**1.667**	**385**	**702**	**4**	**5**	**91**	**84**	**5**	**11**
POF			S5					S5			
2002	599	-671	1.333	357	375	2	4	97	93	1	3
2008	680	-846	1.838	612	552	1	3	98	94	1	3
Total	**639**	**-759**	**1.607**	**494**	**463**	**1**	**3**	**98**	**94**	**1**	**3**

Fonte: POF 2002-2003/POF 2008-2009.

A tabela 2 apresenta a distribuição por faixa de renda dos segmentos de famílias com poupança financeira S1 positiva, nula e negativa. São definidas dez faixas limitadas pelos decis da distribuição da renda. Em linha com a evidência empírica para os Estados Unidos apresentada em Dynan, Skyner e Zeldes (2004), segundo a qual as famílias mais ricas têm maior propensão a poupar, a tabela revela que a proporção de famílias com poupança financeira positiva aumenta com a faixa de renda, o contrário se verificando com a proporção de famílias com poupança financeira nula. No entanto, cabe observar que, entre a faixa mais rica e mais pobre das famílias, a queda na proporção de famílias com poupança nula ultrapassa 45 pontos percentuais (p.p.), enquanto o aumento na proporção das famílias com poupança positiva mal alcança 10 p.p. Esta diferença de magnitude

é explicada pelo avanço da participação do segmento de famílias com poupança negativa nas faixas de renda mais elevadas, a qual aumenta 37 p.p. entre as faixas extremas, sugerindo que a disponibilidade de renda não é o elemento crítico na decisão entre poupar ou "despoupar". Este resultado é coerente com a hipótese de que as famílias mais ricas estão menos sujeitas à restrição de crédito e, consequentemente, são mais prováveis de contratar empréstimos financeiros para a compra de imóveis, veículos e outros bens duráveis.

Em seu conjunto, as estatísticas apresentadas nas tabelas 1 e 2 deixam evidente que a presença de famílias com poupança nula está longe de ser um resultado residual, e também não é um fato restrito às faixas mais pobres da população. Assim, a metodologia usada na projeção da taxa de poupança agregada das famílias precisa levar em conta a existência de uma possível heterogeneidade no efeito das variáveis preditivas entre segmentos de famílias com poupança nula e não nula. Isto é feito neste trabalho por meio do quarto modelo – teórico estendido –, analisado na subseção 2.1, o qual se baseia no procedimento de Heckman para tratamento do problema do viés de seleção amostral.

TABELA 2
Distribuição por faixa de renda dos segmentos de famílias com poupança financeira (S1) positiva, nula e negativa

Número do decil	1	2	3	4	5	6	7	8	9	10	**Total**
Decil da distribuição da renda familiar	565	813	1.056	1.329	1.659	2.079	2.661	3.635	5.716	Máximo	-
Famílias com poupança positiva (%)	4,0	5,2	5,7	6,9	7,8	8,8	10,6	11,4	12,3	13,8	**8,6**
Famílias com poupança nula (%)	92,9	88,9	86,1	82,3	78,4	75,4	68,4	63,7	57,5	45,9	**74,4**
Famílias com poupança negativa (%)	3,1	5,9	8,2	10,8	13,8	15,8	21,0	24,9	30,2	40,3	**17,1**
Total	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**

Fonte: POF 2002-2003/POF 2008-2009.

5 PREDITORES DO MODELO TEÓRICO DA TAXA DE POUPANÇA

O objetivo do trabalho é avaliar a estabilidade de modelos de projeção usados na extrapolação da taxa de poupança agregada das famílias. Como esta estabilidade depende da especificação do modelo, esta seção discute as variáveis explicativas – demográficas e socioeconômicas – incluídas nos modelos de projeção teórico e teórico estendido apresentados na subseção 2.1. Browning e Lusardi (1996) e Deaton (2005) oferecem um resumo abrangente da literatura que fundamenta a seleção destas variáveis. A tabela 3 apresenta a média amostral delas para diferentes segmentos da população.

Uma vez que a taxa de poupança é a razão entre a poupança e a renda, o ponto de partida teórico de qualquer discussão em torno de seus determinantes remonta ao modelo do ciclo da vida desenvolvido por Friedman (1953) e Brumberg e Modigliani (1954), segundo o qual as famílias suavizam consumo no tempo como resultado de uma alocação intertemporal ótima de seus recursos. Obviamente que a validade deste resultado vai depender do grau de acesso das famílias ao mercado de crédito. Uma importante implicação desta teoria é que o consumo das famílias não responde a choques transitórios na sua renda, mas se move em linha com choques permanentes. Estes resultados centrais da teoria foram logo contestados pela evidência empírica, a qual revela a existência de um razoável paralelismo entre a renda corrente e o consumo corrente no ciclo da vida. Em um importante trabalho, Gourinchas e Parker (2000) observam que o paralelismo tende a desaparecer apenas na fase final do ciclo da vida. Por conta desta discussão, a idade do chefe da família e a renda corrente familiar são fortes candidatas a variáveis explicativas. A inclusão de um termo quadrático para a idade torna o modelo suficientemente flexível para capturar a evolução empírica da taxa de poupança no ciclo da vida.

Uma vez que a existência de restrições institucionais para o acesso ao crédito altera drasticamente as predições teóricas em relação ao comportamento da taxa de poupança ao longo do ciclo da vida, o modelo inclui uma variável *dummy* para indicar a realização de operações financeiras pelos membros da unidade familiar, tais como aplicações e resgates em cadernetas de poupança, fundos mútuos, ações e títulos de renda fixa. A ideia subjacente é que a extensão do acesso aos mercados financeiros reflete a força da restrição de crédito enfrentada pela unidade familiar.

Uma importante variável explicativa para a taxa de poupança é o nível de educação da família. Esta variável traz informação sobre a renda permanente da família. Conforme a teoria do ciclo da vida, mudanças na renda permanente têm efeito duradouro sobre o consumo e a poupança. Além disso, esta variável pode ter efeitos adicionais sobre a poupança, na medida em que famílias mais escolarizadas são mais eficientes na administração de seu patrimônio. Neste sentido, alguns estudos, tais como Bernheim e Scholz (1993), Browning e Lusardi (1996) e Attanasio e Székely (1998), encontraram evidência empírica de um efeito positivo da educação sobre a poupança. Contudo, usando dados da POF, estudo do Banco Central do Brasil (2013) não encontra correlação entre a escolaridade e a taxa de poupança das famílias. Neste trabalho, o nível de educação é medido como o número de anos de estudo do chefe da família. Quatro classes de escolaridade são definidas com base no número de anos de estudo do chefe da família, denotado por "e": primário incompleto (e<4), secundário incompleto (4<=e<8), colegial incompleto (8<=e<11) e universitário (11=<e). Também é incluída uma variável que reflete o capital humano agregado da família, dado pela média do número de anos de estudo dos membros da família que trabalham além do chefe.

Como a base de dados foi produzida por duas POFs realizadas em períodos distintos, é possível introduzir uma variável de coorte a partir do ano de nascimento do chefe da família. A ideia é que pessoas nascidas em diferentes períodos podem ter hábitos de consumo e poupança distintos. Para tanto, denotando por "a" o ano de nascimento do chefe a família, são definidas cinco categorias para esta variável: a<1950, 1950<=a<1960, 1960<=a<1970, 1970<=a<1980 e a>=1980, que estão representadas por variáveis indicadoras de pertinência a cada um destes grupos. Os estudos que analisam se gerações distintas têm diferentes propensões a poupar chegam a resultados mistos. Em alguns casos, o efeito coorte mostra-se significativo: Paxson (1996) para os Estados Unidos; Jappelli e Modigliani (1998) para a Itália; e Moreira e Silveira (2015) para o Brasil; enquanto para outros o efeito é inexistente: Paxson (1996) para o Reino Unido.

Algumas extensões da teoria foram desenvolvidas para conciliá-la com a evidência empírica. Uma primeira linha de estudo argumenta que, se o chefe da família planeja suavizar o consumo dos membros individuais, e não o consumo agregado da família, suas decisões de consumo e poupança precisam levar em conta a evolução do tamanho e da composição da família no ciclo da vida. Consistentes com esta argumentação, alguns estudos empíricos revelam que o paralelismo desaparece após a correção da renda e do consumo por estas variáveis. Contudo, outros estudos, como Fernandez-Villaverde e Krueger (2007), observam que o paralelismo, embora reduzido, permanece mesmo controlando por demográficas. Com base nesta discussão, as seguintes variáveis demográficas foram incluídas no modelo para capturar o efeito de mudanças no tamanho e na composição da família: número de crianças (menos de 15 anos), número de jovens (entre 15 e 22 anos), número de adultos (entre 22 e 65 anos) e número de idosos (acima de 65 anos).

Uma segunda linha de estudo argumenta que uma importante razão para as famílias pouparem é o motivo precaução. É razoável esperar que o tamanho desta modalidade de poupança varie diretamente com a incerteza em relação à renda do trabalho (redução de salário ou perda do emprego) e com a probabilidade de ocorrência inesperada de gastos improrrogáveis, principalmente despesas com saúde, os quais podem ser interpretados como choques negativos na renda corrente. Browning e Lusardi (1996) argumentam que esta incerteza tem um impacto negativo sobre o consumo, principalmente quando interage com a existência de restrições de liquidez. Eles também encontram evidência de que a poupança precaucionária é mais importante para jovens e pobres. Carrol (1992) propõe que a resposta ótima das famílias à incerteza da renda futura é a rápida formação de um estoque de ativos líquidos (*buffer stock*) na fase inicial do ciclo da vida, após o qual o consumo tende a mover-se em paralelo com a renda.

Gourinchas e Parker (2000) encontram evidência de que o comportamento das famílias é compatível com este argumento até a meia idade. Já na fase final do ciclo da vida, as famílias começam a poupar rapidamente para aposentadoria, comportando-se mais em linha com o modelo-padrão. Freitas (2010) conclui que o efeito da poupança precaucionária no Brasil é estatisticamente significante. Na mesma linha, Ribeiro (2015) encontra evidência de que o motivo precaução é a principal razão para a poupança das famílias, inclusive para os mais velhos. Moreira e Silveira (2015) concluem que a maior estabilidade da renda tem efeito negativo sobre a taxa de poupança.

No que diz respeito ao efeito do motivo precaução sobre a taxa de poupança, foram incluídas variáveis *dummies* para indicar a existência ou não de pelo menos um membro da família que receba renda como aposentado, como empregado público, como empregado no setor formal da economia e como beneficiário de programas de transferências governamentais do tipo Bolsa Família. Existe uma *dummy* diferente para cada uma destas fontes de renda. A ideia é que a maior estabilidade destas fontes de renda torna a poupança precaucionária menos relevante, pois é possível contar com uma renda mínima em qualquer circunstância. Além disso, em um contexto institucional caracterizado pela existência de um regime de previdência pública englobando parcela significativa da população, não é possível ignorar o fato de que o incentivo a poupar das famílias contribuintes pode ser fortemente reduzido devido à perspectiva de uma renda vitalícia futura, especialmente se existem programas de renda mínima para a velhice, como é o caso do Brasil (Lei Orgânica da Assistência Social – Loas). Também como indicadoras da estabilidade da renda familiar, foram introduzidas variáveis *dummies* para informar se o chefe da família e seu cônjuge recebem renda. Quanto maior o número de membros trabalhando, menor a probabilidade de uma família ficar sem qualquer fonte de renda. Com o mesmo objetivo, outra variável *dummy* é introduzida para informar se o número de recebimentos mensais da família ao longo do ano é maior que doze. Esta variável pretende avaliar a estabilidade financeira da unidade familiar, uma vez que indica se a família é ou não desprovida de qualquer fonte de renda em algum mês do ano.

Finalmente, o modelo inclui variáveis *dummies* indicando a propriedade de imóvel próprio e de veículo. Estes bens duráveis merecem tratamento diferenciado devido ao seu alto valor unitário e menor taxa de depreciação temporal, de forma que sua propriedade pode alterar as decisões de poupança das famílias. Eles podem ser usados como garantia para empréstimos pessoais, suavizando a restrição de crédito, ou como reserva de valor para fins de poupança precaucionária, reduzindo a demanda por ativos financeiros com este objetivo.

TABELA 3
Taxas de poupança agregada e variáveis explicativas

POF	2002	2008	**Total**	POF	2002	2008	**Total**
Perfil econômico e financeiro				*Variáveis demográficas*			
Aposentadoria (%)	29,7	34,1	**32**	Idade do chefe	46,3	48,2	**47,3**
Emprego público (%)	29,2	27,3	**28,3**	Idade do chefe 2	23,4	25,3	**24,4**
Transferência (%)	6,5	11,7	**9,2**	Sexo do chefe (%)	24	27,7	**25,9**
Emprego formal (%)	54,8	54,4	**54,6**	Nº crianças (idade<=15)	1,1	0,9	**1,0**
Nº rendimentos (%)	74,9	70,6	**72,7**	Nº jovens (15<idade<=22)	1,1	0,9	**1,0**
Cônjuge com renda (%)	68,8	69,8	**69,3**	Nº adultos 22<idade<=65)	1,5	1,5	**1,5**
Imóvel próprio (%)	77,6	77,4	**77,5**	Nº idosos (idade>65)	0,2	0,2	**0,2**
Veículo próprio (%)	48,5	50,1	**49,3**	Capital humano	2,3	2,7	**2,5**
Log renda	0,45	0,45	**0,46**	*Poupança (%)*			
Operação financeira(%)	19,8	24,7	**22,5**	Financeira: S1	0,7	4,1	**2,4**
Classes de escolaridade (anos de estudo do chefe = e)				S1 e imóveis: S2	4,9	9,7	**7,3**
4<=e<8 (%)	12,5	11,8	**12,2**	S2 e veículos: S3	10,1	14,7	**12,5**
8<=e<11 (%)	22,9	26,7	**24,8**	S3 e bens duráveis: S4	14,9	19,9	**17,4**
11<=e (%)	19,9	20,5	**20,2**	S4 e saúde e educação: S5	24,3	29,2	**26,8**

Fonte: POF 2002-2003/POF 2008-2009.
Elaboração dos autores.
Obs.: 1. O valor em percentual de uma variável explicativa indica a proporção das famílias da amostra pertencentes ao segmento da população caracterizado pela variável.
2. Idade do chefe 2 é o quadrado da idade do chefe, dividido por 100.

6 RESULTADOS

Esta seção discute os resultados da aplicação das decomposições de Blinder-Oaxaca e de Neuman-Oaxaca sobre a diferença entre as taxas de poupança agregada das famílias brasileiras estimadas com as versões das POFs de 2002 e de 2008. O objetivo é verificar a confiabilidade das projeções da taxa de poupança agregada em 2008 a partir de modelos econométricos estimados com os dados da POF de 2002. Para testar a robustez dos resultados, a decomposição é aplicada sobre quatro modelos alternativos, os quais explicam a taxa de poupança familiar a partir de diferentes abordagens, conforme explicado na subseção 2.1. Inicialmente, a tabela 4 apresenta as estatísticas R^2 dos quatro modelos para cada POF. Os resultados mostram que o modelo teórico, o qual inclui apenas 26 variáveis explicativas, tem um ajuste superior aos modelos de renda-estado e de renda-perfil. Em geral, o modelo teórico estendido melhora marginalmente a capacidade explicativa do modelo teórico.

TABELA 4
Ajustamento (*R2*) dos modelos de poupança

Medida de poupança	S1	S2	S3	S4	S5
POF	Modelo renda-estado: 270 variáveis				
2002	0,022	0,055	0,139	0,110	0,187
2008	0,098	0,202	0,205	0,177	0,210
POF	Modelo renda-perfil: 320 variáveis				
2002	0,018	0,048	0,120	0,102	0,155
2008	0,088	0,135	0,152	0,133	0,160
POF	Modelo teórico: 26 variáveis				
2002	0,024	0,099	0,195	0,170	0,222
2008	0,150	0,326	0,291	0,261	0,267
POF	Modelo teórico estendido				
2002	0,036	0,124	0,187	0,162	0,214
2008	0,226	0,417	0,294	0,267	0,271

Fonte: POF 2002-2003/POF 2008-2009.

A tabela 5 apresenta os resultados da decomposição de Blinder-Oaxaca para os modelos renda-estado, renda-perfil e teórico. São mostrados, para cada modelo, o efeito-explicativa *Q* e o erro de projeção, o qual é a soma do efeito-coeficiente *P* e do efeito-interação *I*, tal como descritos na subseção 2.3. Para os dois primeiros modelos também é apresentado o efeito-constante, o qual é a parte do efeito-coeficiente relativo à diferença entre as constantes estimadas com as POFs de 2008 e 2002. Adicionalmente, a primeira parte da tabela apresenta as taxas de poupança agregada calculadas com as duas POFs e ainda a respectiva diferença entre elas.

TABELA 5
Decomposição de Blinder-Oaxaca da diferença entre as taxa de poupança de 2008 e 2002

Taxa de poupança agregada					
Medida de poupança	S1	S2	S3	S4	S5
POF 2008 (A)	4,14***	9,73***	14,69***	19,91***	29,20***
	(0,20)	(0,25)	(0,26)	(0,25)	(0,23)
POF 2002 (B)	0,68***	4,85***	10,15***	14,88***	24,25***
	(0,01)	(0,16)	(0,18)	(0,18)	(0,17)
Diferença: (A) - (B)	3,46***	4,88***	4,54***	5,02***	4,95***
	(0,05)	(0,30)	(0,31)	(0,30)	(0,29)

(Continua)

Decomposição de Blinder-Oaxaca: modelo renda-estado					
Medida de poupança	S1	S2	S3	S4	S5
Efeito-explicativa	-0,12***	-0,36***	-0,96***	-0,85***	-1,13***
	(0,02)	(0,07)	(0,10)	(0,10)	(0,12)
Erro de projeção:					
Efeito-coeficiente	3,79***	5,68***	5,68***	6,10***	6,22***
	(0,04)	(0,32)	(0,32)	(0,32)	(0,29)
Efeito-interação	-0,21***	-0,44***	-0,19	-0,23	-0,14
	(0,06)	(0,14)	(0,14)	(0,14)	(0,12)
Decomposição de Blinder-Oaxaca: modelo renda-perfil					
Medida de poupança	S1	S2	S3	S4	S5
Efeito-explicativa	0,02	-0,27***	-0,77***	-0,75***	-0,76***
	(0,03)	(0,06)	(0,09)	(0,08)	(0,10)
Erro de projeção:					
Efeito-coeficiente	3,47***	5,44***	5,47***	5,95***	5,79***
	(0,20)	(0,30)	(0,31)	(0,31)	(0,29)
Efeito-interação	-0,03	-0,28**	-0,16	-0,18	-0,09
	(0,09)	(0,11)	(0,11)	(0,11)	(0,10)
Decomposição de Blinder-Oaxaca: modelo teórico					
Medida de poupança	S1	S2	S3	S4	S5
Efeito-explicativas	-0,07	-0,48***	-0,74***	-0,82***	-0,59***
	(0,07)	(0,16)	(0,20)	(0,19)	(0,21)
Erro de projeção:					
Efeito-coeficiente	4,00***	5,85***	5,64***	6,15***	5,82***
	(0,27)	(0,30)	(0,31)	(0,30)	(0,28)
Efeito-interação	-0,47**	-0,49**	-0,36	-0,31	-0,28
	(0,19)	(0,24)	(0,23)	(0,22)	(0,21)

Fonte: POF 2002-2003/POF 2008-2009.
Obs.: 1. Desvio-padrão entre parênteses.
2. ***, ** e * = significativo a 1%, a 5% e a 10%, respectivamente.

Os principais resultados da tabela 5 são praticamente robustos em relação às medidas de poupança adotadas. Inicialmente, cabe destacar que a taxa de poupança agregada teve um aumento significativo no intervalo entre as duas POFs, o qual oscila entre 3,46 e 5,02 p.p., dependendo da medida adotada. O método de Oaxaca-Blinder decompõe este aumento no efeito-observável e no erro de projeção

não observável. Neste sentido, o primeiro fato importante é que, com exceção do efeito-observável relativo à medida de poupança S1, os dois principais termos da decomposição – o efeito-explicativa e o efeito-coeficiente – são significativos, com o primeiro sempre positivo e o segundo sempre negativo. O efeito-interação também é sempre negativo, embora tenha se revelado significativo apenas para as medidas S1 e S2 de poupança.

Embora significativo para quase todas as medidas de poupança, o efeito-explicativa tem magnitude pequena e sinal oposto em relação à variação da taxa de poupança agregada. Já o efeito-coeficiente é significativo para todas as medidas de poupança. Além disso, este termo é da mesma ordem de grandeza e sinal que a variação total da taxa de poupança agregada. O efeito-interação também apresenta pequena magnitude em relação à variação da taxa de poupança agregada.

Em linha com os resultados anteriores, a magnitude do efeito-coeficiente positivo, em valor absoluto, é de ordem de grandeza superior à magnitude do efeito-observável negativo, o que explica o aumento da taxa de poupança agregada. Em geral, a magnitude do efeito-observável fica abaixo de 1 p.p., enquanto a magnitude do efeito-coeficiente supera 4 p.p., chegando até 6 p.p. nas medidas mais abrangentes de poupança. Em suma, não apenas o efeito-coeficiente é estatisticamente significativo, como também sua magnitude supera de longe a magnitude do efeito-explicativa, respondendo quase que inteiramente por toda a diferença entre as taxas de poupança estimadas pelas versões da POF de 2002 e 2008. Além disso, o efeito-explicativas tem direção oposta ao aumento da taxa de poupança agregada, enquanto o efeito-coeficiente tem a mesma direção. No seu conjunto, estes resultados constituem evidência empírica contrária à estabilidade dos parâmetros dos modelos de projeção.

Como mostra a tabela 1, uma proporção considerável das famílias apresenta taxa de poupança nula, especialmente no caso da poupança financeira (S1), fato que poderia explicar parte dos resultados encontrados anteriormente. Para lidar com esta possibilidade, o trabalho introduz o quarto modelo, teórico estendido, descrito na subseção 2.1. A estrutura não linear deste modelo requer uma extensão da decomposição de Oaxaca-Blinder conhecida como decomposição de Neuman-Oaxaca, a qual foi descrita na subseção 2.4. Como esta decomposição está circunscrita à subamostra de famílias com poupança não nula, é necessário que seus termos sejam ajustados pela participação na renda agregada do conjunto das famílias com poupança não nula, a fim de se obter a decomposição relativa à amostra total. A primeira parte da tabela 6 apresenta a decomposição de Neuman-Oaxaca para as subamostras de famílias com poupança não nula, enquanto a segunda parte apresenta o ajuste desta decomposição para a amostra total.

TABELA 6
Decomposição de Neuman-Oaxaca da diferença entre as taxa de poupança de 2008 e 2002

Taxa de poupança agregada: subamostra de famílias com poupança ≠ 0					
Medida de poupança	S1	S2	S3	S4	S5
POF 2008 (A)	8,45***	15,84***	20,84***	20,93***	29,44***
	(0,42)	(0,40)	(0,35)	(0,27)	(0,24)
POF 2002 (B)	2,24***	9,33***	14,76***	15,61***	24,47***
	(0,24)	(0,29)	(0,25)	(0,19)	(0,18)
Diferença: (A) - (B)	6,21***	6,51***	6,08***	5,32***	4,97***
	(0,48)	(0,50)	(0,43)	(0,33)	(0,30)
Decomposição de Neuman-Oaxaca: subamostra de famílias com poupança ≠ 0					
Medida de poupança	S1	S2	S3	S4	S5
Efeito- explicativa	-1,59***	-0,83	-1,23***	-0,46	-0,54*
	(0,55)	(0,53)	(0,45)	(0,34)	(0,32)
Erro de projeção	7,80**	7,34**	7,30**	5,79**	5,52**
	(1,03)	(1,02)	(0,88)	(0,67)	(0,62)
Taxa de poupança agregada: amostra completa					
Medida de poupança	S1	S2	S3	S4	S5
POF 2008 (A)	4,14***	9,73***	14,69***	19,91***	29,20***
	(0,21)	(0,27)	(0,27)	(0,26)	(0,24)
POF 2002 (B)	0,68***	4,85***	10,15***	14,88***	24,25***
	(0,08)	(0,16)	(0,18)	(0,18)	(0,18)
Diferença: (A) - (B)	3,46***	4,88***	4,54***	5,02***	4,95***
	(0,22)	(0,31)	(0,32)	(0,31)	(0,30)
Decomposição de Neuman-Oaxaca: amostra completa					
Medida de poupança	S1	S2	S3	S4	S5
Efeito-explicativa	-1,33***	0,2	-0,54*	0,26	-0,06
	(0,29)	(0,31)	(0,32)	(0,33)	(0,32)
Erro de projeção	4,78**	4,68**	5,08**	4,76**	5,00**
	(0,52)	(0,61)	(0,64)	(0,65)	(0,62)

Fonte: POF 2002-2003/POF 2008-2009.
Obs.: 1. Desvio-padrão entre parênteses.
2. ***, ** e * = significativo a 1%, a 5% e a 10%, respectivamente.

O primeiro fato relevante na subamostra de famílias com poupança não nula é que a taxa de poupança agregada, para todas as medidas de poupança, continua aumentando significativamente entre 2002 e 2008. No entanto, ao contrário do

que ocorre com a amostra completa, agora este aumento diminui monotonicamente com a abrangência da medida de poupança, caindo de 6,21 p.p., na medida S1, para 4,97 p.p., na medida S5.

No que tange à decomposição de Neuman-Oaxaca, os resultados levam a conclusões qualitativamente semelhantes aos obtidos com os três modelos anteriores. Mesmo considerando apenas o subconjunto das famílias que têm poupança não nula nos dois anos, o efeito-explicativa é significativo, mas tem o sinal contrário ao esperado. A taxa de poupança observada, em todos os casos, aumenta, e a taxa projetada diminui. Resultado mostra que, mesmo corrigindo para um possível erro de especificação, verifica-se que o padrão de comportamento das famílias mudou entre os anos 2002/2008. Os resultados extrapolados para o conjunto das famílias elimina a relevância do efeito-explicativa e confirma a inadequação da especificação.

É importante salientar a robustez dos resultados obtidos em relação à especificação dos modelos de projeção. Além do modelo renda-estado, mais próximo ao procedimento adotado pelo IBGE para projetar a taxa de poupança agregada, três modelos alternativos foram testados com este objetivo, os quais incluíram na sua especificação um número bem mais extenso de variáveis explicativas sugeridas pela literatura. Além disso, o modelo teórico estendido foi especificado, de forma a levar em conta o fato de que uma proporção significativa das famílias tem poupança nula. Apesar deste esforço, verifica-se que, historicamente, a diferença entre as taxas de poupança agregada estimadas com duas POFs consecutivas é explicada em maior extensão por mudanças nos coeficientes estimados, restando um papel secundário para as mudanças nas médias amostrais das características preditivas. Mesmo quando estatisticamente significativo, o tamanho do efeito-explicativa observado é economicamente irrelevante e tem o sinal errado, face ao tamanho dos efeitos não observados em todos os quatro modelos de projeção.

Intuitivamente, este resultado sugere que aquela diferença deve ser atribuída a mudanças no padrão de comportamento das famílias no intervalo de realização da POF, o que está refletido nas diferentes estimativas dos coeficientes produzidas pelas pesquisas. Este resultado depõe desfavoravelmente contra o emprego de modelos econométricos estimados com dados defasados da POF para projetar a taxa de poupança corrente e futura.

Por fim, cabe ponderar que este resultado é muito forte e também poderia ser explicado por erros de mensuração, uma vez que as versões da POF de 2002 e 2008, embora muito próximos conceitualmente, não são exatamente iguais. Assim, embora as variáveis tenham sido construídas com o cuidado de se homogeneizar os conceitos, não há como garantir uma exata equivalência de medida, mas vale mencionar que o objetivo das POFs é a medida de despesa das famílias, o que sugere que o próprio IBGE teve interesse em manter a consistências destas medidas

centrais da pesquisa. Além disso, o primeiro modelo explica a taxa de poupança apenas com o estado e o decil de renda do domicílio, quantidades que certamente não estão sujeitas ao problema de erro de medida. O segundo, o terceiro e o quarto modelos, em ordem crescente, estão vulneráveis a questões de consistência. No entanto, a instabilidade dos resultados foi verificada de forma semelhante em todos os modelos. A robustez do resultado indica que a instabilidade dos parâmetros parece não depender da consistência entre as duas POFs.

7 CONCLUSÃO

O trabalho conclui que modelos estimados com dados defasados da POF não foram historicamente capazes de produz projeções confiáveis da taxa de poupança agregada das famílias brasileiras. Este resultado é robusto em relação aos quatro modelos de projeção selecionados, cujas especificações incluem diferentes conjuntos de variáveis preditivas com disponibilidade anual de dados. Os três modelos lineares são avaliados diretamente segundo a decomposição de Blinder-Oaxaca. A existência de uma proporção elevada de famílias com poupança nula motiva o uso do modelo de Heckman como a quarta alternativa, a qual é avaliada por meio de uma extensão da decomposição anterior conhecida como decomposição de Neuman-Oaxaca.

Mais especificamente, ambas as decomposições mostram que a diferença entre as taxas de poupança agregada estimadas com as versões da POF de 2002 e 2008 decorreu principalmente de mudanças nos parâmetros estimados dos modelos de projeção – o que reflete uma mudança do padrão de comportamento das famílias –, restando um papel secundário para o efeito de mudanças nas médias amostrais das características preditivas. Em decorrência da baixa frequência da POF, este resultado compromete o uso de modelos estimados com dados desta pesquisa na projeção da taxa de poupança agregada das famílias brasileiras.

REFERÊNCIAS

ATTANASIO, O.; SZÉKELY, M. Ahorro de los hogares y distribución del ingreso en México. **Nueva Época**, Toluca, v. 8, n. 2, p. 267-338, 1998.

BANCO CENTRAL DO BRASIL. Taxa de poupança: uma análise regional. **Boletim Regional do Banco Central do Brasil**, v. 7, n. 1, p. 91-93, 2013.

BERNHEIM, B.; SCHOLZ, J. Private saving and public policy. *In*: POTERBA, J. **Tax policy and the economy**. Massachusetts: MIT Press, 1993.

BLINDER, A. S. Wage discrimination: reduced form and structural estimates. **The Journal of Human Resources**, v. 8, n. 4, p. 436-455, 1973.

BROWNING, M.; LUSARDI, A. Household saving: micro theories and micro facts. **Journal of Economic Literature**, v. 34, n. 4, p. 1797-1855, 1996.

BRUMBERG, R.; MODIGLIANI, F. Utility analysis and the consumption function: an interpretation of the cross-section data. *In*: KURIHARA, K. (Ed.). **Post-Keynesian economics**. New Brunswick: Rutgers University Press, 1954.

CARROLL, C. The buffer stock theory of saving: some macroeconomics evidence. **Brookings Papers on Economic Activity**, v. 2, p. 61-156, 1992.

DEATON, A. Franco Modigliani and the life-cycle theory of consumption. **BNL Quarterly Review**, v. 58, n. 233-234, p. 91-107, 2005.

DYNAN, K.; SKYNER, J.; ZELDES, S. Does the rich save more? **Journal of Political Economy**, v. 112, n. 2, p. 397-444, 2004.

FERNANDEZ-VILLAVERDE, J.; KRUEGER, D. Consumption over the life cycle: facts from consumer expenditure survey data. **The Review of Economics and Statistics**, v. 89, n. 3, p. 552-565, 2007.

FREITAS, P. G. **Poupança precaucional no Brasil**: uma análise com microdados da POF. 2010. Dissertação (Mestrado) – Insper, São Paulo, 2010.

FRIEDMAN, M. **A theory of the consumption function**. Princeton: Princeton University Press, 1953.

GOURINCHAS, P. O.; PARKER, J. Consumption over the life cycle. **Econometrica**, v. 70, n. 1, p. 47-89, 2002.

HECKMAN, J. J. Sample selection as a specification error. **Econometrica**, v. 47, n. 1, p. 153-161, 1979.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Sistema de Contas Nacionais**. Rio de Janeiro: IBGE, 2008. (Série Relatórios Metodológicos, n. 24).

JANN, B. Standard errors for the Blinder-Oaxaca decomposition. *In:* GERMAN STATA USERS GROUP MEETING, 3., 2005, Berlin. **Annals** […]. Berlin: ETH, 2005. Disponível em: https://goo.gl/hckpFp. Acesso em: 27 fev. 2018.

_____. The Blinder-Oaxaca decomposition for linear regression models. **The Stata Journal**, v. 8, n. 4, p. 453-479, 2008.

JAPPELLI, T.; MODIGLIANI, F. **The age-saving profile and the life-cycle hypothesis**. Napoli: Centre for Studies in Economics and Finance, 1998. (Working Paper, n. 9).

MOREIRA, A. B.; SILVEIRA, M. **Determinantes da taxa de poupança das famílias brasileiras**: evidência microeconômica com as POFs 2002-2003 e 2008-2009. Brasília: Ipea, 2015. (Texto para Discussão, n. 2070).

NEUMAN, S.; OAXACA, R. **Wage differentials in the 1990s in Israel**: endowments, discrimination, and selectivity. Bonn: IZA, 2004. (Working Paper, n. 10).

OAXACA, R. Male-female wage differentials in urban labor markets international. **Economic Review**, v. 14, p. 693-709, 1973.

OAXACA, R.; RANSOM, M. Calculation of approximate variances for wage decomposition differentials. **Journal of Economic and Social Measurement**, v. 24, p. 55-61, 1998.

PAXSON, C. Saving and growth: evidence of micro data. **European Economic Review**, v. 40, p. 255-288, 1996.

RIBEIRO, P. F. **Consumo, poupança precaucionária, lei de crédito consignado e impacto sobre alocação de ativos e distribuição de consumo**. 2015. Tese (Doutorado) – Escola de Economia de São Paulo, São Paulo, 2015.

BIBLIOGRAFIA COMPLEMENTAR

ATTANASIO, O. Consumption demand. *In:* TAYLOR, J. B.; WOODFORD, M. (Eds.). **Handbook of macroeconomics**. Amsterdam: North Holland-Elsevier, 1999.

A EFICIÊNCIA TÉCNICA DAS DISTRIBUIDORAS DE ELETRICIDADE DO BRASIL DURANTE 2003-2012: UMA APLICAÇÃO DE FUNÇÕES DE DISTÂNCIA ESTOCÁSTICAS[1]

Mariana De Santis[2]
Fernando Damonte[3]
Elías Curi[4]

Neste trabalho é estimada a eficiência dos custos operacionais e de capital das empresas de distribuição de eletricidade do Brasil no período 2003-2012, empregando funções de distância estocásticas com dados em painel e controlando pela heteroscedasticidade. Além de incluir os efeitos estocásticos, as estimativas do estudo são controladas pelo efeito das variáveis exógenas, que resultaram estatisticamente significantes. Conclui-se que a eficiência estimada do setor aumentou durante o período e que os efeitos das variáveis não gerenciáveis, como as diferenças regionais no preço do trabalho, a concentração dos consumidores e o ambiente socioeconômico em que as empresas operam, devem ser considerados na análise.

Palavras-chave: eficiência; dados em painel; distribuição de eletricidade.

TECHNICAL EFFICIENCY OF ELECTRICITY DISTRIBUIDORS IN BRAZIL DURING 2003-2012: AN APPLICATION OF STOCHASTIC DISTANCE FUNCTIONS

The efficiency of operational and capital costs of Brazilian electricity distribution companies in the period 2003-2012 is estimated in this paper by using stochastic distance functions with panel data models and controlling for heteroscedasticity. Additionally to including stochastic effects, the estimates of the study are controlled by the effect of exogenous variables that resulted statistically significant. The study concludes that the efficiency of the industry increased in the period and the effects of non-managerial variables such as regional differences in the price of labor, the concentration of consumers and socio-economic environment in which companies operate should be considered in the analysis.

Keywords: efficiency; panel data; utilities.

JEL: L94; D24.

1 INTRODUÇÃO

O setor de distribuição de energia elétrica do Brasil está em constante expansão e desenvolvimento. Composto por 63 distribuidoras encarregadas da administração e da operação de redes de baixa e média tensões, além de um conjunto de

1. Os autores agradecem profundamente a colaboração de José Marcio Goulart Junior e Marianela Restelli na revisão e na correção deste trabalho.

2. Especialista em serviços públicos.
3. Especialista em serviços públicos.
4. Especialista em serviços públicos.

cooperativas de eletrificação rural[5] que abastecem mais de 78 milhões de clientes. A grande extensão territorial do Brasil gera uma ampla diversidade climática, com uma enorme gama de relevos e zonas que vão desde selváticas e pouco povoadas até metropolitanas e com alta densidade populacional. Estas características impactam, em distinta medida, a realidade e os custos das distribuidoras, as quais podem ser muito díspares, dependendo do tamanho, da localização da área de concessão, da quantidade e dos tipos de consumidores, entre outras variáveis ambientais influentes.

A demanda elétrica cresceu de maneira sustentada ao longo dos anos, incentivando ao desenvolvimento de práticas regulatórias cada vez mais sofisticadas e específicas. A partir dos anos 1990, como parte de uma tendência mundial de restruturação, o Brasil entra em um processo de privatização emoldurado no chamado Plano Real. Este processo afetou de maneira direta o mercado de distribuição elétrica, mediante o Programa Nacional de Privatização, dando lugar à desverticalização do setor e ao começo da participação de empresas privadas no mercado elétrico.

A presença de distribuidoras de caráter não governamental, conjuntamente às distribuidoras estatais já existentes, deu nascimento, em 1997, à Agência Nacional de Energia Elétrica (Aneel), cuja principal finalidade é melhorar a eficiência do setor energético. Para o cumprimento desse objetivo, a Aneel leva a cabo revisões tarifárias periódicas (a cada quatro anos), baseadas em incentivos e aplicando metodologias que foram evoluindo com o tempo e são constantemente aprimoradas.

A aplicação de diferentes modelos para medir a eficiência por parte da Aneel tem gerado no setor um crescente interesse pelo estudo de diversas metodologias alternativas. Apesar da vigência de uma metodologia inovadora e baseada nos últimos antecedentes bibliográficos, a influência das variáveis que caracterizam o ambiente em que operam as distribuidoras não foi especificamente considerada. Contudo, essa diversidade ambiental cobra particular importância em um país como o Brasil, em que existe grande heterogeneidade geográfica, econômica e social.

Além de haver consenso técnico e acadêmico em que as variáveis ambientais influem nos custos da prestação do serviço elétrico, não existem estudos conclusivos que sejam referências no setor e que consigam explicar adequadamente seu efeito na eficiência das empresas. Nesse sentido, considera-se que ainda não foram utilizadas todas as abordagens e as metodologias disponíveis, entre as quais destacamos como exemplo os métodos paramétricos e a captura do efeito da heteroscedasticidade e dos efeitos fixos. Assim, em caso de lograr capturar o efeito das variáveis ambientais, seria possível medir a evolução da eficiência das empresas no tempo.

Contudo, no serviço elétrico existem possibilidades de substituição entre capital e trabalho. Este fenômeno toma cada vez maior relevância com a evolução

5. Dados do Anuário Estatístico de Energia Elétrica 2016 da Empresa de Pesquisa Energética (EPE).

tecnológica, que permite a automação e aumenta a eficiência na prestação do serviço. No entanto, são escassos os estudos de eficiência no Brasil que incluem explicitamente ambos os custos, com fins de capturar esse efeito tão importante.

Neste trabalho estima-se a eficiência dos custos das empresas de distribuição de eletricidade do Brasil no período 2003-2012, empregando a abordagem de funções de distância estocásticas com dados em painel, controlando pelo efeito das variáveis não gerenciáveis. Para isso, consideram-se como insumos os custos operacionais e os custos de capital. Desta maneira, a análise inclui as remunerações a todos os insumos empregados no fornecimento do serviço, e a eficiência estimada corresponde à totalidade dos custos de produção desta indústria. Além de incluir os efeitos estocásticos, as estimativas do estudo são controladas pelo efeito das variáveis exógenas, que resultarem estatisticamente significantes.

Este trabalho está organizado em cinco seções, incluindo esta introdução. Na seção 2 são apresentados os principais estudos regulatórios sobre a estimação da eficiência no setor de distribuição elétrica no Brasil. Na seção 3 é exposto o marco teórico sobre fronteiras de eficiência e é descrita a estratégia empírica adotada para o cálculo de eficiência no presente artigo, detalhando as variáveis consideradas, as técnicas usadas para a estimação e os resultados obtidos. Na seção 4 é realizada uma reflexão final, junto à contribuição do estudo na determinação dos níveis de eficiência do mercado de distribuição elétrica do Brasil. Por fim, a seção 5 traz as reflexões finais.

2 ESTUDOS SOBRE A EFICIÊNCIA NO SETOR ELÉTRICO DO BRASIL

A literatura sobre estimações de eficiência no setor elétrico é muito vasta, em razão da ampla variedade de técnicas de estimação e estudos acadêmicos realizados para diferentes países. No entanto, a evidência nos países latino-americanos é limitada devido a problemas de disponibilidade de dados e à relativamente recente introdução de técnicas de *benchmarking* para a definição de tarifas. Entre os trabalhos mais relevantes para o Brasil, pode-se citar Mota (2004), que utiliza a análise envoltória de dados (DEA, do inglês *data envelopment analysis*) para decompor os índices de Malmquist e funções de distância orientadas aos insumos, comparando dados das empresas de distribuição elétrica do Brasil em relação a distribuidoras similares nos Estados Unidos, nos anos de 1994 e 2000. Uma contribuição de Mota (2004) é o emprego dos custos operacionais e totais como insumos para obter estimativas da eficiência nos custos, incluindo em alguns modelos variáveis que podem ser consideradas exógenas, como a demanda máxima, a densidade de clientes e a proporção de clientes residenciais no total de clientes. Os resultados dos modelos de fronteira estocástica com dados em painel indicam uma diferença tecnológica significativa a favor das empresas do Brasil, incluindo

somente os custos operacionais como a totalidade dos custos. Não obstante, a taxa de mudança tecnológica estimada foi maior para as empresas americanas. Contudo, o estudo de Mota foi realizado com dados de quinze anos atrás, tal que as conclusões já não seriam aplicáveis à realidade atual.

Posteriormente, Tovar, Ramos-Real e De Almeida (2009) analisaram a decomposição da produtividade em eficiência técnica, eficiência de escala e mudança técnica, utilizando dados de dezessete empresas brasileiras entre 1998 e 2005. Os autores estimaram funções de distância empregando o modelo de Battese-Coelli. Consideraram-se como *outputs* a quantidade de eletricidade distribuída e o número de consumidores, e como *inputs* as variáveis físicas (trabalho, comprimento de rede e perdas de energia na rede). Uma das conclusões finais sugere que poderiam se justificar as fusões de distribuidoras, por seu potencial de gerar ganhos futuros na produtividade, mas se deve notar que não controla nenhum efeito ambiental. Na mesma linha, Tovar, Ramos-Real e De Almeida (2009) estimam a evolução da produtividade na indústria de distribuição de eletricidade no Brasil entre 1998 e 2005, com o objetivo de avaliar o impacto das reformas realizadas neste período no setor, utilizando um painel de dezessete empresas (cerca de 55% do consumo de eletricidade no Brasil em 2005), com dados extraídos de relatórios da Associação Brasileira de Distribuidores de Energia Elétrica (Abradee), complementados com as informações fornecidas nos relatórios anuais das empresas em estudo. Os autores calcularam o índice de produtividade total dos fatores de Malmquist empregando DEA para analisar a mudança na produtividade do setor. As variáveis consideradas como produtos das distribuidoras foram à energia distribuída e ao número de clientes, enquanto os insumos incluídos foram o número de empregados, a extensão da rede (como *proxy* do capital físico) e as perdas totais. Finalmente, foi incorporado como variável ambiental o tamanho da área de serviço. Além da desvantagem de trabalhar com insumos físicos, o trabalho de Tovar, Ramos-Real e De Almeida (2009) controla a heterogeneidade das distribuidoras somente com a variável "área de serviço", o que não é suficiente para capturar o impacto de outras variáveis não gerenciáveis que influem nos custos, como as características climáticas ou as diferenças salariais.

Mais recentemente, Silva (2011) analisa a correspondência entre os índices de eficiência dos custos operacionais e de manutenção (Opex, do inglês *operational expediture*) reconhecidos pelo regulador brasileiro, mediante a comparação da empresa modelo (um enfoque construtivo de determinação dos custos) com os *scores* derivados dos métodos DEA e análise de fronteira estocástica (SFA, do inglês *stochastic frontier analysis*). Com relação à mudança de eficiência técnica pura, o estudo identifica que os serviços públicos de distribuição estão se afastando da fronteira de eficiência e não melhoraram seu rendimento, independentemente do tamanho da empresa. Silva detalha que os posicionamentos das distribuidoras,

mediante empresa modelo e SFA, não se correlacionam de maneira significativa. O autor também roda modelos DEA em três estágios para incorporar o impacto de variáveis ambientais. No entanto, somente considera os Opex na análise e compara-os com os valores do método de "empresa de referência", cujo uso já é praticamente obsoleto, e os dados correspondem a um período muito antigo.

Uma das publicações mais recentes sobre cálculos da eficiência de distribuidoras elétricas do Brasil é a de Leme *et al.* (2014), quem consideraram as variáveis não gerenciais como indispensáveis na análise de eficiência comparativa. Estes autores procuraram realizar uma contribuição à metodologia levada em consideração pela Aneel no terceiro ciclo de revisões tarifárias, sugerindo o emprego do desenho experimental para escolher as variáveis ambientais relevantes no cálculo do Opex reconhecido na receita. O cálculo foi realizado para as 29 distribuidoras consideradas como de grande porte (maiores) do Brasil, com dados publicados pela agência reguladora. Os *scores* de eficiência foram calculados em uma primeira etapa com um modelo DEA com rendimentos não decrescentes a escala e mínimos quadrados corrigidos, replicando a metodologia da Aneel. Na segunda etapa, foi avaliada a contribuição das variáveis ambientais com o critério bayesiano de informação, para escolher aquelas mais relevantes. Os autores simularam situações com presença e ausência de variáveis exógenas até dois fatores contextuais (efeitos individuais e conjuntos de até duas variáveis). Concluíram que a variável "consumidores por área" não resultou significativa por ela mesma, assim como já tinha acontecido na metodologia realizada pela Aneel. No entanto, as inter-relações entre "salário médio-consumidores por área" e "salário médio-índice de complexidade" apresentaram significância estatística. A contribuição do estudo é que, além das dependências lineares, as dependências não lineares devem ser consideradas na análise. As conclusões comparativas exibem uma diferença nos Opex anuais permitidos de 2,1% entre os métodos, equivalente a US$ 205 milhões, dos quais US$ 187,5 milhões representam déficits anuais nas empresas. Embora o estudo inclua insumos monetários e variáveis ambientais dos períodos mais recentes, ele é baseado na metodologia do terceiro ciclo adotada pela Aneel, já modificada em 2015.

O trabalho de Souza *et al.* (2014) propõe a utilização de uma técnica DEA com rendimentos constantes à escala e restrição nos pesos contingentes para estimar a eficiência dos custos operacionais das distribuidoras elétricas do Brasil. A contribuição do estudo é o agrupamento das empresas tendo em conta os níveis de utilidade. Os autores indicam que a metodologia proposta permite reduzir as assimetrias do modelo da Aneel sem grupamentos nem restrições de pesos e com rendimentos variáveis à escala. A análise considera os Opex como *input* e a energia distribuída, as unidades de consumo e as redes de distribuição como *outputs*, mas não considera variáveis ambientais; portanto, não contabiliza as características contextuais das distribuidoras e só considera questões intrínsecas a elas.

A revisão acadêmica descrita leva-nos à conclusão de que a literatura sobre eficiência na distribuição de energia elétrica no Brasil, ainda interessante, está referida à eficiência dos custos operacionais, mas sem incorporar o capital físico, além de trabalhar com informação antiga e, na maioria dos casos, não aplicar metodologias apropriadas de estimação de fronteiras estocásticas com dados em painel.

3 MARCO TEÓRICO: FRONTEIRAS DE EFICIÊNCIA

Para estimar a eficiência técnica das distribuidoras de energia elétrica do Brasil foi empregada uma abordagem paramétrica, aplicando as mais recentes técnicas de estimação para dados em painel, em razão de dispor de dados de numerosas empresas em um período de dez anos. Foi especificada uma função de distância estocástica orientada aos insumos, abordagem que apresenta vantagens com relação às funções de produção e custos, já que permite incorporar vários produtos obtidos com os mesmos insumos, tanto que a função de produção somente admite um único produto, além de não assumir a conduta minimizadora de custos e evitar o problema da endogeneidade dos preços dos insumos (Coelli *et al.*, 2003). Esta abordagem tem sido aplicada por Farsi e Filippini (2004), Farsi, Filippini e Greene (2005), Growitsch, Jamasb e Pollitt (2005), Goto e Tsutsui (2008), Cullmann (2009) e Pérez-Reyes e Tovar (2009) em estudos sobre eficiência do setor de distribuição de eletricidade em diversos países utilizando dados em painel.

A partir das estimativas da função distância são obtidas estimativas da eficiência técnica, isto é, a capacidade das empresas para obter um nível determinado de produto com a mínima quantidade possível de insumos. Estritamente, para que uma firma seja tecnicamente eficiente, é necessário que ela esteja alocada na isoquanta, que define o limite tecnológico do conjunto das diferentes combinações de insumos, os quais permitem obter um determinado vetor de produtos. Esta medida de eficiência não leva em conta os preços dos insumos nem a proporção em que eles são empregados, diferentemente da eficiência econômica, que indica a capacidade das empresas para obter um determinado nível de produção a custo mínimo.[6]

A função distância pode ser considerada como uma função de produção com mais de um produto, o que permite salvar a limitação de trabalhar com um único produto. Resulta muito apropriada para analisar empresas que produzem serviços, como as distribuidoras de eletricidade, cuja produção resulta muito melhor quantificada considerando mais de uma dimensão; em geral, clientes atendidos, cumprimento das redes e energia distribuída são as mais usuais.

6. A eficiência econômica implica eficiência técnica e também eficiência locativa, definida como a habilidade para combinar otimamente os insumos, os dados e os seus preços.

3.1 Função de distância empírica

Define-se a função de distância como a distância entre a localização ou a posição de uma firma e o limite das possibilidades de produção. De acordo com Kumbhakar e Lovell (2001), uma função de distância orientada aos insumos, $D_I(y,x)$, define-se da seguinte maneira:

$$D_I(y,x) = max\{\lambda: x/\lambda \in L(y)\},$$

onde D_I é a distância de cada firma até a fronteira; x é um vetor de n insumos; y é um vetor de m produtos; λ é uma constante maior ou igual à unidade; e $L(y)$ representa as combinações de insumos que permitem obter y. A função de distância orientada para os insumos representa a máxima quantidade em que o vetor de insumos de uma empresa deve ser radialmente contraído para seguir produzindo o mesmo vetor de produtos. Os conceitos de distância indicados foram desenvolvidos na década de 1950,[7] e a sua utilização em estudos de eficiência é muito difundida. Nos casos em que se tenta avaliar a eficiência de firmas que devem ajustar o uso dos insumos para fazer frente a uma demanda exógena, como ocorre com as distribuidoras de eletricidade,[8] recomenda-se utilizar o enfoque da função de distância orientada para os insumos, pois no setor de energia elétrica o número de consumidores atendidos e os demais produtos das distribuidoras constituem fatores exógenos, ou seja, a distribuidora não possui maior controle/ingerência sobre a quantidade de usuários. Portanto, a função de distância orientada para os insumos adapta-se melhor ao contexto da distribuição de energia elétrica, razão pela qual é escolhida para os trabalhos descritos a seguir.

Formalmente, segundo Kumbhakar e Lovell (2001), a eficiência técnica orientada aos insumos ET_I pode ser definida da seguinte maneira:

$$ET_I(y,x) = min\{\theta: D_I(y,\theta x) \geq 1\}.$$

A expressão anterior indica a contração radial de todos os insumos no vetor x, que deveria ocorrer para ficar alocado na isoquanta que define o limite tecnológico. Quando a posição da firma está em dita isoquanta, já não é necessário realizar nenhum ajuste nos insumos, resultando máxima a eficiência técnica e nula a distância à fronteira. A eficiência técnica orientada para os insumos é igual à recíproca à função distância, podendo se expressar da seguinte forma:

$$ET_I(y,x) = \frac{1}{D_I(y,x)} = \frac{x/\lambda^*}{x} = \theta, \quad (1)$$

7. Debreu (1951) e Farrell (1957) são os trabalhos pioneiros.

8. Geralmente as distribuidoras de energia elétrica estão obrigadas a atender todos os clientes de uma área determinada.

onde θ representa a eficiência técnica. A igualdade anterior indica que: quanto maior for a distância ao limite do conjunto de combinações viáveis de insumos, menor será a eficiência técnica orientada aos insumos. Note-se que ET_I assume valores entre 0 e 1. O valor 1 corresponde à situação na qual a firma está alocada na isoquanta e, portanto, não é necessário reduzir as quantidades de insumos, já que estão sendo empregados nas mínimas quantidades possíveis.

A função de distância possui a propriedade de homogeneidade de grau um nos insumos, que estabelece que, caso o uso dos insumos seja incrementado em uma determinada quantidade, mantendo-se constante a quantidade de produtos obtidos, a distância à fronteira será incrementada também nessa quantidade. Esta propriedade pode se expressar do seguinte modo:

$$D_I(y, \lambda x) = \lambda D_I(y, x),$$

sendo $\lambda > 0$. Fazendo $\lambda = \frac{1}{x_1}$ e substituindo na igualdade anterior:

$$D_I(y, \frac{x}{x_1}) = \frac{1}{x_1} D_I(y, x). \quad (2)$$

Tomando logaritmo natural em ambos os membros e aplicando *-lnx*$_1$, obtém-se:

$$-lnx_1 = lnD_I(y, \frac{x}{x_1}) _ lnD_I\,(y, x). \quad (3)$$

O sinal negativo do logaritmo natural da função de distância representa o desvio da fronteira determinística de produção, o qual depende de efeitos aleatórios, erros de medida e de ineficiências estocásticas não negativas.

Portanto, pode-se escrever:

$$-lnD_I(y, x) = v_{it} - u_{it}. \quad (4)$$

Os erros v_{it} representam os desvios aleatórios em relação à fronteira de produção, os quais podem ser positivos ou negativos, sendo sua média igual a zero; enquanto os erros u_{it} representam desvios por baixo da fronteira de produção atribuídos à ineficiência, razão pela qual sempre apresentam sinal negativo.

Para completar, a expressão na equação (3) é expressa em forma funcional como $lnD_I(y, \frac{x}{x_1})$. Para isso, pode-se supor uma estrutura Cobb-Douglas, obtendo-se a seguinte expressão:

$$-lnx_{1it} = \alpha_i + \sum_m \beta_m\, lny_{mit} + \sum_{n-1} \beta_n \ln\,(\frac{x_{n-1it}}{x_1}) + v_{it} - u_{it}. \quad (5)$$

A estimação da função especificada anteriormente permite o cálculo da eficiência técnica orientada para os insumos, a qual, como já detalhado previamente, é igual à inversa da função distância. Portanto, pode-se escrever:

$$ET_I(y,x) = \frac{1}{D_I(y,x)} = \frac{1}{E[\exp(u_{it})|v_{it}-u_{it}]}. \quad (6)$$

4 ESTRATÉGIA EMPÍRICA

Para estimar as funções de distância e a eficiência das empresas distribuidoras de eletricidade do Brasil, utilizaram-se dados de 58 distribuidoras correspondentes ao período 2003-2012, na medida em que estiveram disponíveis as observações de todas as variáveis incluídas na análise. Os dados foram retirados do *site* da Aneel.

4.1 Variáveis incluídas na análise

A seguir são definidas as variáveis que representam os insumos (*inputs*), os produtos (*outputs*) fornecidos pelas empresas da indústria e as variáveis ambientais ou exógenas que afetam ou deslocam a fronteira, seja favorável, seja desfavoravelmente. Essas variáveis são aquelas tradicionalmente consideradas neste tipo de estudo, em concordância com o trabalho pioneiro de Neuberg (1977) e o clássico de Philip e Weyman-Jones (1996).

4.1.1 *Inputs*

No presente estudo, foram considerados como insumos aos Opex e à anuidade do custo de capital (Ck). A consideração simultânea destes dois tipos de custos separados permite analisar a complementariedade e/ou substituição existente entre ambos, e constitui um avanço em relação aos estudos de eficiência que consideram unicamente o Opex ou o custo total agregado (Totex). A inclusão do custo de capital no modelo permite aproximar a quantidade e a qualidade do capital físico, com a vantagem de possibilitar explicitar a complementariedade com os custos operacionais. Dita análise não é possível quando a eficiência é somente estimada para os custos operacionais, resultando favorecidas aquelas empresas que adotaram uma combinação de insumos intensiva em capital.

Quando são comparadas empresas heterogêneas, com diferentes realidades e trajetórias, como as distribuidoras de eletricidade do Brasil, é muito provável que se observem situações bastante dissimiles. Em consequência, não seria correto estimar a eficiência destas empresas considerando somente um grupo de custos por separado (por exemplo, somente os Opex), já que se estaria obviando esta situação e, em consequência, distorcendo os resultados.

Pelas razões descritas, no presente estudo é abordada a análise da eficiência com modelos que levam em consideração tanto os custos operacionais quanto os custos de capital.

Custos operacionais

Os custos operacionais, para fins de revisão tarifária, correspondem aos custos com: a) pessoal; b) materiais; c) serviço de terceiros; d) outros custos operacionais; e) tributos; e f) seguros relativos à atividade de distribuição e comercialização. No Manual de Contabilidade do Setor Elétrico correspondem às subcontas 615.03.X.X.XX e 615.05.X.X.XX. Os dados são referentes aos valores históricos informados e foram atualizados em moeda de junho de 2013 com o Índice Geral de Preços do Mercado (IGPM).

Custos de capital

Os custos de capital foram estimados como uma anuidade constante da base de ativos regulatórios. Para o cálculo dessa anuidade, é considerada a seguinte equação:

$$Ck = BC\ x\ \frac{TCC}{1-\frac{1}{(1+TCC)^n}},$$

onde:

- *Ck*: anuidade do custo de capital;
- *BC*: base de capital. Corresponde à base de ativos bruta, sobre a qual é aplicado o fator de recupero de capital, que considera tanto a depreciação quanto a rentabilidade dele;
- *TCC*: taxa de custo de capital. Neste artigo é considerada a taxa WACC depois de impostos;
- *n*: anos de vida útil contábil.

A base foi construída a partir dos valores da base de remuneração bruta definidos pela Aneel nas revisões tarifárias que ocorreram entre 2003 e 2013. Foi considerado o total dos ativos, sejam próprios, sejam doados, aplicando sobre eles a equação detalhada previamente. Este critério é consistente com a ideia de que a eficiência é medida em relação a todos os recursos envolvidos nos processos de prestação do serviço elétrico, independentemente da propriedade e da situação contábil deles, sempre que estejam em serviço. A aplicação efetiva dos resultados deste enfoque tem que ser finalmente efetuada sobre a base de ativos próprios, sob o suposto de que a ineficiência obtida aplica de igual maneira tanto aos ativos próprios quanto aos doados e demais *inputs* dos modelos.

Todos os valores econômicos sofreram correção monetária pelo IGPM, ficando expressos a preços de junho de 2013. Não foram considerados os bens 100% depreciados, já que o custo de capital deles já foi recuperado na sua totalidade pelas distribuidoras. Tampouco foram considerados os valores do índice de aproveitamento aplicados sobre alguns ativos.

A taxa WACC depois de impostos foi aplicada segundo corresponde a cada distribuidora, mas considerando que as mudanças da taxa para os diferentes ciclos tarifários são efetuadas no mesmo ano para todas as distribuidoras, especificamente no primeiro ano de cada ciclo de revisões tarifárias.[9] Isso faz com que o cálculo seja coerente na aplicação da taxa.

É importante efetuar também uma distinção entre o custo de capital utilizado no presente artigo e as despesas de capital, chamadas de Capex (*capital expenditures*), na literatura do tema. As Capex representam os investimentos anuais efetuados pelas empresas distribuidoras para fornecer o serviço, seja por expansão dele, seja para reposição de ativos antigos. Estes investimentos anuais não são adequados para representar o nível de custo de capital das empresas, já que não trazem em conta o custo de ativos anteriores ainda em serviço.

Contudo, para efetuar uma estimativa do custo de capital que permita uma correta comparação entre as empresas, sugere-se o cálculo de uma anuidade constante da base bruta total de ativos das distribuidoras. Aqui é importante destacar também que, no caso de comparar o custo de capital líquido das empresas, estaríamos considerando como ineficientes aquelas empresas que possuem redes mais novas e, em consequência, um maior custo de capital calculado sobre o capital líquido. Paralelamente, seriam consideradas como eficientes empresas com redes velhas ou obsoletas, para as quais é observado um menor custo de capital na sua base líquida. Portanto, é fundamental nivelar essa questão, e a melhor forma é calculando uma anuidade constante de custo de capital.

4.1.2 *Outputs*

As empresas de distribuição prestam um serviço que não é perfeitamente quantificável e deve ser aproximado mediante diferentes variáveis. Neste estudo são utilizados os produtos tradicionalmente utilizados na bibliografia: a quantidade de consumidores, o cumprimento da rede e a energia distribuída ponderada pelos níveis de tensão, como representativas da produção das distribuidoras de eletricidade. Os dados foram obtidos das tabelas publicadas pela Aneel no marco da Consulta Pública nº 11/2013.

9. Como a aplicação da taxa WACC é atualizada em ocasião da revisão tarifária de cada distribuidora e elas não acontecem em um mesmo ano, pode-se gerar inconsistência nos critérios adotados. Para salvar a situação, considerou-se que a taxa WACC é atualizada no mesmo ano para todas as distribuidoras.

4.1.3 Variáveis ambientais

As variáveis ambientais são aquelas que não estão sob o controle gerencial das empresas, ou seja, são exógenas e afetam positiva ou negativamente a função de produção subjacente. Podem ser consideradas na especificação das funções a estimar, para controlar seu efeito sobre o nível dos custos ou, alternativamente, nos erros do modelo, como será explicado mais adiante. No presente estudo foram consideradas aquelas variáveis que os especialistas do setor assinalam como relevantes. Os dados foram obtidos das estimativas da eficiência no marco da Consulta Pública nº 11/2013 ou proporcionados pela Abradee:

- *salários:* indicador da relatividade de custo de mão de obra entre as distribuidoras, obtido como a projeção do nível médio salarial de ocupações relacionadas à distribuição de energia na área de concessão da distribuidora, expresso em moeda de junho de 2013;
- *chuva:* média de precipitação acumulada anual, em mm, no período de 1999 a 2012, na área de atuação de cada distribuidora. Os dados são obtidos do Instituto Nacional de Pesquisas Espaciais (Inpe);
- *descargas atmosféricas:* densidade de descargas atmosféricas médias, em km², por área de atuação, para o período de apuração 1998-2009. Informação fornecida pelo Operador Nacional do Sistema Elétrico (NOS) – Mapa de Densidade de Descargas Atmosféricas;
- área de atuação: delimitação de uma "área de atuação" para cada distribuidora, em km², que consiste na envoltória em torno da rede de alta e média tensões, considerando um 1 km de *buffer*;
- *declividade:* indicador de declividade média do relevo da área de atuação de cada concessionária (%), construído a partir da média aritmética das declividades dos terrenos na área de atuação (SRTM 2000 – Mapa de Declividade 200 metros);
- *vegetação remanescente:* indicador de incidência elevada de vegetação natural na área de atuação de cada concessionária, que implique maiores custos operacionais. Expresso em km²;
- *consumidores por quilômetro de rede:* quantitativo de consumidores da concessionária no mês de dezembro de cada ano dividido pelos quilômetros de rede totais de cada ano;
- *consumidores por área de atuação:* quantitativo de consumidores da concessionária no mês de dezembro de cada ano dividido pela área de atuação de cada empresa para cada ano;

- *coeficiente de Gini:* indicador do grau de desigualdade na distribuição da renda domiciliar *per capita* entre os indivíduos, obtido a partir do Censo Demográfico 2010. É calculado sobre a renda nominal mensal dos domicílios particulares permanentes, por situação do domicílio.

4.1.4 Estatística descritiva

Como já antedito, as empresas que compõem o setor elétrico brasileiro apresentam uma grande heterogeneidade. A tabela 1 detalha, além do universo de dados utilizados nos modelos, a grande dispersão das variáveis-chave que existe entre as concessionárias.

TABELA 1
Descrição das variáveis

Variável	Unidade	Média	Desvio-padrão	Mínimo	Máximo	Dados
Opex	R$ mil	191.262	281.013	703	1.873.440	515
Custos de capital	R$ mil	259.722	409.688	174	1.963.210	515
Km total	Km	48.852	76.251	43	498.353	515
Consumidores	-	1.066.550	1.474.590	2.169	7.460.090	515
Consumidores/área	Consumidores/km^2	82	187	3	1.493	515
Área de atuação	Km^2	46.536	77.298	15	425.628	515
Declividade	%	6,90	4,04	1,13	23,30	515
Descargas atmosféricas	Descargas/km^2	6,82	3,05	0,54	15,43	515
Vegetação	Km^2	12.859	21.881,3	0,6	92.193	515
Chuva	Mm/ano	1.437	329	697,5	2.196,7	515
Índice de Gini	-	0,5	0,0	0,4	0,6	515
Salários	R$/mês	4.067	699,9	2.372	7.271,2	515
Mercado ponderado	MWh	2.497.390	3.661.480	5.778	20.726.600	515

Elaboração dos autores.

4.2 Modelo empírico

Estimou-se a fronteira estocástica com os dados de toda a amostra, utilizando uma função distância orientada aos insumos. A partir dos erros de ineficiência de dita estimação, obtiveram-se os *scores* de eficiência segundo a equação (6). O modelo estimado com as variáveis mencionadas no item anterior é uma função de distância do tipo Cobb-Douglas, conforme a seguinte expressão:

$$-lnCK_{it} = \alpha_i + \beta_{UC} lnUC_{it} + \beta_{ener} ln\ ener_{it} + \beta_{km} lnkm_{it} + \beta_{opex} \ln\left(\frac{OPEX_{it}}{CK_{it}}\right) + \gamma_{salar} ln\ salario_{it} + \gamma_{chuva} ln\ chuva_{it} + \gamma_{desc} ln\ desc_{it} + \gamma_{decl} ln\ decl_{it} + \gamma_{dens} ln\ dens_{it} + \gamma_{areal} ln\ area_{it} + \gamma_{desig} ln\ desig_{it} + v_{it} - u_{it}. \quad (7)$$

Nela, os coeficientes β acompanham as variáveis que representam os *outputs* e os γ as que representam os fatores ambientais fora do controle gerencial. A constante α_i tem um papel diferente, que depende do modelo de dados em painel escolhido para a estimação, como se assinalará mais adiante. O termo v é o erro aleatório do modelo e o termo u é o erro atribuído à ineficiência gerencial, que representa a distância de cada empresa à fronteira, não justificada pelas variáveis do modelo (as distribuições dos termos u e v são definidas mais adiante). Assim, a partir da estimação de u, é obtido o *score* de eficiência de cada empresa como a inversa da distância:

$$ET_I(y,x) = \frac{1}{D_I(y,x)} = \exp(-u).$$

Note-se que, no modelo, foi escolhida a variável *Ck* para garantir a homogeneidade de grau um nos insumos da função distância, e é por isso que aparece no lado esquerdo da função e também como denominador da variável *Opex* no lado direito. Os resultados não mudariam caso se utilize como divisor a variável *Opex*.

4.2.1 Técnicas de estimação

Uma vantagem relevante das técnicas econométricas é a possibilidade de incluir variáveis ambientais, tanto na função quanto nas distribuições dos erros do modelo. Contrariamente, os métodos não paramétricos (como a DEA, por exemplo), embora fáceis de resolver, apresentam dificuldades no momento de incorporar as variáveis exógenas que não constituem nem insumos nem produtos, e também no tratamento com dados em painel. Apesar de uma extensa literatura sugerir complementar os modelos DEA com estimações paramétricas em uma segunda etapa (Simar e Wilson, 2007), ou aplicar DEA a unidades de decisão que se desempenham em meios homogêneos (eficiência condicionada, proposta por Badin, Daraio e Simar, 2012), no presente estudo foram priorizadas as vantagens dos modelos de verdadeiros efeitos fixos com heteroscedasticidade para o tratamento das variáveis exógenas, já que se trata de uma técnica mais específica e adequada para o tratamento de dados em painel.

Rodaram-se os modelos que supõem eficiência invariante no tempo: modelos de efeitos fixos tradicionais, modelos de efeitos aleatórios tradicionais e a variante de Battese-Coelli; e modelos de eficiência variante no tempo. Estes últimos consistiram em modelos de verdadeiros efeitos fixos, nos quais é depurado o erro de

ineficiência de cada empresa do impacto da heterogeneidade não observada, de maneira que cada ordenada à origem reflete o impacto de variáveis não incluídas no modelo, mas que afetam o desempenho da empresa e constituem atributos invariantes no tempo. Adicionalmente, estimaram-se os modelos de verdadeiros efeitos fixos com heterogeneidade explícita no desvio-padrão e na média dos erros. Estes últimos apresentaram alguns problemas de convergência ao máximo valor da função de verossimilhança, pelo que seus resultados não serão considerados. O método de verdadeiros efeitos aleatórios não foi aplicado em razão de que se trata de um método instável.

As variáveis ambientais podem entrar no modelo modificando a altura da função. Assim, aquelas distribuidoras que enfrentem maiores salários terão a mesma função que as restantes, mesmo que deslocada para cima. E se a densidade de clientes for um fator que diminui custos, as distribuidoras com maior densidade terão sua função distância deslocada para baixo em comparação com as outras. Ou seja, as variáveis ambientais nas especificações anteriores modificam a altura da função. Existem, além disso, outras abordagens para o tratamento do efeito que as variáveis ambientais exercem sobre os custos de produção, que consistem em inclui-las nas distribuições dos erros do modelo. Desse jeito, é possível provar se os erros do modelo são independentes das variáveis explicativas ou se, pelo contrário, sua variância ou sua média é modificada pelo valor que assumem as variáveis ambientais.

Neste trabalho é estimada uma função distância que incorpora as variáveis ambientais ($z_{it} = ln\ salario_{it}$, $ln\ chuva_{it}$, $ln\ desc_{it}$, $ln\ decl_{it}$, $ln\ dens_{it}$, $ln\ area_{it}$, $ln\ desig_{it}$) na função, na variância dos erros separadamente e/ou conjuntamente. Assim, o modelo estimado pode se expressar, em linhas gerais, como:

$$-lnCK_{it} = \alpha_i + \beta_{UC} lnUC_{it} + \beta_{ener} ln\ ener_{it} + \beta_{km} lnkm_{it} + \beta_{opex} \ln\left(\frac{OPEX_{it}}{CK_{it}}\right) + \gamma_{salar} ln\ salario_{it} + \gamma_{chuva} ln\ chuva_{it} + \gamma_{desc} ln\ desc_{it} + \gamma_{decl} ln\ decl_{it} + \gamma_{dens} ln\ dens_{it} + \gamma_{areal} ln\ area_{it} + \gamma_{desig} ln\ desig_{it} + v_{it} - u_{it}\ .$$

Os erros v_{it} apresentam uma distribuição normal, com média zero e variância constante, e os erros u_{it} são maiores ou iguais a zero e distribuem-se de acordo com uma seminormal truncada em zero com variância constante.

Note-se que, nestes modelos, os efeitos fixos α_i "justificam" um maior ou menor custo. Quando a ineficiência é uma característica não observável que permanece sem mudanças no período em análise, então o erro gerencial vê-se em queda. Uma possível alternativa para evitar a subestimação da ineficiência é incluir o efeito das variáveis ambientais na variância dos dois erros:

$$-lnCK_{it} = \alpha + \beta_{UC} lnUC_{it} + \beta_{ener} ln\ ener_{it} + \beta_{km} lnkm_{it} + \beta_{opex} \ln\left(\frac{OPEX_{it}}{CK_{it}}\right) + v_{it} - u_{it},$$

onde:

$$v_{it} \sim N\left(0, \sigma_{v_{it}}^2\right), \sigma_{v_{it}}^2 = \sigma_v^2 . \exp\left(\delta' z_{it}\right),$$

$$u_{it} \sim \left|N(0, \sigma_{u_{it}}^2)\right|, \sigma_{u_{it}}^2 = \sigma_u^2 . \exp\left(\alpha_i + \varphi' z_{it}\right).$$

Nestes modelos as variâncias dos erros não são constantes, mas são determinadas pelas variáveis ambientais z. Além disso, a variância dos u_i também é uma função das variáveis não observadas de cada empresa, capturadas nos efeitos fixos α_i.

Além de todos os métodos possuírem suas respectivas vantagens e desvantagens, considera-se que, para o caso da comparação dos custos de empresas de distribuição de eletricidade, o método de verdadeiros efeitos fixos, que incorpora o impacto das variáveis ambientais na variância dos erros, evita a possibilidade de sobrestimar a eficiência, no caso em que a incapacidade gerencial tiver ficado constante no período considerado. Este efeito é reconhecido por Greene (2003), quando assinala que a separação do termo de erro entre um efeito fixo e a componente não gerenciável é um fator positivo destes modelos. Porém, no caso em que a empresa tiver ficado ineficiente, sem reagir aos incentivos, o modelo percebe isto como uma característica que não muda no tempo, similar ao que acontece na localização ou no tamanho da área que é considerado como um efeito fixo. Em consequência, o erro de ineficiência é depurado da incapacidade de gerência, sendo a ineficiência estimada menor que a real. Estes modelos permitem também um tratamento mais apropriado das variáveis ambientais, já que possibilitam estimações da eficiência para cada momento do tempo e corrigem a heteroscedasticidade. Ademais, permitem introduzir a heterogeneidade não observada na variância dos erros, evitando sobrestimar a eficiência. No entanto, a seguir são apresentados resultados de modelos com estes métodos, mas também de outros modelos mais simples que têm sido testados: efeitos aleatórios e efeitos fixos. As estimativas são realizadas por máxima verossimilhança e a decomposição do erro total, em aleatório e gerencial, é feita pelo método de Jondrow *et al.* (1982).

4.2.2 Resultados

Nesta seção são apresentados os resultados dos modelos paramétricos estimados a partir da função distância, anteriormente especificados. Para a estimação, todas as variáveis foram expressas em logaritmos naturais e cada observação foi dividida pela média geométrica da respectiva série de dados. Desta maneira, suaviza-se o efeito dos valores extremos (*outliers*), e os parâmetros obtidos na estimação da função são interpretados como elasticidades, valoradas na média da série (Coelli *et al.,* 2003).

A seguir são apresentados os resultados dos coeficientes dos modelos estimados que, embora sejam de menor interesse quando comparados com as estimativas de eficiência, constituem o primer estágio do processo de estimação. A tabela 2 contém os resultados dos modelos tradicionais de efeitos aleatórios supondo uma distribuição seminormal para o erro atribuído à ineficiência gerencial, com média zero e desvio-padrão constante (modelos homoscedásticos) nas colunas 2 e 3, e de efeitos fixos na última coluna. A tabela 3 corresponde aos modelos de verdadeiros efeitos fixos, os quais incorporam os efeitos fixos como constantes que modificam a altura da função distância[10] e supõem homoscedasticidade nas variâncias dos erros. Finalmente, a tabela 4 apresenta os resultados dos modelos de verdadeiros efeitos fixos, os quais consideram os desvios-padrão dos erros como função dos efeitos fixos das empresas e/ou de variáveis ambientais (modelos heteroscedásticos).

TABELA 2
Modelos tradicionais

Variável	Modelo efeitos aleatórios			Modelo efeitos aleatórios Battese-Coelli			Modelo efeitos fixos tradicionais		
	Coeficiente	Desvio-padrão	Coeficiente/desvio-padrão	Coeficiente	Desvio-padrão	Coeficiente/desvio-padrão	Coeficiente	Desvio-padrão	Coeficiente/desvio-padrão
Opex/Ck	0,40310	0,01542	26,1378	0,37942	0,01834	20,6870	0,38005	0,02038	18,6467
Salário	-0,18404	0,06486	-2,8373	-0,29614	0,05894	-5,0240	-0,16783	0,05659	-2,9659
Densidade (consumidores/km_rede)	0,35088	0,05951	5,8962	0,11945	0,03217	3,7128	-	-	-
Constante	0,68102	0,07883	8,6396	0,43182	0,04739	9,1114	-0,42366	0,14987	-2,8268
Sigma	0,84697	0,09682	8,7478	0,41100	0,01366	30,0765	-	-	-
Lambda	10,54212	4,55160	2,3161	5,23381	0,03833	136,5335	-	-	-
Eta	-	-	-	0,04186	0,00627	6,6733	-	-	-
R^2 ajustado	-	-	-	-	-	-		0,9988862	
Log L		406,71			449,51		-	-	-
N		515			515			515	
Empresas		58			58			58	

Elaboração dos autores.

10. Os coeficientes não são apresentados por razões de espaço, mas podem ser solicitados aos autores.

TABELA 3
Modelos de verdadeiros efeitos fixos homoscedásticos

Variável	Modelo verdadeiros efeitos fixos			Modelo verdadeiros efeitos fixos			Modelo verdadeiros efeitos fixos		
	Coeficiente	Desvio-padrão	Coeficiente/desvio-padrão	Coeficiente	Desvio-padrão	Coeficiente/desvio-padrão	Coeficiente	Desvio-padrão	Coeficiente/desvio-padrão
Km de rede	-0,18821	0,01250	-15,05726958	-0,09744	0,02153	-4,5268	-0,12749	0,02226	-5,7265
Consumidores	-0,29913	0,02362	-12,66237956	-0,50631	0,02983	-16,9745	-0,39288	0,03134	-12,5379
Mercado	-0,56126	0,02669	-21,02580086	-0,43214	0,02284	-18,9194	-0,51481	0,02625	-19,6111
Opex/Ck	0,42737	0,02057	20,77569698	0,45310	0,01985	22,8268	0,43340	0,01976	21,9329
Salário	-	-	-	-0,67558	0,04222	-16,0026	-0,66516	0,04220	-15,7614
Densidade (consumidores/km_rede)	-	-	-	0,06981	0,01071	6,5205	0,06069	0,01070	5,6739
Desigualdade (coef. de Gini)	-	-	-	-	-	-	0,84030	0,10343	8,1246
Sigma	0,2536	0,8336	0,3043	0,25440	0,00838	30,3478	0,21092	0,00707	29,8292
Lambda	0,0001	0,0276	0,0025	0,83294	0,05933	14,0395	0,00006	0,02761	0,0022
Log L		180,3284			249,15		-	266,41	-
N		515			515			515	
Empresas		58			58			58	

Elaboração dos autores.

TABELA 4
Modelos de efeitos fixos heteroscedásticos (EFHet)

Variável	Modelo 1			Modelo 2			Modelo 3		
	Coeficiente	Desvio-padrão	Coeficiente/desvio-padrão	Coeficiente	Desvio-padrão	Coeficiente/desvio-padrão	Coeficiente	Desvio-padrão	Coeficiente/desvio-padrão
Km de rede	-0,01908	0,04365	-0,43722	-0,05629	0,03262	-1,72571	-0,07197	0,04137	-1,73986
Consumidores	-0,61875	0,06289	-9,83861	-0,61662	0,06080	-10,14194	-0,49684	0,07079	-7,01888
Mercado	-0,38343	0,05693	-6,73510	-0,36504	0,05545	-6,58264	-0,45948	0,05608	-8,19281
Opex/Ck	0,46555	0,03137	14,83931	0,44834	0,02819	15,90700	0,44264	0,02861	15,46993
Salário	-0,55187	0,08706	-6,33932	-0,53202	0,07728	-6,88422	-0,48935	0,08136	-6,01451
Densidade (consumidores/área)	0,15679	0,02720	5,76491	0,12795	0,02094	6,11112	0,12931	0,02658	4,86447
Chuva	-	-	-	-	-	-	-0,13297	0,10069	-1,32058
Desigualdade (coef. de Gini)	-	-	-	-	-	-	-0,95407	0,21055	-4,53138
Declividade	-	-	-	-	-	-	-	-	-
Constante	0,12393	0,02201	5,62971	0,13342	0,01959	6,81133	0,12723	0,02020	6,29892

(Continua)

(Continuação)

Variável	Modelo 1			Modelo 2			Modelo 3		
	Coeficiente	Desvio-padrão	Coeficiente/desvio-padrão	Coeficiente	Desvio-padrão	Coeficiente/desvio-padrão	Coeficiente	Desvio-padrão	Coeficiente/desvio-padrão
Parâmetros da variância do erro v									
Constante	-4,56096	0,11798	-38,65773	-4,83863	0,13753	-35,18167	-4,67199	0,13357	-34,97685
Chuva	-	-	-	-2,16025	0,57907	-3,73054	-	-	-
Coeficiente de Gini	-	-	-	-8,82147	3,18487	-2,76980	-	-	-
Salário	-	-	-	-	-	-	-	-	-
Densidade (consumidores/área)	-	-	-	-	-	-	-	-	-
Declividade	-	-	-	-	-	-	-	-	-
Parâmetros da variância do erro u[1]									
Constante	-5,50878	2,38156	-2,31310	-4,13596	1,31740	-3,13948	-5,94812	2,73919	-2,17149
Salário	-	-	-	-	-	-	-	-	-
Densidade (consumidores/área)	-	-	-	-	-	-	-	-	-
Chuva	-	-	-	-	-	-	-	-	-
Coeficiente de Gini	-	-	-	-	-	-	-	-	-
Log L		336,62			347,56			352,95	
AIC		-541,2			-559,1			-569,9	
N		515			515			515	
Empresas		58			58			58	

Variável	Modelo 4			Modelo 5			Modelo 6		
	Coeficiente	Desvio-padrão	Coeficiente/desvio-padrão	Coeficiente	Desvio-padrão	Coeficiente/desvio-padrão	Coeficiente	Desvio-padrão	Coeficiente/desvio-padrão
Km de rede	-0,12732	0,02675	-4,75976	0,00038	0,03528	0,01084	-	-	-
Consumidores	-0,45549	0,05002	-9,10542	-0,71953	0,06524	-11,02884	-0,71925	0,05513	-13,04720
Mercado	-0,44684	0,04341	-10,29317	-0,31940	0,05696	-5,60715	-0,31929	0,05645	-5,65667
Opex/Ck	0,45309	0,02443	18,54433	0,43489	0,03033	14,34048	0,43786	0,02930	14,94324
Salário	-	-	-	-0,52042	0,08601	-6,05080	-0,51919	0,08701	-5,96689
Densidade (consumidores/área)	-	-	-	0,17013	0,02239	7,59981	0,16534	0,01763	9,37654
Chuva	-	-	-	-	-	-	-	-	-
Desigualdade (coef. de Gini)	-	-	-	-	-	-	-	-	-
Declividade	-	-	-	-0,12054	0,05242	-2,29948	-0,11396	0,03322	-3,43019
Constante	0,11733	0,01759	6,67011	0,13184	0,02093	6,29857	0,13104	0,02045	6,40915

(Continua)

(Continuação)

Variável	Modelo 4			Modelo 5			Modelo 6		
	Coeficiente	Desvio-padrão	Coeficiente/desvio-padrão	Coeficiente	Desvio-padrão	Coeficiente/desvio-padrão	Coeficiente	Desvio-padrão	Coeficiente/desvio-padrão
Parâmetros da variância do erro v									
Constante	-4,82758	0,21946	-21,99755	-4,83634	0,15419	-31,36540	-4,81960	0,15216	-31,67521
Chuva	-1,88995	0,71868	-2,62977	-2,10240	0,59673	-3,52324	-1,98748	0,57345	-3,46580
Coeficiente de Gini	-11,57967	4,56962	-2,53406	-8,27681	3,06345	-2,70180	-8,84605	2,93509	-3,01390
Salário	-4,37757	1,01944	-4,29410	-	-	-	-	-	-
Densidade (consumidores/área)	0,64135	0,16466	3,89502	-	-	-	-	-	-
Declividade	-	-	-	0,46212	0,50619	0,91294	-	-	-
Parâmetros da variância do erro u[1]									
Constante	-22,12834	81981900	-0,00000	-5,00246	2,20555	-2,26813	-4,71814	1,54014	-3,06344
Salário	2,76921	3,47098	0,79782	-	-	-	-	-	-
Densidade (consumidores/área)	-6,84062	2,96995	-2,30327	-	-	-	-	-	-
Chuva	44,06630	34941800	0,00000	-	-	-	-	-	-
Coeficiente de Gini	5,85825	9,57111	0,61208	-	-	-	-	-	-
Log L		300,64			359,18			357,09	
AIC		-456,9			-578,4			-578,2	
N		515			515			515	
Empresas		58			58			58	

Elaboração dos autores.
Nota: [1] Não são expostos os coeficientes dos efeitos fixos das 58 distribuidoras por razões de espaço de exposição.

Como pode ser observado nas tabelas anteriores, todos os coeficientes estimados apresentam o sinal esperado e são, salvo poucas exceções, estatisticamente significativos. A relação direta entre os custos e as variáveis que aproximam o produto das distribuidoras é confirmada: comprimento da rede, quantidade de unidades consumidoras e quantidade de energia distribuída. Note-se que estes coeficientes apresentam sinal negativo, igual ao custo do lado direito, indicando que estas variáveis movem-se no mesmo sentido ou têm relação direta. Como assinalado anteriormente, os coeficientes estimados representam elasticidades e, como as variáveis estão expressas em desvios, podem ser interpretados como elasticidades na média da amostra. Assim, os resultados do modelo 3 (EFHet) indicam que, para cada 1% de incremento na quantidade de consumidores, os custos incrementam-se, *ceteris paribus*, em 0,497%.

A variável comprimento de rede apresenta um nível de significância muito baixo em alguns modelos, talvez pela elevada correlação com as unidades consumidoras e com a energia distribuída, como no modelo 5 (EFHet). O modelo 6 (EFHet) não apresenta o comprimento de rede como variável.

Em todas as estimativas é confirmada a relação direta esperada entre o custo de capital e as variáveis que representam a produção das distribuidoras: comprimento de rede, quantidade de unidades consumidoras e volume de energia vendida. Note-se que os coeficientes destas variáveis têm sinal negativo, igual ao custo de capital, no lado direito da equação estimada. Isto significa que, entre os custos e cada um dos *outputs* existe uma relação direta, já que ambas variáveis têm o mesmo sinal. Como mencionado, o coeficiente das variáveis representa a elasticidade do custo em relação a cada *output*. Assim, por exemplo, em função do modelo de efeitos aleatórios da tabela 2, o coeficiente dos quilômetros de rede é *0,106*, o que indica que, ante um incremento de 1% no comprimento da rede, ambos os custos incrementam-se na média da indústria, *ceteris paribus*, em *0,106*%. Note-se que, pela propriedade de homogeneidade de grau um da função distância, é indiferente considerar, no lado direito da equação, Ck ou Opex. Assim, se, no lugar de considerar como variável dependente do modelo o Ck, houvesse considerado o Opex (bem como a sua inversa de Opex/Ck a Ck/Opex), os coeficientes de todas as variáveis seriam iguais, com exceção da relação Ck/Opex, que seria igual a um menos o valor do coeficiente Opex/Ck.

Outro resultado muito relevante é que a razão Opex/Ck sempre tem associado um coeficiente positivo. Isto indica que, caso se incremente o Opex mantendo sem mudanças as variáveis que representam o *output* e as ambientais, o custo de capital diminui. Essa relação inversa entre os dois insumos do modelo confirma que existe uma relação de substituição entre eles, que, na média das empresas do estudo, é em torno de *0,4*, interpretando-se que é possível diminuir em *0,4%* o custo de capital por cada *1%* de incremento no Opex.

Analisando as variáveis ambientais ou exógenas, algumas destas variáveis não resultaram significativas sobre os custos em nenhum dos casos, como as descargas e a vegetação. No caso das descargas, pode ser devido à elevada correlação identificada com a variável chuva. Como foram apresentadas anteriormente, as variáveis ambientais foram incorporadas como uma variável que afeta diretamente a função e como explicativas das variâncias dos erros. Na primeira abordagem (tabelas 2 e 3), as variáveis densidade da rede e salário médio da região resultaram altamente significativas e com o sinal esperado na maioria das estimações.

A densidade, medida como consumidores por unidade de superfície da área de concessão, apresenta sinal contrário ao custo em todas as estimações e é estatisticamente significativa. Sua elasticidade oscila entre 0,13 e 0,17, segundo

o modelo especificado, e indica que um incremento de 1% na concentração dos consumidores do serviço reduz os custos, *ceteris paribus*, entre 0,13% e 0,17%.

O nível de salários do setor é uma variável ambiental que tanto empresas quanto regulador reconhecem como um determinante relevante dos custos, especialmente operativos. Seu nível flutua de uma região a outra, refletindo o custo de oportunidade dos trabalhadores. Em zonas nas quais há mais alternativas de trabalho, os trabalhadores têm um maior custo de oportunidade ao empregar-se no setor da distribuição de eletricidade. Por isso, as empresas localizadas nas zonas mais prósperas e desenvolvidas pagam, em geral, salários mais elevados do que as demais. Esta hipótese se cumpre em todas as estimações, pois em todos os casos é observada uma relação direta entre o nível do custo e o salário, com elevado nível de significância estatística. O valor estimado da elasticidade do custo em relação ao salário varia com o modelo entre 0,17 e 0,55, resultando mais elevada nos modelos de verdadeiros efeitos fixos.

As variáveis restantes não apresentaram significância, ou sua inclusão determinou a não convergência ao máximo valor da função de verossimilhança, razão pela qual não são apresentados esses resultados. Uma exceção a se destacar são as variáveis de declividade e desigualdade econômica – aproximada com o coeficiente de Gini (modelos 3 e 6 da tabela 4, respectivamente). A desigualdade induz a um incremento nos custos, com uma elasticidade média de *0,95*, similar à declividade, cuja elasticidade-custo é de *0,12*. Este resultado indicaria que as distribuidoras que operam em zonas de maior pobreza estão mais expostas a situações de roubo de energia (perdas não técnicas), devendo instalar redes e equipes de maior custo, incorrendo em maiores custos operativos e de manutenção.

As estimativas do parâmetro $\sigma(\sqrt{\sigma_v^2 + \sigma_u^2})$ são estatisticamente distintas de zero, como também aquelas do parâmetro $\lambda\left({}^{\sigma_u}/_{\sigma_v}\right)$, podendo-se afirmar que a variabilidade da ineficiência gerencial não é nula e que é, além disso, superior àquela atribuída ao erro não sistemático (note-se que, nos modelos com heteroscedasticidade, estes parâmetros não são únicos; eles devem ser calculados para cada empresa, já que dependem dos valores de suas variáveis ambientais e de suas respectivas ordenadas à origem).

Em relação ao efeito das variáveis ambientais na dispersão dos erros, aprecia-se que o coeficiente de Gini, os salários da região e o nível de precipitações afetam negativamente a variância do erro aleatório, tanto que a densidade tem o efeito inverso, confirmando a heteroscedasticidade (tabela 4). Estes resultados expõem que a variância do erro não sistemático não é constante; ela se modifica de acordo com as variáveis exógenas das empresas, definidas como significativas. Contrariamente, a dispersão dos erros sistemáticos ou de ineficiência não depende das variáveis

ambientais nem dos efeitos fixos a_{ii}[11], com exceção do modelo 4 (EFHet), em que uma menor densidade contribui para uma maior dispersão, podendo-se concluir que sua dispersão em torno da média é constante.

Para confirmar os resultados descritos no parágrafo anterior, realizaram-se os testes de hipótese com base na razão de verossimilhança, para provar se os modelos restringidos são estatisticamente iguais aos não restringidos, em termos das variáveis ambientais incorporadas na função distância e/ou nas variâncias dos erros do modelo. As duas primeiras linhas da tabela 5 demonstram as provas de igualdade a zero dos coeficientes de variáveis ambientais para os modelos apresentados nas tabelas 3 e 4. Em todos os casos a hipótese é recusada, devido aos elevados níveis de confiança da hipótese nula. Adicionalmente, os valores do critério de informação de Akaike sugerem que os modelos 5 e 6, que consideram a heterogeneidade, são os melhores.

TABELA 5
Testes de razão de verossimilhança: modelo de verdadeiros efeitos fixos

Modelos sem heteroscedasticidade							
Modelo	H_0	Parâmetros	*Log L*	Estatístico λ	Restrições	Prob.	Decisão
1	$\gamma_w=\gamma_{dens}=0$	66	249,15	137,6424	2	0	Rejeição
2	$\gamma_w=\gamma_{dens}=\gamma_{desig}=0$	67	266,41	172,1596	3	0	Rejeição
Modelos com heteroscedasticidade							
Modelo	H_0	Parâmetros	*Log L*	Estatístico λ	Restrições	Prob.	Decisão
1	$\gamma_w=\gamma_{dens}=0$	66	336,62	172,6526	2	0	Rejeição
2	$\gamma_w=\gamma_{dens}=\gamma_{prec}=\gamma_{desig}=0$	68	352,95	205,3312	4	0	Rejeição
3	$\delta_{prec}=\delta_{desig}=0$	68	347,56	194,5464	4	0	Rejeição
4	$\delta_w=\delta_{dens}=\delta_{prec}=\delta_{desig}=0$; $\varphi_w=\varphi_{dens}=\varphi_{prec}=\varphi_{desig}=0$	72	300,64	100,7048	8	0	Rejeição
5	$\gamma_w=\gamma_{dens}=\gamma_{decliv}=0$; $\delta_{prec}=\delta_{desig}=\delta_{decliv}=0$	70	359,18	217,775	6	0	Rejeição
6	$\gamma_w=\gamma_{dens}=\gamma_{decliv}=0$; $\delta_{prec}=\delta_{desig}=0$	68	357,09	213,607	4	0	Rejeição

Elaboração dos autores.

4.2.3 Análise da eficiência

A partir das estimações apresentadas nas tabelas 3, 4 e 5, obtiveram-se as porcentagens de eficiência técnica de cada uma das 58 distribuidoras incluídas na amostra. Como já assinalado, os modelos de verdadeiros efeitos fixos com heteroscedasticidade resultaram os mais adequados. No entanto, também são apresentados os *scores* de eficiência obtidos com os outros modelos, a modo comparativo. Conforme já mencionado, os modelos tradicionais de efeitos fixos e efeitos aleatórios (tradicional

11. As ordenadas à origem das empresas não resultaram estatisticamente diferentes de zero, podendo concluir que a variância dos erros u_{it} é constante e igual para todas as empresas.

e a versão Battese-Coelli, 1992) supõem que a eficiência manteve-se constante no período analisado. A variante do modelo de efeitos aleatórios de Battese e Coelli estima uma taxa média de evolução da eficiência para todas as empresas da amostra (o parâmetro *eta*).[12] A estimativa do parâmetro *eta* resultou positiva, indicando um crescimento médio anual da eficiência de cada empresa de 2,16%. Apesar de a opção de Battese e Coelli poder parecer superior ao modelo tradicional de efeitos aleatórios, tem a limitação de supor que a taxa *eta* é constante e igual para todas as empresas, suposto que parece pouco provável se levarmos em consideração que as 58 empresas apresentam características díspares em função de suas áreas de concessão, tipo e quantidade de consumidores, assim como da sua própria evolução. Segundo os modelos de verdadeiros efeitos fixos, a eficiência do setor apresentou taxas médias de crescimento anuais positivas, que vão de 0,49% a 1,25%.

Análise dos scores

A seguir é apresentada uma análise sobre a eficiência técnica das empresas da amostra, a partir das estimativas apresentadas na seção anterior. Em primeiro termo, consideram-se os *scores* de eficiência de cada distribuidora de energia elétrica no final do período. A tabela 6 apresenta os *scores* do período final, para cada um dos modelos realizados. Para a maioria das empresas, foi considerado o ano 2012. Para alguns casos em que esse dado não estava disponível, foi utilizado o ano 2011.

TABELA 6
***Scores* de eficiência do período final**
(Em %)

Empresa	Modelos tradicionais		Verdadeiros efeitos fixos homoscedásticos	Efeitos fixos heteroscedásticos					
	Efeitos aleatórios	Efeitos aleatórios BC	Modelo 3	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
AES- Sul	0,59	0,75	0,90	1,00	1,00	1,00	1,00	1,00	1,00
AME	0,17	0,47	0,92	0,70	0,79	0,87	0,96	0,79	0,78
Bandeirante	0,21	0,50	0,92	0,87	0,88	0,86	0,93	0,89	0,90
Boa Vista	0,29	0,50	0,91	0,71	0,72	0,80	0,77	0,79	0,71
Caiuá	0,55	0,75	0,89	0,99	0,96	0,97	0,96	0,94	0,94
Ceal	0,51	0,67	0,90	0,82	0,85	0,84	0,93	0,84	0,85
CEB	0,20	0,51	0,92	0,92	0,96	0,96	0,90	0,92	0,92
CEEE	0,43	0,63	0,90	0,91	0,92	0,90	0,92	0,91	0,90
Celesc	0,38	0,63	0,92	0,96	0,95	0,95	0,95	0,98	0,97
Celg	0,75	0,77	0,92	1,00	1,00	1,00	0,98	1,00	1,00

(Continua)

12. A taxa de crescimento da eficiência de um período a outro é dada por $u_i e^{-h(t-T)} (e^{-h}-1)$.

(Continuação)

Empresa	Modelos tradicionais		Verdadeiros efeitos fixos homoscedásticos	Efeitos fixos heteroscedásticos					
	Efeitos aleatórios	Efeitos aleatórios BC	Modelo 3	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
Celpa	0,73	0,75	0,91	0,99	0,99	0,99	0,96	0,98	0,98
Celpe	0,76	0,91	0,91	1,00	1,00	1,00	1,00	1,00	1,00
Celtins	0,85	0,65	0,92	0,92	0,93	0,93	0,85	0,90	0,90
Cemar	0,93	0,82	0,90	1,00	1,00	1,00	0,99	1,00	1,00
Cemat	0,81	0,72	0,90	0,97	0,95	0,98	0,94	0,93	0,93
Cemig	0,46	0,57	0,92	0,89	0,90	0,87	0,89	0,92	0,92
Cepisa	0,95	0,82	0,91	0,96	0,96	0,96	0,97	0,96	0,96
CFLO	0,66	0,88	0,90	1,00	1,00	1,00	0,99	1,00	1,00
Chesp	0,98	0,88	0,90	0,99	0,97	0,98	0,89	0,97	0,97
Jaguari	0,23	0,55	0,91	0,91	0,89	0,93	0,96	0,86	0,87
Mococa	0,63	0,77	0,90	1,00	0,95	0,98	0,97	0,96	0,96
Santa Cruz	0,60	0,81	0,91	1,00	1,00	1,00	0,98	0,98	0,98
Nacional	0,51	0,68	0,91	0,92	0,90	0,91	0,94	0,92	0,92
Cocel	0,72	0,76	0,91	1,00	1,00	0,96	0,98	1,00	1,00
Coelba	0,80	0,82	0,91	0,99	1,00	0,98	1,00	1,00	1,00
Coelce	0,51	0,76	0,92	0,96	0,93	0,93	0,95	0,91	0,89
Cooperaliança	0,42	0,62	0,89	0,92	0,94	0,91	0,93	0,95	0,95
Copel	0,72	0,78	0,92	0,96	0,96	0,96	1,00	0,95	0,94
Cosern	0,58	0,73	0,92	0,97	0,96	0,97	0,97	0,96	0,96
CPEE	0,59	0,74	0,91	0,97	0,96	0,97	0,96	0,96	0,96
Piratininga	0,36	0,63	0,90	0,99	0,98	0,97	0,98	0,97	0,98
CPFL Paulista	0,22	0,56	0,91	0,93	0,93	0,93	0,99	0,92	0,93
CSPE	0,65	0,78	0,91	1	0,96	0,98	0,97	0,96	0,96
Demei	0,42	0,88	0,89	0,99	0,99	1	1	0,93	0,94
DME-PC	0,27	0,53	0,88	0,62	0,61	0,62	0,79	0,66	0,65
Ene. Borborema	0,74	0,92	0,90	0,98	0,98	0,97	1	0,98	0,98
Vale Paranapanema	0,48	0,67	0,91	0,91	0,91	0,91	0,94	0,94	0,94
Bragantina	0,61	0,77	0,90	1,00	0,97	0,98	0,97	0,96	0,95
Joao Cesa	0,45	0,80	0,90	0,90	0,87	0,82	0,95	0,88	0,88
Urussanga	0,28	0,58	0,89	0,77	0,74	0,73	0,92	0,73	0,72
Elektro	0,44	0,61	0,92	0,95	0,95	0,94	0,95	0,95	0,95
Electroarce	0,75	0,73	0,88	0,90	0,83	1,00	0,85	0,77	0,76
Electrocar	0,86	0,91	0,90	1,00	0,99	1,00	1,00	0,99	0,98
Electropaulo	0,08	0,37	0,92	0,72	0,77	0,73	0,90	0,73	0,74
Santa Maria	0,61	0,69	0,92	0,93	0,91	0,92	0,95	0,94	0,94

(Continua)

(Continuação)

Empresa	Modelos tradicionais		Verdadeiros efeitos fixos homoscedásticos	Efeitos fixos heteroscedásticos					
	Efeitos aleatórios	Efeitos aleatórios BC	Modelo 3	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
Ene. Minas Gerais	0,64	0,69	0,91	0,91	0,91	0,90	0,93	0,95	0,95
Enersul	0,66	0,62	0,91	0,89	0,88	0,88	0,87	0,85	0,85
Ene. Nova Friburgo	0,40	0,68	0,91	0,90	0,90	0,93	0,95	0,97	0,97
Ene. Paraíba	0,94	0,91	0,90	1,00	1,00	1,00	1,00	1,00	1,00
Escelsa	0,44	0,61	0,91	0,90	0,90	0,90	0,95	0,93	0,94
Ene. Sergipe	0,64	0,76	0,90	0,95	0,95	0,95	0,98	0,94	0,94
Hidropan	0,52	0,74	0,89	0,86	0,82	0,86	0,96	0,80	0,81
Iguaçu	0,56	0,74	0,92	0,95	0,93	0,94	0,94	0,94	0,94
Light	0,16	0,47	0,91	0,82	0,85	0,85	0,90	0,86	0,87
Muxfeldt	0,64	0,98	0,91	1,00	1,00	1,00	1,00	1,00	1,00
RGE	0,57	0,75	0,91	1,00	1,00	1,00	1,00	1,00	1,00
Sulgipe	0,97	0,82	0,90	0,88	0,87	0,87	0,88	0,86	0,87
Nova Palma	0,85	0,84	0,91	0,97	0,95	0,97	0,96	0,96	0,96

Elaboração dos autores.

Como pode ser observado, dez distribuidoras estão alocadas na fronteira ao final do período, com base nas estimativas dos modelos EFH 5 e 6. São elas: AES-Sul, Celg, Celpe, Cemar, CFLO, Cocel, Coelba, ENE-Paraíba, Muxfeldt e RGE.

Para analisar a evolução da eficiência de cada uma das distribuidoras e da indústria, compararam-se os *scores* da tabela apresentada previamente (período final) com aqueles obtidos pelos diferentes métodos em anos prévios (período inicial). Foi tomado o ano 2003 como período inicial. No caso de falta de dados, foi considerado o ano mais próximo com informação disponível. A comparação efetuada é exposta na tabela 7, na qual são detalhadas as porcentagens de variação de eficiência entre os dois períodos. Pela lógica do modelo de efeitos aleatórios, a variação em todos os casos é zero, pelo que não é incorporada na tabela de resultados. Pode-se ver que, na maioria dos casos, as empresas manifestaram incrementos no nível de eficiência, isto é, variações positivas entre o período 2003-2012. Das 58 distribuidoras, somente oito tiveram uma variação negativa na eficiência técnica. Foram elas: Cemat, Cooperaliança, Demei, DME-PC, Urussanga, Eletroacre, Sulgipe e Novapalma. Cabe destacar que algumas empresas apresentaram variação nula em alguns de seus *scores*, tendo uma eficiência máxima nos dois períodos. Isto é, não sofreram uma estagnação, senão que se mantiveram nos pontos máximos de eficiência. Exemplos delas são: AES-Sul, Celg, Celtins, Celpe, CFLO, Chesp, Coelba, Demei, Electrocar, ENE-Paraíba, Muxfeldt e RGE.

TABELA 7
Variações de eficiência do período inicial ao período final
(Em %)

Empresa	Modelos tradicionais	Verdadeiros efeitos fixos homoscedásticos	Efeitos fixos heteroscedásticos					
	Efeitos aleatórios BC	Modelo 3	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
AES-Sul	13,8	2,5	0,0	0,0	0,0	0,0	0,0	0,0
AME	29,0	6,4	23,6	11,8	44,4	27,4	10,6	12,5
Bandeirante	32,0	4,5	23,2	25,5	17,5	19,2	22,2	23,1
Boa Vista	26,7	4,8	20,3	2,0	25,3	7,8	1,3	2,1
Caiuá	11,9	5,0	0,1	2,9	1,7	17,0	7,7	6,2
Ceal	14,8	1,5	4,1	6,7	6,3	17,1	5,2	5,2
CEB	31,0	17,0	62,1	20,2	38,8	3,2	16,0	16,4
CEEE	19,9	5,3	10,8	13,6	10,4	14,8	15,7	14,4
Celesc	20,3	5,0	6,1	3,2	12,7	16,4	0,5	0,7
Celg	9,2	7,3	0,0	0,0	0,0	19,9	0,0	0,0
Celpa	12,3	5,2	0,1	1,7	0,2	27,9	2,6	1,7
Celpe	3,3	6,6	0,0	0,0	0,0	0,0	0,0	0,0
Celtins	18,5	10,6	29,9	33,5	29,0	7,6	24,7	29,6
Cemar	7,1	1,5	0,0	0,0	0,0	6,0	0,0	0,0
Cemat	16,0	1,0	-0,2	-0,7	-0,2	13,8	-3,2	-2,8
Cemig	29,3	5,5	22,8	22,7	18,6	29,3	15,1	17,0
Cepisa	6,8	1,9	0,6	0,8	0,7	14,7	0,6	0,5
CFLO	5,3	1,7	0,0	0,0	0,0	3,4	0,0	0,0
Chesp	5,2	1,7	0,0	0,2	-0,1	-4,7	0,0	0,0
Jaguari	26,9	9,8	38,3	40,0	28,1	15,2	40,4	37,8
Mococa	10,7	7,9	0,0	14,5	1,2	14,5	7,1	8,2
Santa Cruz	8,9	7,1	0,0	0,0	0,0	11,1	1,0	0,6
Nacional	16,9	3,4	8,6	14,9	6,2	16,0	9,0	10,4
Cocel	13,6	8,6	0,0	0,0	6,5	15,4	0,0	0,0
Coelba	8,2	4,9	0,1	0,0	1,3	0,0	0,0	0,0
Coelce	9,7	5,7	7,7	5,6	19,9	8,0	17,4	12,7
Cooperaliança	21,3	-0,8	-1,4	-1,8	-3,5	10,4	-1,2	-1,3
Copel	12,1	7,0	6,1	4,2	10,9	0,0	9,4	7,5
Cosern	13,3	8,7	4,9	18,5	5,4	17,0	8,0	10,3
CPEE	13,0	4,7	1,4	5,4	2,1	17,3	4,1	4,3
Piratininga	23,3	3,3	0,1	0,4	0,6	5,2	0,9	0,7
CPFL Paulista	25,8	5,3	12,9	15,2	9,7	2,5	17,3	16,7

(Continua)

(Continuação)

Empresa	Modelos tradicionais	Verdadeiros efeitos fixos homoscedásticos	Efeitos fixos heteroscedásticos					
	Efeitos aleatórios BC	Modelo 3	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
CSPE	10,4	7,5	0,0	5,9	1,6	8,7	5,9	5,9
Demei	4,3	0,1	0,0	0,0	0,0	0,0	-1,8	-1,2
DME-PC	24,4	-2,1	-14,0	-12,9	-17,6	-5,9	-12,9	-13,3
Ene. Borborema	3,0	2,9	0,5	0,1	0,8	0,0	0,3	0,3
Vale Paranapanema	17,1	8,3	22,5	30,0	21,8	28,2	12,8	17,1
Bragantina	10,9	6,6	0,0	2,7	1,3	20,1	8,3	6,7
Joao Cesa	9,3	1,0	1,9	2,6	1,9	6,4	1,9	1,9
Urussanga	24,4	0,0	-5,1	-2,4	-5,6	22,2	-5,4	-5,4
Elektro	21,8	4,2	8,2	9,9	7,2	22,1	9,3	9,2
Electroarce	11,5	-1,8	-5,7	-13,3	0,0	-3,1	-19,4	-19,7
Electrocar	3,2	2,2	0,0	0,1	0,0	0,0	0,2	0,3
Electropaulo	48,5	8,0	31,0	15,2	27,4	46,3	10,5	11,3
Santa Maria	16,0	7,7	20,4	26,0	17,7	16,1	7,1	12,0
Ene. Minas Gerais	15,8	6,1	19,9	25,1	19,5	31,7	7,9	12,1
Enersul	24,3	2,3	4,2	8,7	4,5	15,1	7,1	6,7
Ene. Nova Friburgo	16,6	3,2	9,9	16,5	6,5	10,0	2,4	4,4
Ene. Paraíba	3,3	2,0	0,0	0,0	0,0	0,0	0,0	0,0
Escelsa	21,7	6,4	19,8	21,9	15,8	9,4	9,7	12,8
Ene. Sergipe	13,3	0,8	0,1	0,5	0,7	12,8	0,3	0,2
Hidropan	11,0	0,7	0,9	2,1	-2,5	4,3	0,6	0,9
Iguaçu	12,4	4,4	7,9	8,3	11,4	20,0	3,4	4,5
Light	41,7	6,2	31,3	22,3	28,0	31,2	29,7	26,0
Muxfeldt	0,6	2,9	0,0	0,0	0,0	0,0	0,0	0,0
RGE	13,9	5,0	0,0	0,0	0,0	0,0	0,0	0,0
Sulgipe	8,0	1,1	-0,4	1,6	1,9	6,6	-0,3	-0,2
Nova Palma	6,2	-0,3	-0,2	-1,4	-0,3	-0,9	-0,9	-1,1
Total indústria	**16**	**4**	**8**	**7**	**8**	**12**	**5**	**6**

Elaboração dos autores.

Com base nos modelos heteroscedásticos 5 e 6, os "melhores modelos", as empresas na indústria da distribuição de energia elétrica do Brasil incrementaram a sua eficiência entre 5% e 6% no período 2003-2012. Foram também calculadas as médias dos *scores* de todas as empresas para cada um dos métodos. Na tabela 8, podem ser apreciados os valores obtidos ao início e ao final do período, com cada modelo. Em seguida, é apresentado um gráfico que explicita dita comparação.

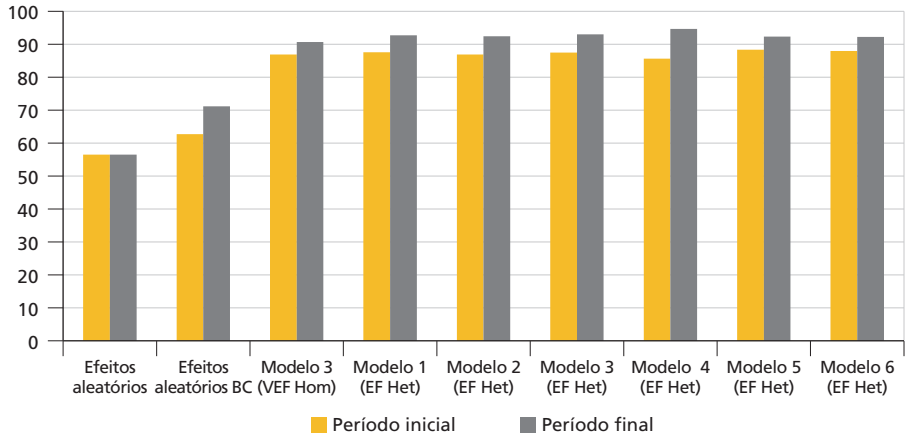
TABELA 8
Comparação dos *scores* de eficiência médios dos diferentes períodos
(Em %)

Período	Modelos tradicionais		Verdadeiros efeitos fixos homoscedásticos	Efeitos fixos heteroscedásticos					
	Efeitos aleatórios	Efeitos aleatórios BC	Modelo 3 (VEF Hom)	Modelo 1 (EF Het)	Modelo 2 (EF Het)	Modelo 3 (EF Het)	Modelo 4 (EF Het)	Modelo 5 (EF Het)	Modelo 6 (EF Het)
Período inicial	56,45	62,72	86,90	87,54	86,85	87,44	85,58	88,28	87,94
Período final	56,45	71,10	90,67	92,72	92,38	92,92	94,66	92,32	92,22

Elaboração dos autores.

A amplitude do intervalo dos *scores* de eficiência do período inicial está entre 56,45% do modelo de efeitos aleatórios e 88,28% do modelo 5 de efeitos fixos heteroscedásticos. No período final, a amplitude de intervalo fica entre 56,45% do modelo de efeitos aleatórios e 94,65% do modelo 4 de efeitos heteroscedásticos. Para todos os modelos, o *score* de eficiência média aumentou no período final, com relação ao período inicial. Pode-se observar isso no gráfico 1, exceto para o modelo de efeitos aleatórios, que é mantido constante devido à sua lógica de cálculo. Pode-se concluir que, independentemente do modelo empregado para estimar a fronteira à eficiência média na indústria da distribuição de energia elétrica do Brasil, incrementou-se essa eficiência no período analisado.

GRÁFICO 1
***Scores* de eficiência médios da indústria ao início e ao final do período**
(Em %)



Elaboração dos autores.

5 REFLEXÕES FINAIS

O presente estudo contribui para se conhecer com maior profundidade a eficiência do setor de distribuição de energia elétrica do Brasil nos últimos anos. Deve-se destacar que é um dos poucos trabalhos que incorpora em forma desagregada os custos de operação e os custos de capital, e que corrobora uma significativa substituibilidade entre ambos os custos (Ck e Opex). Este aspecto é de suma importância, já que em muitos estudos somente é estimada a eficiência dos custos de operação, deixando de lado a contraparte, isto é, os custos destinados a manter o capital investido. Portanto, a eficiência estimada neste trabalho é global. O eixo principal do estudo radica na consideração de variáveis ambientais dentro dos modelos, as quais permitam contemplar as disparidades existentes entre as empresas que excedem a capacidade gerencial delas.

A análise baseia-se nos dados de 58 distribuidoras brasileiras para o período 2003-2012. As estimativas de eficiência são obtidas mediante a construção de fronteiras estocásticas com dados em painel, que utilizam funções de distância do tipo Cobb-Douglas. Aplicaram-se métodos de estimação que supõem uma eficiência invariante no tempo, como os efeitos fixos tradicionais e os efeitos aleatórios, e sua variante Battese-Coelli. Por outra parte, como uma das particularidades do presente trabalho, incorporaram-se modelos de eficiência não variante no tempo, que consideram verdadeiros efeitos fixos e são capazes de depurar o impacto não desejado da ineficiência de cada empresa do impacto das heterogeneidades não observadas. Foi possível testar hipóteses significativas a respeito da mudança de eficiência no período, por meio da estimação de modelos de dados em painel com eficiência variante no tempo, demonstrando que a variância do erro aleatório é afetada por variáveis ambientais, mas não assim a eficiência gerencial. Este resultado indica que a capacidade gerencial das empresas é independente dos fatores ambientais, sugerindo que os tomadores de decisões nas empresas possuem capacidade de adaptação aos diferentes entornos. Este resultado contradisse os resultados expostos na literatura, em que há numerosos estudos que associam a ineficiência gerencial às particularidades do entorno das firmas.

Os resultados obtidos revelam uma relação direta esperada entre o custo de capital e as variáveis que representam a produção das distribuidoras, como são os quilômetros de rede, a quantidade de unidades consumidas e o volume de energia vendida. No concernente às variáveis ambientais, principal interesse do estudo, algumas estimativas resultaram sumamente significativas e outras não tanto.

A variável "densidade" foi significativa e apresentou um comportamento esperado e negativo em relação aos custos, afirmando que, quanto maior o número de consumidores por área, menores os seus custos unitários. Por outra parte, a variável "salário" também apresentou significância estatística em todos os casos

e uma relação direta e esperada com os custos, demonstrando que é necessário incorporar esta variável na estimação da eficiência. As variáveis ambientais restantes apresentaram efeitos diferentes, dependendo do modelo especificado. "Declividade", "desigualdade" (estimada pelo coeficiente de Gini) e "chuvas" somente resultaram estatisticamente significativas em alguns casos. Neles, observam-se empresas que possuem maiores custos quanto mais inclinadas são suas geografias; empresas cujas áreas de concessão encontram-se em zonas com maior pobreza e que estão expostas com maior probabilidade aos roubos; e, no último caso, que é mais custoso operar em áreas com maiores precipitações. As variáveis "descargas" e "vegetação" não apresentam significância estatística em nenhum dos casos, ou melhor, não apresentam os sinais esperados.

O trabalho demonstra que, no período considerado, a eficiência do setor apresenta um crescimento sustentado, fenômeno que merece uma análise posterior mais aprofundada, para determinar em que medida essa maior eficiência é resultado da regulação baseada em incentivos que vem sendo implementada no Brasil durante os últimos anos.

O estudo demonstra também que os custos eficientes do setor estão influenciados por variáveis não controláveis gerencialmente, em concordância com resultados de trabalhos previamente relevados. Este resultado permite recomendar a inclusão destes efeitos nos estudos da indústria de distribuição de energia elétrica, especialmente naqueles destinados a desenho de tarifas e incentivos para o setor.

REFERÊNCIAS

BADIN, L.; DARAIO, C.; SIMAR, L. How to measure the impact of environmental factors in a nonparametric production model. **European Journal of Operational Research**, v. 223, p. 818-833, 2012.

BATTESE, G.; COELLI, T. Frontier production function, technical efficiency and panel data with application to paddy farmers in India. **Journal of Productivity Analysis**, v. 3, p. 153-169, 1992.

COELLI, T. *et al.* **Una introducción a las medidas de eficiencia para reguladores de servicios públicos e de transporte**. Bogotá: Alfaomega Colombiana, 2003.

CULLMANN, A. Parametric and nonparametric efficiency analysis in electricity distribution: a European perspective. *In:* SCHOOL ECONOMICS AND MANAGEMENT, 7., 2009, Berlin. **Annals** […]. Berlin: Technische Universität Berlin, 2009.

DEBREU, G. The coefficient of resource utilization. **Econometrica**, v. 19, p. 273-292, 1951.

FARRELL, M. The measurement of productive efficiency. **Journal of the Royal Statistical Society A**, v. 120, p. 253-281, 1957.

FARSI, M.; FILLIPINI, M. Regulation and measuring cost efficiency with panel data models: application to electricity distribution utilities. **Review of Industrial Organization**, v. 25, n. 1, p. 1-19, 2004.

FARSI, M.; FILLIPINI, M.; GREENE, W. **Application of panel data models in benchmarking analysis of the electricity distribution sector**. Zurich: Cepe, 2005. (Cepe Working Paper, n. 39).

GOTO, M.; TSUTSUI, M. Technical efficiency and impacts of deregulation: an analysis of three functions in U.S. electric power utilities during the period from 1992 through 2000. **Energy Economics**, v. 30, p. 15-38, 2008.

GREENE, W. **Distinguishing between heterogeneity and inefficiency**: stochastic frontier analysis of the World Health Organization's Panel Data on National Health Care Systems. New York: New York University, 2003.

GROWITSCH, C.; JAMASB, T.; POLLIT, M. **Quality of service, efficiency and scale in networks industries**: an analysis of European electricity distribution. Germany: Halle Institute of Economic Research, 2005.

JONDROW, J. *et al.* On the estimation of technical inefficiency in the stochastic frontier production function model. **Journal of Econometrics**, v. 19, p. 233-238, 1982.

KUMBHAKAR, S.; LOVELL, C. A. **Stochastic frontier analysis**. Cambridge: Cambridge University Press, 2001.

LEME, R. *et al.* Design of experiments applied to environmental variables analysis in electricity utilities efficiency: the Brazilian case. **Energy Economics**, v. 45, p. 111-119, 2014.

MOTA, R. L. **Comparing Brazil and USA electricity distribution performance**: what was the impact of privatisation? Cambridge: CWPE, 2004. (Working Papers in Economics CWPE, n. 0423).

NEUBERG, L. G. Two issues in the municipal ownership of electricity power distribution systems. **The Bell Journal of Economics**, v. 8, p. 303-323, 1977.

PÉREZ-REYES, R.; TOVAR, B. Measuring efficiency and productivity change (PTF) in the Peruvian electricity distribution companies after reforms. **Energy Policy**, v. 37, p. 2249-2261, 2009.

PHILIP, B.; WEYMAN-JONES, T. G. Cost functions and cost efficiency in electricity distribution: a stochastic frontier approach. **Bulletin of Economic Research**, v. 48, p. 41-64, 1996.

SILVA, H. **Cost efficiency in periodic tariff reviews**: the reference utility approach and the role of interest groups. Gainesville: University of Florida, 2011.

SIMAR, L.; WILSON P. W. Estimation and inference in two-stage ,semi-parametric models of production processes. **Journal of Econometrics**, v. 136, p. 31-64, 2007.

SOUZA, M. V. *et al.* An application of data envelopment analysis to evaluate the efficiency level of the operational cost of Brazilian electricity distribution utilities. **Socio-Economic Planning Sciences**, v. 48, p. 169-174, 2014.

TOVAR, B.; RAMOS-REAL, F.; DE ALMEIDA, E. F. Firm size and productivity: evidence from the electricity distribution industry in Brazil. **Energy Policy**, v. 39, p. 826-833, 2009.

BIBLIOGRAFIA COMPLEMENTAR

BATTESE, G.; COELLI, T. Technology gap, efficiency and stochastic metafrontier function. **International Journal of Business and Economics**, v. 1, p. 87-93, 2002.

BATTESE, G.; COELLI, T.; O´DONNELL, C. A metafrontier production function for estimation of technical efficiencies and technology potentials for firms operating under different technologies. **Journal of Productivity Analysis**, v. 21, p. 91-103, 2004.

DE WITTE, K.; KORTELAINEN, M. **Blaming the exogenous environment?** Conditional efficiency estimation with continuous and discrete exogenous variables. Munich: MPRA, 2009. (MPRA Paper, n. 14034). Disponível em: https://goo.gl/z63cLb. Acesso em: 1º mar. 2019.

EPE – EMPRESA DE PESQUISA ENERGÉTICA. **Anuário Estatístico de Energia Elétrica 2014, ano-base 2013**. Rio de Janeiro: EPE, 2014. Disponível em: https://goo.gl/CFEA3K. Acesso em: 1º mar. 2019.

FÄRE, R. J.; GROSSKOPF, S.; KNOX LOVELL, C. A. **The measurement of efficiency of production**. Boston: Kluwer Academic Publishers, 1985.

FARSI, M.; FETZ, A.; FILLIPINI, M. **Benchmarking and regulation in the electricity distribution sector**. Zurich: Cepe, 2007. (Cepe Working Paper, n. 54).

GALÁN, J. E.; POLLITT, M. G. Inefficiency persistence and heterogeneity in Colombian electricity utilities. **Energy Economics**, v. 46, p. 31-44, 2014.

GIANNAKIS, D.; JAMASB, T.; POLLITT, M. Benchmarking and incentive regulation of quality service: an application to the UK electricity distribution networks. **Energy Policy**, v. 33, p. 2256-2271, 2005.

GREENE, W. The econometric approach to efficiency analysis. *In:* FRIED, H.; LOVELL, K.; SCHMIDT, S. (Eds.). **The measurement of efficiency**. Oxford: Oxford University Press, 2008.

HANEY, A. B.; POLLITT, M. **Efficiency analysis of energy networks-an international survey of regulators**. Cambridge: University of Cambridge, 2009.

HUANG, Y.; CHEN, K.; YANG, C. Cost efficiency and optimal scale of electricity distribution firms in Taiwan: an application of metafrontier analysis. **Energy Economics**, v. 32, p. 15-23, 2010.

JAMASB, T.; POLLITT, M. **Benchmarking and regulation of electricity transmission and distribution utilities**: lessons from international experience. Cambridge: University of Cambridge, 2000.

KOPSAKANGAS, M.; SVENTO, R. Estimations of cost-effectiveness of the Finnish electricity distribution utilities. **Energy Economics**, v. 30, p. 212-229, 2008.

LIN, B.; DU, K. Technology gap and China´s regional energy efficiency: a parametric metafrontier approach. **Energy Economics**, v. 40, p. 529-536, 2013.

PITT, M.; LEE, L. The measurement and sources of technical inefficiency in the Indonesian weaving industry. **Journal of Development Economics**, v. 9, p. 43-64, 1981.

RAO, D. S. P.; O´DONNELL, C.; BATTESE, G. **Metafrontier functions for the study of inter-regional productivity differences**. Brisbane: The University of Queensland, 2003. (Working Paper, n. 01/2003).

SCHMIDT P.; SICKLES, R. Production frontiers and panel data. **Journal of Business and Economic Statistics**, v. 2, p. 367-374, 1984.

EFEITOS DO PROGRAMA DE QUALIFICAÇÃO PROFISSIONAL BOLSA FUTURO NA EMPREGABILIDADE E SALÁRIO DOS TRABALHADORES

Guilherme Resende Oliveira[1]
Felipe Resende Oliveira[2]
Rafael Terra[3]
Ana Carolina Zoghbi[4]

Este estudo verifica os impactos econômicos de um programa de qualificação profissional do governo do estado de Goiás, Bolsa Futuro, por meio de métodos baseados em diferenças em diferenças e ponderação pelo *propensity score*. Os resultados das estimações sugerem que o Programa Bolsa Futuro aumenta a empregabilidade formal dos participantes entre 3,4 e 4,4 pontos percentuais. Por outro lado, as evidências apontam que a política não eleva a remuneração média dos seus egressos. O programa teve efeito similar ao curso de formação inicial e continuada do Programa Nacional de Acesso ao Ensino Técnico e Emprego (Pronatec).

Palavras-chave: avaliação de impacto; educação profissional; *propensity score*; diferenças em diferenças.

THE EFFECTS OF THE JOB TRAINING PROGRAM BOLSA FUTURO IN EMPLOYMENT AND WAGE OF WORKERS

This study evaluates the economic impacts of a training program ran by the state of Goiás, Bolsa Futuro, using methods based on differences-in-differences and propensity score weighting. The results of the estimates suggest that Bolsa Futuro increases participants' formal occupation between 3.4 and 4.4 percentage points. On the other hand, the evidence indicates that the policy does not raise the average salary. The program has similar effect as Pronatec's FIC course.

Keywords: impact evaluation; technical education; propensity score matching; difference-in-differences, double robust estimator.

JEL: C52; I21.

1 INTRODUÇÃO

Este trabalho analisa os efeitos do Programa Bolsa Futuro sobre a taxa de ocupação de seus participantes e sobre seus rendimentos. O Programa Bolsa Futuro foi criado pela Secretaria de Estado de Desenvolvimento Econômico, Científico e Tecnológico e de Agricultura, Pecuária e Irrigação (SED) de Goiás com o intuito de fornecer qualificação profissional para que os participantes tornem-se mais empregáveis e dependam menos de assistência social. O programa foca em trabalhadores de

1. Professor do Centro Universitário Alves Faria (Unialfa).
2. Professor da Universidade Federal de Mato Grosso (UFMT).
3. Professor da Universidade de Brasília (UnB).
4. Professora da UnB.

baixa renda[5] e realiza uma transferência condicional de R$ 75,00 por mês para diminuir a evasão.

Os cursos do Programa Bolsa Futuro assemelham-se aos de formação inicial e continuada (FIC) do Programa Nacional de Acesso ao Ensino Técnico e Emprego (Pronatec). As principais diferenças consistem na transferência condicional presente no primeiro e ausente no segundo, e na maior utilização de ensino a distância como ferramenta complementar de ensino pelo programa goiano. O orçamento do programa em 2014 – excluindo o gasto com a transferência condicional – foi de R$ 66.675.900 e atendeu 45.797 alunos, o que implicou um custo de R$ 1.455,90 por aluno. Considerando os gastos com a transferência condicional, esse valor sobe para R$ 1.936,28 (Goiás, 2014). Para efeito de comparação, o Pronatec custou R$ 8.010.536.000 e atendeu 4.617.420 alunos, o que resultou em um custo de R$ 1.734,85 por aluno entre 2011 e 2015 (Brasil, 2016b).

Este estudo usa dados dos participantes do Programa Bolsa Futuro combinados com dados da Relação Anual de Informações Sociais do Ministério do Trabalho e Emprego (Rais/MTE), e aplica métodos quase experimentais para verificar o impacto dos beneficiados no mercado de trabalho, mais especificamente sobre a renda e a empregabilidade.[6] Em especial, empregamos os estimadores de diferenças em diferenças, estimadores ponderados pelo inverso do *propensity score* (IPW) e o estimador duplamente robusto que combina diferenças em diferenças com essa ponderação e um ajuste por regressão (IPWRA), semelhante ao proposto por Heckman, Ichimura e Todd (1997).

Os resultados sugerem, em geral, efeitos positivos e significativos do Programa Bolsa Futuro sobre a empregabilidade quando o grupo de controle usado é composto por inscritos no programa e não atendidos por questões de oferta. Não há evidências de que o programa afete a remuneração média dos seus participantes. Por outro lado, quando consideramos alunos do Pronatec como grupo de controle, não encontramos nenhuma diferença entre a empregabilidade ou os salários dos dois grupos, o que sugere que ambos os programas têm efeitos similares.

A transferência condicional de R$ 75/mês por participante de baixa renda do Bolsa Futuro visa reduzir o custo privado de participação no programa – por exemplo, custo de oportunidade, com transporte, alimentação, material, cuidados com os filhos durante o período de curso etc. –, que influencia a evasão do aluno. De fato essa transferência aparentemente determina uma evasão mais baixa no programa goiano, igual a 9,6% até 2014, enquanto a do Pronatec era de 21% entre 2011 e 2014 (Montagner e Muller, 2015). Se considerarmos que os participantes de programas têm chances 3,5 pontos percentuais (p.p.) maiores

5. Participantes do Bolsa Família ou Renda Cidadã de Goiás.
6. Usamos também dados do Pronatec para construir um dos grupos de controle.

de se empregarem (ver tabela 3), e que estes ganham um salário mínimo a mais do que os não empregados, a vantagem dos tratados teria que durar cerca de oito anos para justificar os custos adicionais.[7] Como a vantagem proporcionada pelo programa é pequena, os custos da transferência condicional ficam mais difíceis de serem cobertos pelos benefícios. Assim, é possível concluir que uma revisão do Programa Bolsa Futuro seria desejável, revisando valores e beneficiários da transferência condicional ou buscando maior focalização naquele grupo que potencialmente seria mais beneficiado.

Há várias justificativas para avaliar políticas ativas voltadas ao mercado de trabalho. Ross e Sawhill (1975) e Sayer *et al.* (2011) encontram evidências de que a falta de um emprego aumenta as chances de divórcio. Também aumentam as chances de episódios de depressão (Farré, Fasani e Mueller, 2015; Dooley, Catalano e Wilson, 1994). Dee (2001) e Arkes (2007) associam o desemprego a um aumento no consumo de substâncias psicoativas. Outro problema refere-se ao maior engajamento em atividades criminosas entre os desempregados (Levitt e Lochner, 2001).

A variedade de serviços providos pelos programas de emprego e treinamento é substancial. Butler (1976) classifica esses serviços em: *i)* desenvolvimento de habilidades (*skill development*), em sala de aula ou *on-the-job*;[8] *ii)* experiência de trabalho (*work experience*); *iii)* desenvolvimento da empregabilidade (*employment development*), em que se enquadram a assistência na busca por emprego (*job search assistance*) e colocação (*job matching*); e *iv)* desenvolvimento de empregos (*job development*), em que o setor público cria empregos para aqueles que procuram colocação. Esses serviços têm a função de aumentar a eficiência no mercado de trabalho, pois trabalhadores podem ter dificuldades em buscar qualificação adequada, seja por falta de recursos, seja por mercado de crédito inexistente ou falta de informação (Lalonde, 2003, p. 555).

A informação sobre o mercado de trabalho é custosa para os ofertantes de mão de obra. Agências de emprego e treinamento, por sua vez, observam menores custos de reunir informações sobre o mercado de trabalho. A existência desses serviços justifica-se pela correção de falhas de mercado, e sua provisão pelo setor público justifica-se pelo fato de o treinamento ter características de bens públicos (Kuku *et al.*, 2015).

Por mais que a literatura que avalia programas de treinamento seja extensa,[9,10] esta ainda carece de validade externa, pois os efeitos encontrados são bastante heterogêneos.

7. Considerando uma taxa de juros nula para simplificar. O tempo seria ainda maior se considerarmos taxas de juros maiores.

8. O Bolsa Futuro busca desenvolver habilidades vocacionais em sala de aula; portanto, enquadra-se na categoria de *skill development*.

9. Ver Kiefer (1978; 1979), Long, Mallar e Thornton (1981), Bloom e McLaughlin (1982), Eberwein, Ham e Lalonde (1997), Card e Sullivan (1988), Couch (1992), Dickinson, Johnson e West (1986) e Heckman e Smith (1997).

10. Card, Kluve e Weber (2010), Lalonde (2003), Barnow e Smith (2015) e Heckman, Lalonde e Smith (1999) produzem *surveys* da literatura que permitem ter uma dimensão desta e da quantidade de avaliações já realizada.

Há resultados indicando aumentos na empregabilidade, assim como efeitos não significativos.[11] Há estudos indicando tanto aumentos salariais quanto efeitos não significativos ou até reduções;[12] há grande variação de resultados entre regiões,[13] adultos e jovens,[14] entre homens e mulheres e brancos e não brancos.[15] Por fim, os efeitos do programa também variam de acordo com a duração do treinamento.[16]

Há diversas dificuldades para avaliar os programas. Primeiramente, há um claro problema de autosseleção dos trabalhadores (Heckman e Smith, 1997). Em segundo lugar, a seleção de candidatos menos vulneráveis pode partir dos próprios agentes responsáveis pela escolha de candidatos se houver incentivos financeiros atrelados às suas taxas de sucesso (fenômeno denominado *cream skimming*).[17] A terceira dificuldade refere-se à queda verificada nos rendimentos no período pré-tratamento (*ashenfelters' dip*), que pode prejudicar a mensuração dos ganhos dos programas de treinamento e emprego devido à falta de contrafactuais adequados.[18] Outro problema surge do enclausuramento do aluno. Nesse período os alunos encontram menos emprego – pois estão frequentando o curso –, o que subestima o efeito do programa sobre rendimento e emprego nos períodos imediatamente subsequentes ao treinamento.[19] Há também efeitos confundidores (*counfounding effects*) entre serviços. Treinamento e serviços de colocação podem ser ofertados conjuntamente, inviabilizando a separação de um efeito e outro.[20] Ademais, alguns autores argumentam que não é o treinamento em si que aumenta a empregabilidade e a renda dos trabalhadores, e sim a informação sobre o mercado de trabalho reunida no programa e o *networking* realizado durante o curso.[21]

Há décadas, experimentos controlados randomizados têm sido apontados como as formas mais adequadas de estimar de maneira consistente os efeitos de políticas ativas voltadas ao mercado de trabalhos.[22] Entretanto, as dificuldades óbvias para a implementação de experimentos fez desenvolver paralelamente uma literatura que buscava estimadores quase experimentais que replicassem os resultados dos experimentos.[23]

11. Ver Bergemann, Fitzenberger e Speckesser (2009).
12. Ver Heckman, Lalonde e Smith (1999, p. 2061).
13. Ver Lalonde (1995, p. 159).
14. Para resultados de programas voltados para jovens, ver Schochet, Burghardt e Glazerman (2001) e Cave e Doolittle (1991).
15. Ver Friedlander, Greenberg e Robins (1997), Ashenfelter (1978) e Gay e Borus (1980).
16. Kluve *et al.* (2012) estimam um retorno positivo e decrescente da duração do programa sobre a empregabilidade dois anos após a entrada neste.
17. Ver Heckman, Smith e Taber (1996).
18. Ver Ashenfelter (1978) e Heckman e Smith (1999).
19. Ver Gerfin e Lechner (2002).
20. Ver Lalonde (1995, p. 159).
21. Ver Heckman, Lalonde e Smith (1999).
22. Ver Ashenfelter e Card (1985), Burtless e Orr (1986) e Burtless (1995).
23. Ver Lalonde (1986), Dehejia e Wahba (1999), Heckman, Ichimura e Todd (1998) e Smith e Todd (2005).

No Brasil, poucos programas de treinamento e emprego puderam ser avaliados, seja por falta de informações (ou a confidencialidade), seja pelo reduzido número de programas do tipo. Algumas avaliações quantitativas foram feitas sobre treinamento vocacional em geral,[24] sobre o Plano Nacional de Qualificação do Trabalhador (Planfor),[25] sobre o Plano Setorial de Qualificação (PSQ),[26] sobre o Pronatec[27] e sobre o ensino técnico em geral.[28] Tais avaliações baseiam-se em métodos quase experimentais. Somente Hermeto e Rios-Neto (2007) avaliam um programa de forma experimental.[29] Ao analisar o Programa Bolsa Futuro, este estudo busca contribuir com a literatura sobre programas de treinamento no Brasil, de modo a ampliar a validade externa dos efeitos já estimados.

Este artigo está dividido em cinco seções, incluindo esta introdução. A seção 2 apresenta o Programa Bolsa Futuro e sua inserção no contexto goiano. A seção 3 descreve a amostra e discute a estratégia empírica utilizada na análise dos dados. A seção 4 apresenta e discute os resultados obtidos. Por fim, a seção 5 traz as considerações finais do trabalho.

2 ASPECTOS INSTITUCIONAIS

Programas de treinamento no Brasil costumam ser denominados como "educação profissional continuada". O arcabouço institucional básico da educação profissional e tecnológica (EPT) foi estabelecido pela Lei nº 9.394/1996, mais conhecida como Lei de Diretrizes e Bases da Educação. Esta lei propõe que a EPT organize-se em eixos tecnológicos abrangendo os seguintes cursos: de formação inicial e continuada ou qualificação profissional; de educação profissional técnica de nível médio; e de educação profissional tecnológica de graduação e pós-graduação.

Neste sentido, duas das metas do Plano Nacional de Educação (PNE) referem-se diretamente à EPT (metas 10 e 11). A primeira visa expandir a oferta de matrículas de educação de jovens e adultos, nos ensinos fundamental e médio, na forma integrada à educação profissional; e a segunda busca triplicar as matrículas da EPT de nível médio, pelo menos 50% da expansão no segmento público. Logo, é claro o incentivo governamental para estimular a EPT no país.

24. Reis (2015), que encontra efeitos positivos desses programas sobre salário e empregabilidade.

25. Silveira Neto (2002) encontra efeitos positivos do programa sobre a renda e o emprego dos desocupados, enquanto Firpo, Fogel e Jales (2014) e Fernandes, Menezes-Filho e Zylberstajn (2002) não encontram efeitos significativos.

26. Ver Petterini (2011). O autor encontra efeitos positivos do programa sobre o emprego.

27. Ver O'Connell *et al.* (2017) e Barbosa Filho, Porto e Liberato (2015). Enquanto o primeiro trabalho encontra efeitos positivos do Pronatec-MDIC sobre a ocupação, o segundo não encontra evidências significativas do programa geral.

28. Em geral, os estudos encontram efeitos positivos do ensino técnico sobre salários e emprego. Ver Biondi e Felício (2015), Musse e Machado (2013), Severnini e Orellano (2010), Vasconcelos *et al.* (2010), Águas (2011) e Gontijo e Amaral (2015).

29. O programa em questão era Plano Estadual de Qualificação Profissional de Minas Gerais. Os resultados mostram que os tratados encontram emprego mais rapidamente.

Uma das modalidades da educação profissional destina-se aos concluintes dos ensinos fundamental e médio, além dos matriculados no último (realizando os cursos concomitantemente).

Atualmente, Goiás possui mais de 1,2 milhão de pessoas com ensino médio completo, além de quase meio milhão de concluintes do ensino fundamental e alunos que cursam essas etapas de ensino, totalizando quase 2 milhões de indivíduos aptos a inscreverem-se no ensino profissional (Pnad/IBGE, 2013). Cerca de 60% da população goiana (com 10 anos ou mais) apresentava escolaridade inferior ao ensino médio completo. O Censo/IBGE de 2010 mostrava uma taxa de alfabetização geral de aproximadamente 90%, com municípios apresentando valores entre 76% e 96%. O rendimento mensal do trabalho principal apontado pela Pesquisa Nacional por Amostra de Domicílios (Pnad) em 2013 era por volta de dois salários mínimos. Sessenta e cinco por cento dos trabalhadores com carteira assinada e 78% daqueles sem carteira assinada recebiam até dois salários. A pesquisa ainda mostra que aproximadamente 25% dos goianos ocupados maiores que 15 anos eram informais e 6% estavam desempregados.

Em Goiás, o marco legal da educação profissional é a Lei Complementar nº 26/1998, a qual estabeleceu as diretrizes e bases do Sistema Estadual de Educação Profissional, Científica e Tecnológica, que abrange a rede pública estadual e as instituições de educação profissional e tecnológica. A Lei Complementar nº 109/2014 introduziu algumas alterações na lei anterior e criou as unidades de EPT, na forma dos Institutos Tecnológicos do Estado de Goiás (Itegos).

As iniciativas de oferta da educação profissional no estado de Goiás partiram de outros agentes, além do governo estadual, como o Sistema S e o governo federal. A Rede Estadual de Educação Profissional cresceu com o Programa de Expansão da Educação Profissional (Proep), elaborado pelo Ministério da Educação (MEC) nos anos 1990 e conduzido pelas Secretarias Estaduais. Este separou o ensino médio do ensino profissional e lançou as bases do atual arcabouço da Rede de Educação Profissional do estado. Entre as principais instituições ofertantes do estado, destacam-se o Instituto Federal de Educação, Ciência e Tecnologia de Goiás (IFG) e o Instituto Federal Goiano. Do lado da iniciativa privada, o Sistema S é o principal ofertante em Goiás. Suas instituições estão vinculadas às confederações patronais e dedicam-se à formação de trabalhadores para os setores específicos aos quais estão ligados.

A demanda crescente por mão de obra no mercado de trabalho e o crescimento recente da economia goiana acima da média nacional[30] refletiram no aumento do número de matrículas no ensino profissional. De acordo com o Censo Escolar (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep), Goiás possuía aproximadamente 13 mil matrículas em 2008 e 24 mil matrículas em 2014,

30. Dados do IBGE mostram que o produto interno bruto (PIB) goiano cresceu, entre 2003 e 2013, a uma taxa média anual de 4,7%, superior à registrada para a economia brasileira, de 3,5%.

totalizando um terço do total da região Centro-Oeste. Em 2013, Goiás tinha 1.119 docentes ligados à educação profissional, totalizando aproximadamente dezoito alunos por docente. Entre as áreas com maior concentração de docentes estavam saúde e bem-estar social, educação, engenharia, produção e construção, respectivamente. Esses dados referem-se aos cursos técnicos, com carga horária superior a oitocentas horas e, portanto, não entram nas estatísticas do Programa Bolsa Futuro.

O Bolsa Futuro é um programa de treinamento criado pela Lei Estadual nº 17.406/2011 e regulamentado pelo Decreto nº 7.470/2012, com o objetivo de amenizar os problemas ligados à falta de capacitação da mão de obra em Goiás. É coordenado e monitorado pela SED. Seu principal objetivo é expandir a oferta de cursos de educação profissional de formação inicial e continuada, ampliar a estrutura da rede de educação técnica no estado, integrar e expandir a estrutura dos diferentes níveis de ensino e ampliar as oportunidades educacionais dos trabalhadores por meio da melhoria da formação e da qualificação profissional.

A primeira etapa do Programa Bolsa Futuro buscou a redução da pobreza e, portanto, destinou-se aos beneficiários de programas sociais (Programa Renda Cidadã, programa estadual de transferência de renda e Programa Bolsa Família) e às famílias de baixa renda.

Nessa etapa os alunos tinham de cursar obrigatoriamente o núcleo comum e um núcleo específico, de livre escolha. O núcleo comum proporcionava um nivelamento dos conhecimentos de português e matemática e dava uma orientação geral sobre comunicação, postura profissional, responsabilidade socioambiental e informática. Já o núcleo específico era direcionado para qualificar pessoas para várias ocupações no mercado de trabalho de diferentes áreas: comércio, agropecuária, funções de apoio, indústria e infraestrutura (ver cursos adiante).

O aluno deveria ter acima de 16 anos e escolaridade igual a cinco ou mais anos. Apesar da estratégia apresentada, o programa consistia em cursos de qualificação profissional em caráter virtual, isto é, de ensino a distância (EaD), com carga horária recomendada entre cento e sessenta e quatro e duzentos e sessenta e quatro horas. Assim, após a conclusão, os alunos recebiam um certificado de curso de curta duração, emitido pela secretaria responsável do governo do estado de Goiás, com carga horária equivalente ao total de horas do curso. Durante a vigência do curso, alguns alunos de baixa renda receberam mensalmente um incentivo financeiro de R$ 75,00, com o limite de sete parcelas, além da gratuidade do curso.[31]

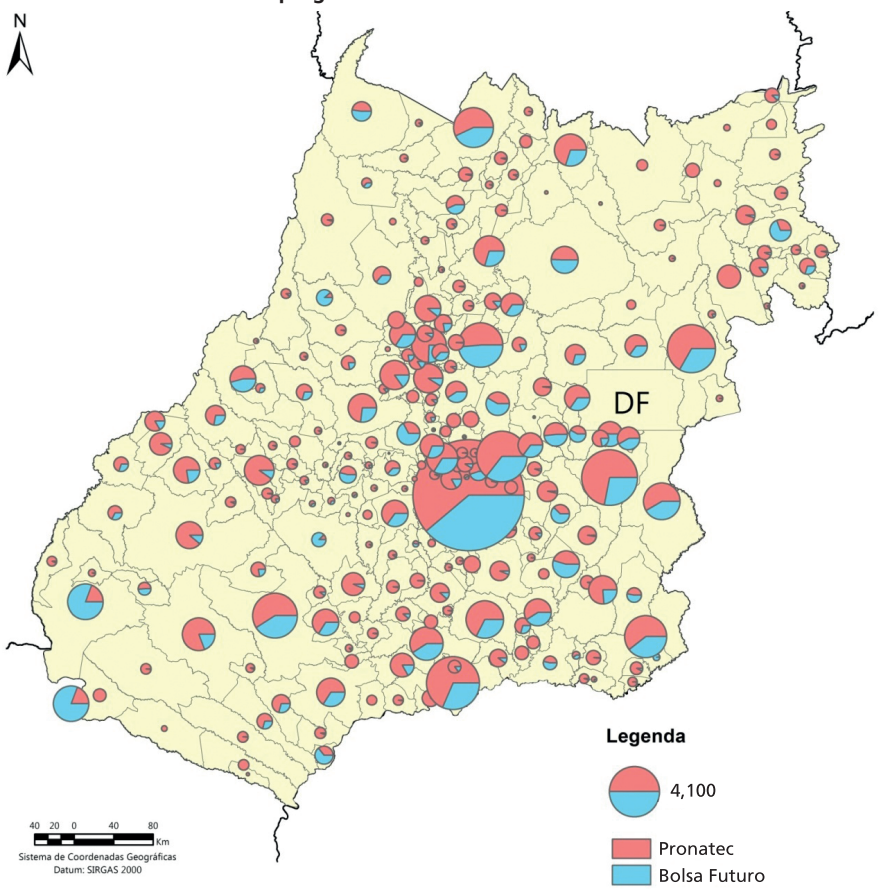
A segunda etapa foi regulamentada pelo Decreto nº 7.959, de agosto de 2013, e reuniu todos os programas de qualificação profissional do governo do estado.

31. Para manter a continuidade do recebimento desse incentivo, deveria ter frequência mensal mínima de 75%. Aos que obtiveram no curso média final igual ou superior a 8,0, existia uma bonificação equivalente a um mês do incentivo financeiro, havendo um incentivo para o aluno dedicar-se aos estudos.

Em linhas gerais, essa etapa manteve o mesmo projeto pedagógico da primeira e permitiu acesso à comunidade em geral, sendo ou não beneficiária dos programas Renda Cidadã e Bolsa Família. Logo, houve uma mudança na regra de seleção (critério de elegibilidade). Portanto, qualquer pessoa que cumprisse o requisito básico poderia se inscrever no *site* do programa e iniciar o curso.

Nessa etapa eram obrigatórias duas horas de aulas presenciais por semana, sendo dois cursos do núcleo comum e um do núcleo específico. Ao total, treze escolas certificadoras – representadas pelos Itegos – e quase cem colégios tecnológicos (Cotecs) estavam espalhados por todo o Estado e faziam parte da rede pública de educação profissional do estado de Goiás, representada no mapa 1. A rede oferecia os cursos e contribuía com o acesso a computadores, especialmente para os alunos que não possuíam o equipamento necessário para o curso.

MAPA 1
Alunos concluintes dos programas Bolsa Futuro e Pronatec entre 2013 e 2014



Fonte: Núcleo Bolsa Futuro/SED.
Elaboração dos autores.

Os cursos oferecidos até 2017 consistiam em: técnico de vendas, secretariado e rotinas administrativas, recepção de hotel e atendente de bar, reprodução animal e produtividade do gado bovino leiteiro, técnicas agrícolas, destilador de álcool, cuidador de idosos e crianças, porteiro e zelador, básico em eletricista e encanador e caldeireiro. Todos os cursos tiveram um alto percentual de conclusão, com destaque para informática, que era prerrequisito para os demais cursos, e, consequentemente, teve maior número de alunos e percentual de conclusão. A respeito da oferta de cursos Bolsa Futuro, apesar de grande parte do curso ser ministrada a distância (EaD), a demanda variou de região a região. O custo do programa em 2014 foi de R$ 1.455,9 por aluno. Considerando os gastos com a transferência condicional, esse valor sobe para R$ 1.936,28 (Goiás, 2014).

A oferta de ensino técnico e profissional em Goiás foi complementada pelo maior programa de educação profissional do Brasil, o Pronatec, criado pelo governo federal em 2011, e pelo programa estadual Bolsa Futuro. A oferta do Pronatec no estado de Goiás é feita pelos institutos federais (IFs), pela rede privada, pelo Sistema S e pelo governo do estado de Goiás, por meio das escolas técnicas e tecnológicas da Rede Estadual de Educação Profissional.

O Pronatec foi regulamentado pela Lei nº 12.513/2011 e tinha a meta de ofertar 8 milhões de vagas a estudantes e trabalhadores diversos em todo o Brasil. O Pronatec baseia-se em cinco iniciativas principais para aperfeiçoar o ensino profissional:

a) ampliação das unidades de ensino para educação profissional, científica e tecnológica;

b) ampliação da integração entre a educação profissional e o ensino médio nas redes estaduais;

c) criação e expansão de cursos técnicos e de qualificação profissional na modalidade a distância (e-Tec Brasil);

d) transferência das contribuições compulsórias recebidas pelo Serviço Nacional de Aprendizagem Industrial (Senai) e pelo Serviço Nacional de Aprendizagem Comercial (Senac) para a criação de vagas gratuitas em cursos técnicos para pessoas de baixa renda;

e) Bolsa Formação: são cursos oferecidos para estudantes e trabalhadores. Os cursos para estudantes têm duração de oitocentas horas, já os cursos para trabalhadores são de cento e sessenta horas. Essa modalidade de curso também é conhecida como cursos FIC (Brasil, 2016a).

Por meio do que veio a ser chamado de Bolsa Formação, a União financia a oferta de cursos presenciais de educação profissional e tecnológica, arcando com os custos relacionados à oferta educacional, ao transporte, à alimentação e até

aos materiais escolares utilizados pelos beneficiários – estratégia similar à bolsa oferecida pelo Bolsa Futuro, exceto pela transferência condicional em dinheiro de R$ 75,00/aluno/mês do último. Os cursos FIC tinham duração de cento e sessenta horas ou mais, e consistiam em 443 cursos distribuídos conforme características científicas e tecnológicas em onze eixos tecnológicos. Nesse caso, a emissão de certificados era de responsabilidade das próprias instituições de ensino ofertantes.

De acordo com dados do Sistema Nacional de Informações da Educação Profissional e Tecnológica (Sistec/MEC), a distribuição de alunos FIC do Pronatec no mapa alcançou ainda mais estudantes do que o Bolsa Futuro na maior parte dos municípios. Neste sentido, ambos programas tiveram grande abrangência espacial no território goiano. Mais do que expandir as vagas dos cursos profissionais, o intuito do Pronatec também era de que essas vagas chegassem ao interior do país, onde a oferta de cursos técnicos é menor, ou seja, democratizar o acesso. Um dos objetivos dos cursos FIC é alcançar trabalhadores em situação de vulnerabilidade social e os beneficiários do Programa Bolsa Família, a fim de criar uma "porta de saída" para o programa. Os dados do Sistec/MEC mostram que aproximadamente 50% das inscrições foram intermediadas pelo Ministério do Desenvolvimento Social (MDS) entre 2011 e 2014. Para o estado de Goiás, esse valor foi de aproximadamente 40%. Apesar da maior abrangência em termos de área do conhecimento, percebe-se grande sobreposição dos dois programas, que tiveram desenhos parecidos em termos de objetivos, público-alvo, distribuição espacial, estratégias de implementação, entre outros fatores. O custo médio do Pronatec foi de R$ 1.734,85 por aluno entre 2011 e 2015 (Brasil, 2016b).

3 DADOS E ESTRATÉGIA EMPÍRICA

A base de dados utilizada neste estudo reúne informações de cadastro do Programa Bolsa Futuro, do Sistec e do Tribunal de Contas da União (TCU), as quais continham as matrículas no Pronatec por Cadastro de Pessoa Física (CPF), curso, data de matrícula e conclusão, e localização da participação no polo da rede de educação. As informações referentes ao mercado de trabalho, como situação empregatícia, renda, setor econômico e outras informações (raça, gênero, idade e escolaridade), encontram-se na Rais/MTE, que reúne informações de todas as plantas instaladas no estado e vincula os empregadores e os empregados. Vale salientar que estes dados são administrativos e englobam apenas o mercado de trabalho formal. Os dados de cadastro são cruzados com a Rais por meio do CPF.

O Programa Bolsa Futuro atendeu 15.646 alunos em 2012, 23.850 alunos em 2013 e 45.797 em 2014. A análise neste trabalho concentra-se em participantes da segunda etapa do Bolsa Futuro no ano de 2014 – até dezembro desse ano.[32] Avaliamos o efeito do treinamento sobre salário e empregabilidade entre janeiro

32. Ver Decretos nº 7.470, de 20 de outubro de 2011, e nº 7.959, de 8 de agosto de 2013.

e dezembro de 2013 e de 2014. A restrição do foco na segunda fase do programa busca aproveitar o maior número de inscritos, tentando manter na amostra uma proporção suficiente de unidades de controle. Como os controles consistiam em inscritos no programa que não frequentaram o curso, que começaram a frequentar e não concluíram ou que só viriam a participar em anos posteriores, a melhor forma de avaliar o efeito do tratamento é focando um ano específico.

Outra razão para focar a segunda etapa da política é que essa é a forma definitiva do programa. A primeira etapa assemelhava-se mais a um programa-piloto, pois atingia menos pessoas e ainda se encontrava em fase de testes. Na primeira etapa o programa era totalmente a distância, mas tal experiência mostrou que alguma aula presencial era necessária para instruções mais básicas, especialmente para aqueles que não tinham o equipamento necessário ou a familiaridade com tal equipamento. A primeira etapa dava preferência para inscritos participantes do Bolsa Família e do Renda Cidadã, enquanto a segunda etapa flexibilizou a preferência por "pessoas com baixa renda" comprovada. Ademais, entre as duas etapas, a duração do benefício aumentou de seis para sete meses.

De acordo com dados do Sistec, concomitantemente ao Programa Bolsa Futuro foram oferecidos cursos FIC do Pronatec em Goiás a 14.066 alunos em 2012, 47.528 alunos em 2013 e 60.682 alunos em 2014. Excluímos do grupo de tratamento aqueles que participaram dos dois programas. Também construímos um grupo de controle com concluintes do Pronatec para comparação com os egressos do Bolsa Futuro.

Apesar de a participação no Programa Bolsa Futuro ser livre, houve uma tentativa de focalização em indivíduos de baixa renda. Esse desenho insere naturalmente um viés de seleção. Diferenças na empregabilidade ou no salário de participantes e não participantes podem ser decompostas em:

$$E[Y_i^T|T] - E[Y_i^C|C] = E[Y_i^T|T] - E[Y_i^C|T] + \underbrace{\{E[Y_i^C|T] - E[Y_i^C|C]\}}_{Viés}, \quad (1)$$

em que os sobrescritos indicam um estado possível; T indica o *status* de tratamento; e C o *status* de controle. $E[Y_i^T|T]$ e $E[Y_i^C|C]$ são o salário ou o *status* de emprego médio observado; e $E[Y_i^T|C]$ e $E[Y_i^C|T]$ são médias contrafactuais. Uma solução para obter uma estimativa consistente do efeito do tratamento sobre os tratados é condicionar em covariadas ou em um *propensity score* $(p(X))$ que satisfaça a hipótese de ignorabilidade do tratamento. Com isso, teríamos $E[Y_i^C|T, p(X)] = E[Y_i^C|C, p(X)]$, e o viés em (1) desapareceria.[33] No quadro 1 apresentamos variáveis observáveis potenciais determinantes da participação no Programa Bolsa Futuro presentes na Rais, e que permitem estimar o *propensity score*. No caso dos estimadores por

33. Ver Rosenbaum e Rubin (1983).

propensity score matching, o efeito do tratamento sobre os tratados ($ATTx$) é obtido pareando cada unidade tratada (no suporte comum) com uma média de um grupo de unidades de controle. O estimador pode ser representado por:

$$E[Y_i^T | T, p(X)] - E\left[Y_i^C \omega(i,j,h)_{KM} \middle| C, p(X)\right], \quad (2)$$

em que $\omega(i,j,h)_{KM}$ é o peso definido por uma função *kernel* para cada unidade de controle j, que declina quanto mais distante da unidade tratada i ele se encontra, e que pode ser representado por:

$$\omega(i,j,h)_{KM} = \frac{\kappa\left(\frac{p(X_j)-p(X_i)}{h}\right)}{\sum_{j\in h}\kappa\left(\frac{p(X_j)-p(X_i)}{h}\right)}, \quad (3)$$

em que $p(X_i)$ é a probabilidade de a unidade i participar do programa; $p(X_j)$ é a probabilidade de tratamento da unidade j; e h é o tamanho da banda – consideramos o padrão h = 0,06. Para que a soma dos pesos das unidades de controle dentro do intervalo h seja igual a 1, a expressão em (3) divide o valor da função *kernel* pela soma dos valores de cada controle dentro da mesma banda h. A função *kernel* é do tipo Epanechnikov.[34] Como o *propensity score* é estimado, o erro-padrão habitual será inconsistente. Assim, calculamos o erro-padrão dos métodos baseados nesse índice por *bootstrap*.

QUADRO 1
Variáveis dependentes e regressores

Variável	Descrição
Emprego	Indicador (*dummy*) de vínculo ativo em 31/12/2014.
Renda	Remuneração do trabalhador em dezembro de 2014 (preços de 2014).[1] Nos modelos econométricos usamos o logaritmo dessa variável.
Regressores	
Emprego D_0	Indicador (*dummy*) de vínculo ativo em 31/12/2013.
Renda D_0	Remuneração do trabalhador em dezembro de 2013 (preços de 2014). Nos modelos econométricos usamos o logaritmo dessa variável.
Tratamento (T_i)	Igual a 1 se concluiu algum curso do Bolsa Futuro, e igual a 0 caso contrário.
Idade	Idade do indivíduo, em anos. É uma *proxy* para experiência.
Escolaridade	*Dummies* de etapas de ensino: ensino fundamental (omitida), ensino médio, ensino superior e pós-graduação.
Homem	Igual a 1 caso masculino, e 0 caso contrário.
Negro	Igual a 1 se negro, e 0 caso contrário.
Setor	*Dummies* de setor de ocupação: agricultura, indústria e serviços, de acordo com a classe de atividade econômica, segundo Classificação Nacional de Atividades Econômicas (CNAE) versão 2.0.

Fonte: Rais/MTE e Sistec/MEC.
Elaboração dos autores.
Nota: [1] Atualizado pelo Índice Nacional de Preços ao Consumidor Amplo (IPCA/IBGE).

34. Dada por $\kappa(u) = \frac{3}{4}(1-u^2)1[|u| \leq 1]$, em que $u = \frac{p(X_j)-p(X_i)}{h}$.

Se as diferenças iniciais entre tratados e controles forem fixas e as tendências de crescimento dos resultados potenciais de tratados e controles forem iguais – $\hat{E}\left[Y_{i1}^{C} - Y_{i0}^{C} \middle| C\right] = \hat{E}[Y_{j1}^{C} - Y_{j0}^{C} | T]$–, o método de diferenças em diferenças pode ser empregado para estimar o efeito do tratamento de forma consistente. Denotando os resultados potencias no período 1 se tratados (controles) como $Y_{i1}^{T}(Y_{i1}^{C})$, e no período inicial como Y_{i0}^{T} (Y_{i0}^{C}) entre os tratados (controles), podemos calcular o efeito do tratamento como:

$$Y_{it} = \delta T_i + \rho D_t + \gamma T_i D_t + X_{it}\beta + \varepsilon_{it}, \quad (4)$$

em que Y_{it} representa o *status* de empregado formal (igual a 1 se empregado formal, e 0 caso contrário) ou o salário médio no trabalho formal do indivíduo, denotado pelo subscrito i. Podemos impor $\beta = 0$, ou não. O vetor X representa as características determinantes do emprego e do salário. O subscrito t denota o período de tempo. A variável D_t assume valor 0 no primeiro período e igual a 1 no segundo. T_i assume valor igual a 1 se o indivíduo pertence ao grupo de tratamento, e valor igual a 0 caso contrário.

No modelo de probabilidade linear, a hipótese de tendências paralelas independe das diferenças entre tratados e controles em termos de salário ou *status* de ocupação. Por outro lado, o uso de modelos Logit ou Probit em um desenho de diferenças em diferenças viola a hipótese de tendências paralelas entre tratados e controles quando há diferenças significantes entre estes.

De fato, a inclusão de covariadas contínuas em um modelo de probabilidade linear com variável dependente binária pode resultar predições fora do intervalo unitário. Mas isso é menos grave do que violar a hipótese de tendências paralelas do desenho de diferenças em diferenças. Além disso, considerando que neste trabalho interessa a causalidade do programa ao invés da previsão salarial ou de empregabilidade, o modelo de probabilidade linear é claramente mais adequado.

O efeito do tratamento pode ser expresso por:

$$\hat{\gamma} = \hat{E}[Y_{i1}^{T} - Y_{i0}^{T} | T] - \hat{E}\left[Y_{j1}^{C} - Y_{j0}^{C} \middle| C\right]. \quad (5)$$

É possível aplicar o método de diferenças em diferenças em uma amostra pareada. O benefício, nesse caso, seria tentar encontrar um grupo de controle mais comparável, que potencialmente teria mais chance de satisfazer a hipótese de tendências paralelas, característica do estimador de diferenças em diferenças.

Os indivíduos que compõem o *pool* de onde são selecionados os potenciais controles são todos aqueles que se matricularam no programa, mas não concluíram. Os erros-padrão para as estimativas que utilizam o *propensity score* são obtidos

por *bootstrap* a fim de incorporar a variância dos *propensity scores* estimados nos cálculos dos valores esperados.

Podemos também usar o inverso do *propensity score* (estimador IPW) como peso, o que permite rebalancear a amostra. Nesse caso, estimamos o efeito do tratamento sobre os tratados, cujo peso é $W(i,j) = T_i - \frac{(1-T_i)}{1-p(X_j)}$, e a expressão é dada por:

$$E\left[W(i,j)[Y_{i1}^T - Y_{i0}^T|T=1]\right] - E\left[W(i,j)\left[Y_{j1}^C - Y_{j0}^C|T=0\right]\right]. \quad (6)$$

Podemos ainda usar um estimador duplamente robusto baseado no ajuste por regressão e no inverso do *propensity score* como ponderação (IPWRA),[35] semelhante ao proposto por Robins, Rotnitzky e Zhao (1995) e Heckman, Ichimura e Todd (1997). Aplicando esse estimador a um desenho de diferenças em diferenças,[36] temos:

$$\begin{aligned} DID_{CAR} = {} & E\left[W(i,j)\left[(Y_{i1}^T - X_{i1}\hat{\beta}) - (Y_{i0}^T - X_{i0}\hat{\beta})|T=1\right]\right] \\ & -E\left[W(i,j)\left[(Y_{j1}^C - X_{j1}\hat{\beta}) - (Y_{j0}^C - X_{j0}\hat{\beta})|T=0\right]\right], \end{aligned} \quad (7)$$

em que $W(i,j)$ é o mesmo peso definido anteriormente; $E[Y_{i1}^T|T=1]$ é o resultado médio dos tratados após o tratamento; $E[X_{i1}\hat{\beta}|T=1]$ é o valor esperado dos tratados no período pós-tratamento, dadas as características X_{i1}, caso não tivessem sido tratados (o coeficiente $\hat{\beta}$ é estimado somente com dados do grupo de controle e extrapolado depois para os tratados). $E[Y_{i0}^T|T=1]$ é o resultado médio esperado para o grupo de tratados no período pré-tratamento, enquanto $E[X_{i0}\hat{\beta}|T=1]$ é o resultado esperado para o grupo de tratados, dadas as características X_{i0} no período pré-tratamento. $E[Y_{j1}^C|T=0]$ é o valor médio ponderado dos controles no período pós-tratamento; $E[X_{j1}\hat{\beta}|T=0]$ é o valor médio ponderado esperado desse mesmo grupo no período pós-tratamento, dadas as características X_{j1}. Finalmente, $E[Y_{j0}^C|T=0]$ é o valor médio ponderado do grupo de controle no período pré-tratamento e $E[X_{j0}\hat{\beta}|T=0]$ é o valor médio ponderado esperado do grupo de controle no período pré-tratamento, dadas as suas características X_{j0}.

O principal grupo de controle escolhido é composto apenas por indivíduos que se inscreveram no programa e não frequentaram o curso, que começaram a frequentar e não concluíram, ou que só viriam a participar em anos posteriores. Denominamos este como "grupo de controle 1". Com isso, buscamos minimizar

35. Esse estimador é robusto à especificação dos regressores que determinam a variável de resultado e à especificação do modelo que determina a autosseleção ao tratamento, daí a denominação de "duplamente robusto". Basta uma das especificações estar correta para que se produzam estimativas consistentes do efeito do tratamento.

36. Esse estimador pode ser usado também para uma amostra seccional somente para o ano pós-tratamento.

as diferenças iniciais e garantir que a hipótese de tendências paralelas seja mais plausível. É possível que alunos matriculados, mas não concluintes, possam ter derivado algum benefício antes mesmo da conclusão do curso – tais como o aumento da produtividade ou um aumento no número de entrevistas das quais participaram. Também é possível que, por terem despendido tempo atendendo ao programa de treinamento, o tempo destinado à busca por emprego tenha se reduzido. No primeiro caso, o viés seria positivo e, no último, seria negativo. A direção desses efeitos, portanto, não é clara, e faltam informações que nos permitam separá-los.

Na tabela 1 apresentamos as médias e as diferenças de covariadas pré-tratamento – referente ao ano de 2013 – entre tratados pelo Bolsa Futuro e três possíveis grupos de controle. O grupo de controle 1 é aquele descrito no parágrafo anterior. O grupo de controle 2 é composto por alunos que cursaram o Pronatec. Este grupo de comparação busca nos informar se o Bolsa Futuro foi mais efetivo do que o Pronatec na promoção do emprego e da renda. O grupo de controle 3 é composto por trabalhadores do estado de Goiás, sem nenhum filtro adicional. Este último grupo serve somente para fins ilustrativos do tamanho do potencial viés quando não restringimos o grupo aos indivíduos mais parecidos com os tratados. Como se pode notar, o grupo de controle 3 é muito diferente dos demais controles e do grupo de tratamento. O grupo de controle 1 – de inscritos desistentes e não concluintes do Bolsa Futuro – é bastante parecido com o grupo de participantes do Bolsa Futuro. Somente em algumas covariadas observamos diferenças estatisticamente significativas em relação aos tratados, e geralmente são pequenas. Por isso, concluímos que a seleção inicial de controles inscritos nos programas reduz drasticamente as diferenças observáveis, o que deve diminuir também as diferenças não observáveis. Ao final da tabela encontramos o número de observações tratadas e de controle do banco de dados usado para as regressões na seção seguinte.

TABELA 1
Diferenças pré-tratamento entre os grupos de tratamento e controle

	(1)	(2)	(3)	(4)			
	Bolsa Futuro	Controle 1 (inscritos)	Controle 2 (Pronatec)	Controle 3 (todos)	(1)-(2)	(1)-(3)	(1)-(4)
Emprego D0	0,696	0,678	0,684	0,73	0,019**	0,012**	-0,033***
	(0,003)	(0,009)	(0,004)	(0,000)	-	-	-
Renda D0	794,096	755,874	797,939	1403,237	38,223**	-3,843	-609,141***
	(6,392)	(16,094)	(8,231)	(1,558)	-	-	-
Negro	0,033	0,038	0,03	0,029	-0,005	0,003	0,003***
	(0,001)	(0,004)	(0,002)	(0,000)	-	-	-

(Continua)

(Continuação)

	(1)	(2)	(3)	(4)			
	Bolsa Futuro	Controle 1 (inscritos)	Controle 2 (Pronatec)	Controle 3 (todos)	(1)-(2)	(1)-(3)	(1)-(4)
Homem	0,367	0,376	0,406	0,558	-0,009	-0,039***	-0,191***
	(0,004)	(0,009)	(0,004)	(0,000)	-	-	-
Idade	30,927	30,198	31,324	33,488	0,729***	-0,397***	-2,56***
	(0,074)	(0,191)	(0,093)	(0,009)	-	-	-
Indústria	0,247	0,222	0,223	0,242	0,025***	0,024***	0,005
	(0,003)	(0,008)	(0,004)	(0,000)	-	-	-
Serviços	0,708	0,73	0,746	0,714	-0,023**	-0,038***	-0,006*
	(0,003)	(0,009)	(0,004)	(0,000)	-	-	-
EnsFund	0,357	0,345	0,356	0,318	0,012	0,001	0,039***
	(0,004)	(0,009)	(0,004)	(0,000)	-	-	-
EnsMed	0,572	0,577	0,57	0,542	-0,005	0,002	0,03***
	(0,004)	(0,01)	(0,004)	(0,000)	-	-	-
EnsSup	0,071	0,077	0,073	0,138	-0,006	-0,002	-0,067***
	(0,002)	(0,005)	(0,002)	(0,000)	-	-	-
PosGrad	0,000	0,001	0,001	0,002	0,000	0,000	-0,002***
	(0,000)	(0,001)	(0,000)	(0,000)	-	-	-
N	39.228	5.574	28.283	3.324.524	-	-	-

Fonte: Núcleo Bolsa Futuro/SED, MEC e Rais/MTE.

Obs.: 1. Dados referentes ao ano de 2013.

2. O grupo de tratamento é composto pelos indivíduos que concluíram o Programa Bolsa Futuro.
3. O grupo de controle 1 é composto somente pelos indivíduos inscritos no Bolsa Futuro e que não iniciaram ou abandonaram por alguma razão, ou que iriam se matricular em algum momento no futuro.
4. O grupo de controle 2 é composto por indivíduos que concluíram o Pronatec e não participaram do Bolsa Futuro.
5. O grupo de controle 3 é composto por todos os trabalhadores do estado de Goiás, e foi considerado somente para ilustrar o quão grande pode ser o viés quando não se escolhe um grupo de controle adequado.
6. Para definição das variáveis, ver o quadro 1.

Podemos concluir, pela tabela 1, que predominam entre os participantes de programas de treinamento aqueles com ensino médio – 57% dos alunos –, enquanto aqueles com ensino fundamental totalizavam 35%. Predominam mulheres – 63% –, trabalhadores do setor de serviços – pouco mais de 70% – e jovens – cerca de 30 anos. A taxa de ocupação entre os participantes em dezembro de 2013 era de cerca de 70%, e o salário médio era pouco maior que um salário mínimo – exceto no caso do grupo de controle 3, cujo salário médio era pouco menor que dois salários mínimos.[37]

37. Um salário mínimo em 2014 era igual a R$ 724,00.

4 RESULTADOS

A tabela 2 mostra os resultados do modelo de diferenças em diferenças com e sem variáveis de controle e apresenta efeitos positivos sobre a empregabilidade, mas não sobre a renda. Os resultados, quando usamos o grupo de controle 1, na coluna (1), sugerem que a empregabilidade dos tratados supera em 3,56 p.p. a dos trabalhadores do grupo de controle 1. Isso representa uma tendência 25% menos declinante do que a do grupo de controle entre os períodos inicial e final – igual a 14,3% negativos. Além disso, alguns gráficos e tabelas do apêndice ilustram as diferenças entre os grupos. Esse resultado difere daquele encontrado por Barbosa Filho, Porto e Liberato (2015) – que encontram efeitos negativos do Pronatec sobre a empregabilidade em Goiás – e aproxima-se mais daquele encontrado por O'Connell *et al.* (2017) – que encontra efeitos positivos do Pronatec MDIC sobre a empregabilidade. Mesmo inserindo controles de escolaridade e área de ocupação na coluna (2), o resultado se mantém. Os resultados do grupo de controle 2 não indicam diferenças estatisticamente significativas entre o Bolsa Futuro e o Pronatec.

Quando analisamos o efeito do programa sobre o logaritmo do salário nas colunas (4) a (6) da tabela 2, não encontramos efeitos estatisticamente significativos. Esse resultado corrobora o encontrado por Barbosa Filho, Porto e Liberato (2015), mas vai contra as estimativas positivas de programas de treinamento encontradas em grande parte da literatura.[38]

TABELA 2
Resultados por diferenças em diferenças

	Empregabilidade			Salário		
	(1) Controle 1 (inscritos)	(2) Controle 1 (inscritos)	(3) Controle 2 (Pronatec)	(4) Controle 1 (inscritos)	(5) Controle 1 (inscritos)	(6) Controle 2 (Pronatec)
Bolsa Futuro	0,0356*	0,0368*	0,133	0,134	0,139	0,383
	(0,0188)	(0,0188)	(0,1156)	(0,1240)	(0,1241)	(0,766)
Ano2014	-0,143***	-0,137***	-0,229**	-0,691***	-0,635***	-0,874
	(0,0181)	(0,0194)	(0,1154)	(0,1193)	(0,1280)	(0,767)
Indústria	-	-0,00439	0,000265	-	0,00253	-0,0102
		(0,0337)	(0,0265)		(0,253)	(0,204)
Serviços	-	0,0577*	0,0585*	-	0,250	0,252
		(0,0341)	(0,0317)		(0,2252)	(0,2117)
Idade	-	-0,00887	-0,00824	-	-0,0798*	-0,0697
		(0,0072)	(0,0072)		(0,0477)	(0,0480)

(Continua)

38. Ver Heckman, Lalonde e Smith (1999, p. 2061).

(Continuação)

	Empregabilidade			Salário		
	(1) Controle 1 (inscritos)	(2) Controle 1 (inscritos)	(3) Controle 2 (Pronatec)	(4) Controle 1 (inscritos)	(5) Controle 1 (inscritos)	(6) Controle 2 (Pronatec)
EnsMed	-	0,0606***	0,0630***	-	0,361***	0,448***
		(0,0146)	(0,0136)		(0,0967)	(0,0903)
EnsSup	-	0,0105	0,00430	-	0,691***	0,552***
		(0,0291)	(0,0268)		(0,1946)	(0,1803)
PosGrad	-	(0,173	-0,123	-	-1,208	-0,572
		(0,2337)	(0,2365)		(1,5487)	(1,5459)
Constante	0,753***	0,948***	0,924***	5,290***	7,282***	6,914***
	(0,0034)	(0,2230)	(0,2226)	(0,0230)	(1,4711)	(1,4742)
R^2	0,0355	0,0383	0,0339	0,0219	0,0241	0,0221
Obs.	34.109	34.109	39.601	34.109	34.109	39.601
Prob>F	0,00	0,00	0,00	0,00	0,00	0,00

Elaboração dos autores.
Obs.: 1. *, ** e *** = p<0,10, p<0,05 e p<0,01, respectivamente.
2. Erro-padrão entre parênteses.

Modelos com diversos tipos de ponderação na tabela 3 reforçam que o Bolsa Futuro aumenta a empregabilidade dos participantes em torno de 3,5 p.p., e ainda permite concluir que este programa não apresenta resultados diferentes do Pronatec – que deve ter tido comportamento similar. Na parte superior da tabela encontram-se os resultados para a variável *emprego*, e na parte inferior os resultados para a variável *salário*. Na primeira linha usamos o método de *propensity score matching*, utilizando a ponderação *kernel* Epanechnikov. Os resultados do modelo com covariadas pré-tratamento, que determinam a probabilidade de tratamento pelo método Probit, encontram-se na tabela A.1 do apêndice. Os resultados do balanceamento das covariadas antes e após o pareamento encontram-se na tabela A.2 e indicam que não há diferenças estatisticamente significantes entre tratados e controles após o pareamento. O percentual de ocupados no ano de 2014 é 3,97 p.p. maior entre os participantes do Bolsa Futuro comparativamente ao grupo de controle 1. Quando usamos uma amostra pareada com seus respectivos pesos em um desenho de diferenças em diferenças, o programa Bolsa Futuro apresenta um efeito de 4,47 p.p.

Pelo método IPW, que atribui peso maior para as unidades de controle quanto maior suas probabilidades de participarem do programa, o efeito do programa é estimado em 3,93 p.p. Usando a mesma ponderação em um desenho de diferenças em diferenças o efeito é próximo, igual a 3,78 p.p. Finalmente, se aplicarmos o estimador duplamente robusto IPWRA, o efeito estimado é de 3,78 p.p., e igual a 3,47 p.p. se aplicado a um desenho de diferenças em diferenças – talvez

nossa estimativa mais robusta. Tais resultados corroboram aqueles da tabela 2 e indicam um efeito consistente, apesar de modesto, de cerca de 3,5 p.p. sobre a empregabilidade.[39]

Os efeitos sobre o salário não se revelaram estatisticamente relevantes, exceto para o estimador PSM, que indica um aumento de 21,55% na renda entre os participantes do Bolsa Futuro comparativamente ao grupo de controle 1.

TABELA 3
Efeito do tratamento usando estimadores ponderados

Efeito do tratamento	Controle 1 (inscritos)	Controle 2 (Pronatec)
	Emprego	
PSM	0,0397**	-0,0103**
	(0,0159)	(0,0047)
Diferenças em diferenças pareado	0,0447**	0,0323***
	(0,0195)	(0,0067)
IPW	0,0393**	0,0523*
	(0,0155)	(0,0335)
Diferenças em diferenças IPW	0,0378*	0,0439
	(0,0193)	(0,0459)
IPWRA	0,0345**	0,0470
	(0,0159)	(0,0352)
Diferenças em diferenças IPWRA	0,0347*	0,0106
	(0,0193)	(0,0317)
	Log (salário)	
PSM	0,2155**	-0,0604
	(0,1065)	(0,0317)
Diferenças em diferenças pareado	0,1906	0,2801***
	(0,1296)	(0,0452)
IPW	0,1477	0,3598
	(0,0969)	(0,2278)
IPWRA	0,1698	0,0836
	(0,0943)	(0,3215)
Diferenças em diferenças IPWRA	0,1311	0,2858
	(0,0908)	(0,2474)

Elaboração dos autores.
Obs.: 1. *, ** e *** = p<0,10, p<0,05 e p<0,01, respectivamente.
2. Erros-padrão obtidos por *bootstrap* entre parênteses.
3. Os regressores usados no *propensity* score foram status de emprego inicial (emprego D0), renda no último emprego (lnrendadezr13), cor (negro), *dummies* de setor (indústria e serviços), idade e *dummies* de escolaridade (ensinos fundamental, médio e superior).
4. As variáveis de controle usadas nos modelos de diferenças em diferenças com dupla robustez foram as mesmas, exceto pelas variáveis que não variam no tempo, que foram retiradas nos modelos.

39. Não apresentamos as avaliações por cursos, pois o número de observações de controle é bastante pequeno em relação ao de tratados e, ao dividir por cursos, há diversos casos em que o número de controles fica excessivamente desbalanceado (muitas observações fora do suporte comum). Os resultados por cursos podem ser obtidos junto aos autores.

Como se pode notar na coluna (2) da tabela 3 – que utiliza o grupo de controle 2, composto por participantes do Pronatec –, exceto nos modelos de diferenças em diferenças pareado e IPW, que apresentaram uma diferença positiva em favor do Bolsa Futuro, e no modelo PSM, que apresenta um pequeno efeito negativo em favor do Pronatec, os demais modelos não sugerem diferenças estatísticas entre participantes dos dois programas. Considerando o modelo de diferenças em diferenças ajustado (modelo diferenças em diferenças IPWRA na tabela 3) como o mais robusto entre os estimados, não podemos dizer que há diferenças estatisticamente significativas entre participantes do Bolsa Futuro e do Pronatec em termos de empregabilidade. Da mesma forma, os modelos não indicam diferenças salariais entre os participantes dos dois programas.

Com esses resultados é possível concluir que o Programa Bolsa Futuro tem um pequeno efeito – cerca de 3,5 p.p. de aumento – na empregabilidade. Uma revisão do Programa Bolsa Futuro seria desejável para melhorar a custo-efetividade do programa, seja revendo o valor da parcela de transferência condicional de R$ 75/mês por participante de baixa renda, seja focalizando as áreas com maior potencial de resultados sobre empregabilidade e salário. Isto porque considerando que os participantes de programas têm cerca de 3,5 p.p. a mais de chances de se empregarem (tabela 3), e que estes ganham um salário mínimo a mais do que os não empregados, a vantagem dos tratados teria que durar no mínimo oito anos para o programa ser custo-efetivo. Com os dados disponíveis ainda não é possível estimar tais efeitos no longo prazo. Mas considerando que estes efeitos tendem a dissipar-se em pouco tempo,[40] é provável que o programa não esteja sendo custo-efetivo.

5 CONCLUSÃO

O desemprego relaciona-se diretamente com divórcio, depressão, consumo de substâncias psicoativas e criminalidade, entre outros desfortúnios. Nesse sentido, políticas ativas voltadas ao mercado de trabalho, como os programas de qualificação de mão de obra, buscam aumentar a eficiência deste, pois podem melhorar o bem-estar dos trabalhadores. Ademais, os trabalhadores podem ter dificuldades em buscar qualificação adequada, seja por falta de recursos, seja por mercado de crédito inexistente ou falta de informação.

A literatura que avalia programas de treinamento ainda carece de validade externa, pois os efeitos encontrados são bastante heterogêneos. Há evidências de que programas de treinamento aumentam, não alteram, ou até mesmo reduzem salários. Há grande variação de resultados entre regiões, adultos e jovens, entre homens e mulheres, brancos e não brancos.

40. Ver Schochet, Burghardt e McConnell (2008).

Este trabalho avalia o Programa Bolsa Futuro criado pelo estado de Goiás para promover o emprego e a renda entre os seus habitantes por meio de treinamento misto presencial e a distância. Os resultados foram obtidos por meio de uma série de técnicas econométricas adequadas para a avaliação de programas. Foram estimados efeitos do programa por diferenças em diferenças, *propensity score matching*, diferenças em diferenças com amostra pareada, regressão com ponderação pelo inverso do *propensity score* (IPW), e um estimador duplamente robusto (IPWRA), além de estimadores de diferenças em diferenças IPW e IPWRA.

Os resultados, em geral, sustentam um efeito positivo do programa sobre a empregabilidade, que varia entre 3,45 p.p. e 4,47 p.p. Não há efeitos significativos sobre os salários. Além disso, não há diferenças na empregabilidade ou no salário em comparação aos indivíduos participantes do Pronatec.

Os efeitos estimados sobre a empregabilidade parecem demasiadamente pequenos para justificar os custos do programa. Uma reformulação da estrutura de custos ou uma focalização no público-alvo mais beneficiado poderia melhorar o retorno do recurso público aplicado no Bolsa Futuro.

Algumas limitações do trabalho devem ser ressaltadas. Primeiramente, focamos a avaliação do mercado de trabalho formal. Muitas pessoas não registradas na Rais estão ocupadas no mercado informal, trabalham como pessoa jurídica, são empregadas domésticas ou são autônomas. Essa é a principal limitação, a qual atualmente não pode ser superada devido à falta de informações individuais sobre o mercado de trabalho e participação no programa de treinamento Bolsa Futuro. Em segundo lugar, analisamos os impactos no curto prazo. Essa limitação tende a desaparecer ao longo do tempo, conforme forem divulgadas informações mais atuais. Finalmente, a existência de poucas observações de controle adequadas para nossa análise prejudica uma avaliação mais detalhada por curso.

REFERÊNCIAS

ÁGUAS, M. **Ensino profissional e rendimentos do trabalho**: uma análise para o Brasil. Rio de Janeiro: Ipea, 2011. (Nota Técnica, n. 47).

ARKES, J. Does the economy affect teenage substance use? **Health Economics**, v. 16, p. 19-36, 2007.

ASHENFELTER, O. Estimating the effect of training programs on earnings. **The Review of Economics and Statistics**, v. 60, n. 1, p. 47-57, 1978.

BARBOSA FILHO, F. H.; PORTO, R.; LIBERATO, D. Pronatec Bolsa-Formação: uma avaliação inicial sobre reinserção no mercado de trabalho formal. *In:* ENCONTRO NACIONAL DE ECONOMIA, 43., 2015, Florianópolis. **Anais** [...]. Florianópolis: Anpec, 2015.

BARNOW, B. S.; SMITH, J. **Employment and training programs**. New York: NBER, 2015. (NBER Working Paper, n. 21659).

BERGEMANN, A.; FITZENBERGER, B.; SPECKESSER, S. Evaluating the dynamic employment effects of training programs in east Germany using conditional difference-in-differences. **Journal of Applied Econometrics**, n. 24, p. 797-823, 2009.

BIONDI, R. L.; FELÍCIO, F. Formação técnica e o mercado de trabalho: a melhoria do matching entre cursos ofertados e as vagas no mercado de trabalho a partir dos diferenciais de salários das ocupações de nível médio. *In:* ENCONTRO NACIONAL DE ECONOMIA, 43., 2015, Florianópolis. **Anais** [...]. Florianópolis: Anpec, 2015.

BLOOM, H.; MCLAUGHIN, M. **CETA training programs**: do they work for adults? Washington: CBO-NCEP, 1982. (Joint Report).

BRASIL. Ministério da Educação. **Guia Pronatec de Cursos FIC**. Brasília: MEC, 2016a.

_____. Tribunal de Contas da União. **Relatório de auditoria de natureza operacional com enfoque em resultado da Iniciativa Bolsa-Formação do Programa Nacional de Acesso ao Ensino Técnico e Emprego – Pronatec**. Brasília: TCU, 2016b.

BURTLESS, G. The case for randomized field trials in economic and policy research. **Journal of Economic Perspectives**, v. 9, n. 2, p. 63-84, 1995.

BURTLESS, G.; ORR, L. Are classical experiments needed for manpower policy? **Journal of Human Resources**, v. 21, n. 4, p. 606-639, 1986.

BUTLER, W. **Employment and training programs**. Washington: Congressional Budget Office, 1976. (Staff Working Paper).

CARD, D.; KLUVE, J.; WEBER, A. Active labour market policy evaluations: a meta-analysis. **Economic Journal**, v. 120, p. 452-477, 2010.

CARD, D.; SULLIVAN, D. Measuring the effect of subsidized training programs on movements in and out of employment. **Econometrica**, v. 56, n. 3, p. 497-530, 1988.

CAVE, G.; DOOLITTLE, F. **Assessing JOBSTART**: interim impacts of a program for school dropouts. New York: Manpower Demonstration Research Corporation, 1991.

COUCH, K. New evidence on the long-term effects of employment and training programs. **Journal of Labor Economics**, v. 10, n. 4, p. 380-388, 1992.

DEE, T. S. Alcohol abuse and economic conditions: evidence from repeated cross-sections of individual-level data. **Health Economics**, v. 10, n. 3, p. 257-270, 2001.

DEHEJIA, R. H.; WAHBA, S. Causal effects in nonexperimental studies: reevaluating the evaluation of training programs. **Journal of the American Statistical Association**, v. 94, n. 448, p. 1053-1062, 1999.

DICKINSON, K.; JOHNSON, T.; WEST, R. An analysis of the impact of CETA on participants' earnings. **Journal of Human Resources**, v. 21, p. 64-91, 1986.

DOOLEY, D.; CATALANO, R.; WILSON, G. Depression and unemployment: panel findings from the Epidemiologic Catchment Area study. **American Journal of Community Psychology**, v. 22, n. 6, p. 745-765, 1994.

EBERWEIN, C.; HAM, J.; LALONDE, R. The impact of classroom training on the employment histories of disadvantaged women: evidence from experimental data. **Review of Economic Studies**, v. 64, n. 4, p. 655-682, 1997.

FARRÉ, L.; FASANI, F.; MUELLER, H. **Feeling useless**: the effect of unemployment on mental health in the Great Recession. Barcelona: GSE, 2015. (Working Paper, n. 838).

FERNANDES, R.; MENEZES-FILHO, N. A.; ZYLBERSTAJN, H. **Avaliando o Planfor**: o Programa do Sindicato dos Metalúrgicos de São Paulo. São Paulo: USP, 2002.

FIRPO, S.; FOGEL, M.; JALES, H. **Evaluating the impact of stratified randomized experiments with an application to a Brazilian Public Training Program**. [S.l.]: [s.n.], 2014. No prelo.

FRIEDLANDER, D.; GREENBERG, D. H.; ROBINS, P. K. Evaluating government training. Programs for the Economically Disadvantaged. **Journal of Economic Literature**, v. 35, n. 4, p. 1809-1855, 1997.

GAY, R.; BORUS, M. Validating performance indicators for employment and training programs. **Journal of Human Resources**, v. 15, p. 29-48, 1980.

GERFIN, M.; LECHNER, M. A microeconometric evaluation of the active labour market policy in Switzerland. **The Economic Journal**, v. 112, p. 854-893, 2002.

GOIÁS. Secretaria de Estado de Gestão e Planejamento. **Estatística de atendimento de cursos de 2003 a 2015 desenvolvidos pelo Itego**. Goiânia: Segplan, 2014.

GONTIJO, B. A.; AMARAL, E. F. Associação do ensino profissionalizante com rendimento e emprego: Minas Gerais (2009 e 2011). **Planejamento e Políticas Públicas**, n. 44, p. 113-143, jan./jun. 2015.

HECKMAN, J.; ICHIMURA, H.; TODD, P. Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. **The Review of Economic Studies**, v. 64, p. 605-654, 1997.

_____; _____; _____. Matching as an econometric evaluation estimator. **The Review of Economic Studies**, v. 65, p. 261-294, 1998.

HECKMAN, J.; LALONDE, R.; SMITH, J. The economics and econometrics of active labor market programs. *In:* ASHENFELTER, O.; CARD, D. (Eds.) **Handbook of Labor Economics**. Amsterdam: North-Holland, 1999. v. 3A.

HECKMAN, J.; SMITH, J. **The determinants of participation in a social program**: evidence from the Job Training Partnership Act. Chicago: University of Chicago, 1997.

_____; _____. A. The pre-program earnings dip and the determinants of participation in a social program: implications for simple program evaluation strategies. **The Economic Journal**, v. 109, p. 313-348, 1999.

HECKMAN, J.; SMITH, J.; TABER, C. What do bureaucrats do? The effects of performance standards and bureaucratic preferences on acceptance into the JTPA Program. *In:* LIBECAP, G. (Ed). **Reinventing government and the problem of bureaucracy**. Greenwich: JAI Press, 1996. (Advances in the Study of Entrepreneurship, Innovation, and Economic Growth).

HERMETO, A. M.; RIOS-NETO, E. L. Uma avaliação experimental dos impactos da política de qualificação profissional no Brasil. **Revista Brasileira de Economia**, v. 61, n. 3, p. 353-378, 2007.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Pesquisa Nacional por Amostra de Domicílios**. Rio de Janeiro: IBGE, 2013.

KIEFER, N. Federally subsidized occupational training and the employment and earnings of male trainees. **Journal of Econometric**, v. 8, p. 111-115, 1978.

_____. Population heterogeneity and inference from panel data on the effects of vocational training and education. **Journal of Political Economy**, v. 81, p. 213-226, 1979.

KLUVE, J. *et al.* Evaluating continuous training programmes by using the generalized propensity score. **Journal of the Royal Statistical Society**, series A, v. 175, n. 2, p. 587-617, 2012.

KUKU, O. *et al.* **Training funds and the incidence of training**: the case of Mauritius. Bonn: IZA, 2015. (IZA Discussion Paper Series, n. 8775).

LALONDE, R. J. Evaluating the econometric evaluations of training programs with experimental data. **The American Economic Review**, v. 76, n. 4, p. 604-620, 1986.

_____. The promise of public sector-sponsored training programs. **The Journal of Economic Perspectives**, v. 9, n. 2, p. 149-168, 1995.

_____. Employment and training programs. *In:* MOFFITT, R. (Ed.). **Means-tested transfer programs in the United States**. Chicago: University of Chicago Press, 2003.

LEVITT, S. D.; LOCHNER, L. The determinants of juvenile crime. *In:* GRUBER, J. (Ed.). **Risky behavior among youths**: an economic analysis. Chicago: University of Chicago Press, 2001.

LONG, D.; MALLAR, C.; THORNTON, C. Evaluating the benefits and costs of the job corps. **Journal of Policy Analysis and Management**. v. 1, n. 1, p. 55-76, 1981.

MONTAGNER, P.; MULLER, L. H. (Orgs.). **Inclusão produtiva urbana**: o que fez o Pronatec/Bolsa Formação entre 2011 e 2014. Brasília: MDS, 2015. (Cadernos de Estudos Desenvolvimento Social em Debate, n. 24).

MUSSE, I.; MACHADO, A. F. Perfil dos indivíduos que cursam educação profissional no Brasil. **Economia e Sociedade**, Campinas, v. 22, n. 1, p. 237-262, 2013.

O'CONNELL, S. D. *et al.* **Can business input improve the effectiveness of worker training?** Evidence from Brazil's Pronatec-MDIC. Washington: World Bank, 2017. (Policy Research Working Paper, n. 8155).

PETTERINI, F. C. Uma avaliação de impacto e retorno econômico do Plano Setorial de Qualificação (Planseq). *In:* ENCONTRO ECONOMIA DO CEARÁ EM DEBATE, 7., 2011, Fortaleza. **Anais** [...]. Fortaleza: Ipece, 2011.

REIS, M. Vocational training and labor market outcomes in Brazil. **The B.E. Journal of Economic Analysis and Policy**, v. 15, n. 1, p. 377-405, 2015.

ROBINS, J. M.; ROTNITZKY, A.; ZHAO, L. P. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. **Journal of the American Statistical Association**, v. 90, n. 429, p. 106-121, 1995.

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

ROSS, H. L.; SAWHILL, I. V. **Time of transition**: the growth of families headed by women. Washington: The Urban Institute, 1975.

SAYER, L. C. *et al.* She left, he left: how employment and satisfaction affect men's and women's decisions to leave marriages. **American Journal of Sociology**, v. 116, n. 6, p. 1982-2018, 2011.

SCHOCHET, P., BURGHARDT, J.; GLAZERMAN, S. **National job corps study**: the impacts of job corps on participants' employment and related outcomes. Princeton: Mathematica Policy Research, 2001.

SCHOCHET, P.; BURGHARDT, J.; MCCONNELL, S. Does job corps work? Impact findings from the national job corps study. **American Economic Review**, v. 98, n. 5, p. 1864-1886, 2008.

SEVERNINI, E. R.; ORELLANO, V. I. O efeito do ensino profissionalizante sobre a probabilidade de inserção no mercado de trabalho e sobre a renda no período pré-Planfor. **EconomiA**, v. 11, p. 155-174, 2010.

SILVEIRA NETO, R. Eficácia e viés de seleção em programas de qualificação em trabalhadores em situação economicamente desvantajosa: evidências para o estado de Pernambuco. **Revista Econômica do Nordeste**, v. 4, n. 1, p. 1-22, 2002.

SMITH, J. A.; TODD, P. E. Does matching overcome LaLonde's critique of nonexperimental estimators? **Journal of Econometrics**, v. 125, n. 1-2, p. 305-353, 2005.

VASCONCELOS, L. *et al.* **Avaliação econômica do ensino médio profissional**. São Paulo: Fundação Itaú Social, 2010. (Relatório de Avaliação Econômica).

BIBLIOGRAFIA COMPLEMENTAR

ASHENFELTER, O.; CARD, D. Using the longitudinal structure of earnings to estimate the effect of training programs. **Review of Economics and Statistics**, v. 67, n. 3, p. 648-660, 1985.

BRASIL. Ministério do Trabalho e Emprego. **Relação Anual de Informações Sociais**. Brasília: MTE, 2011-2014.

CASSIOLATO, M. M.; GARCIA, R. C. **Pronatec**: múltiplos arranjos e ações para ampliar o acesso à educação profissional. Rio de Janeiro: Ipea, 2014. (Texto para Discussão, n. 1919).

MINCER, J. **Schooling, experience, and earnings**. New York: NBER, 1974. (Human Behavior and Social Institutions, n. 2).

NERI, M. **A educação profissional e você no mercado de trabalho**. Rio de Janeiro: FGV/CPS, 2010.

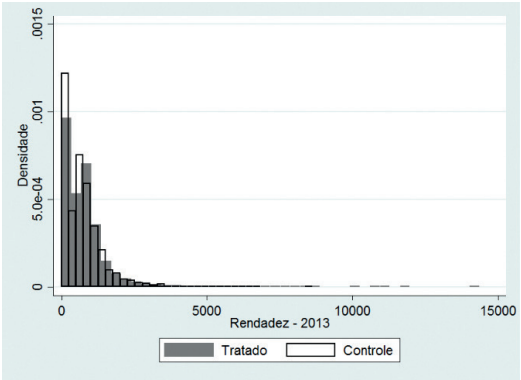
SMITH, J. A.; HECKMAN, J. J. The pre-programme earnings dip and the determinants of participation in a social programme. **Economic Journal**, v. 109, n. 457, p. 313-348, 1999.

WOOLDRIDGE, J. Inverse probability weighted estimation for general missing data problems. **Journal of Econometrics**, v. 141, p. 1281-1301, 2007.

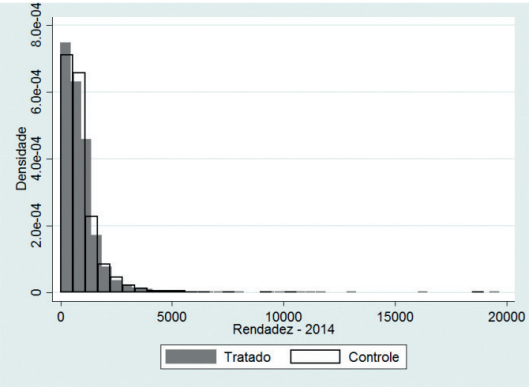
APÊNDICE

GRÁFICO A.1
Histogramas do salário (em 31/12) – por grupo e ano

A.1A – 2013



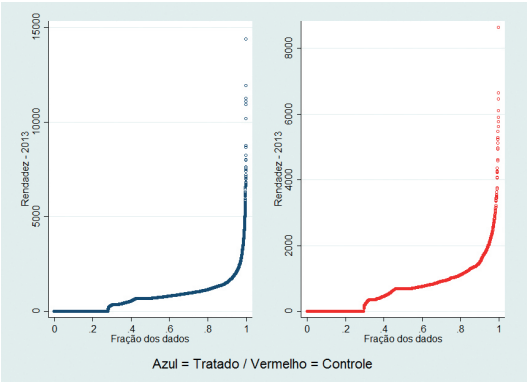
A.1B – 2014



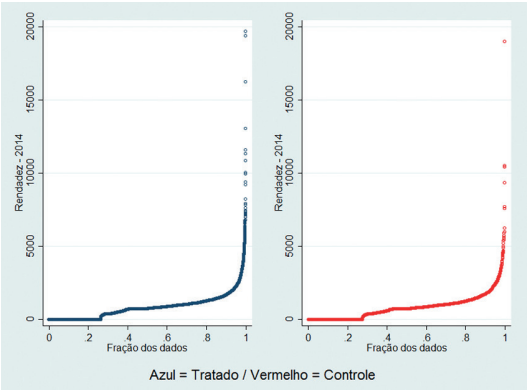
Elaboração dos autores.
Obs.: Gráficos cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

GRÁFICO A.2
Observações da renda (31/12) ordenadas em quantis – por grupo e ano

A.1A – 2013



A.1B – 2014



Elaboração dos autores.
Obs.: 1. Os pontos com renda zero são referentes às observações de indivíduos não empregados.
2. Gráficos cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

TABELA A.1
Regressões Probit usadas para estimar o *propensity score*

	Controle 1 (inscritos)	Controle 2 (Pronatec)
Emprego D0	0,002	-0,039
	(0,058)	(0,024)
Lnrendadezr13	0,006	0,001
	(0,009)	(0,004)
Negro	-0,187**	0,068*
	(0,074)	(0,038)
Indústria	0,024	-0,139***
	(0,075)	(0,036)
Serviços	-0,021	-0,117***
	(0,071)	(0,035)
Idade	0,002	-0,007***
	(0,001)	(0,001)
EnsFudC	0,684	0,507*
	(0,467)	(0,289)
EnsMed	0,647	0,463
	(0,466)	(0,289)
EnsSup	0,513	0,470
	(0,469)	(0,290)
Constante	0,678	0,048
	(0,474)	(0,292)

Elaboração dos autores.

TABELA A.2
Grupos de tratamento e controle, antes e após pareamento (2013)

Variável	Não pareado (N)	Balanceamento 1		Balanceamento 2	
	Pareado (P)	Tratados (Bolsa Futuro)	Controle 1 (inscritos)	Tratados (Bolsa Futuro)	Controle 2 (Pronatec)
Emprego	N	0,69644	0,72956	0,6968	0,7137
	P	0,69644	0,70819	0,6968	0,7072
Lnrendadez	N	4,9072	5,3421	5,046	5,1552
	P	4,9072	4,987	5,046	5,1065
Homem	N	0,36658	0,55772	0,3686	0,4169
	P	0,36658	0,35199	0,3686	0,3748
Negro	N	0,03274	0,02941	0,0309	0,0275
	P	0,03274	0,03247	0,0309	0,0273
Indústria	N	0,24672	0,24169	0,2373	0,2514
	P	0,24672	0,2306	0,2373	0,2402
Serviços	N	0,7077	0,71364	0,7233	0,7153
	P	0,7077	0,72814	0,7233	0,7264
Idade	N	30,927	33,488	30,820	31,899
	P	30,927	31,022	30,820	31,093
EnsFudc	N	0,35707	0,31783	0,3195	0,3046
	P	0,35707	0,35674	0,3195	0,3063
EnsMed	N	0,57155	0,54157	0,5922	0,6039
	P	0,57155	0,56165	0,5922	0,602
EnsSup	N	0,071	0,1385	0,0878	0,0906
	P	0,071	0,08122	0,0878	0,0914
PosGrad	N	0,00038	0,0021	0,0003	0,0007
	P	0,00038	0,00038	0,0003	0,0001

Elaboração dos autores.
Obs.: 1. As linhas referentes à amostra não pareadas (N) apresentam as médias dos grupos de tratados e controles antes da realização do pareamento por *kernel matching*.
2. As linhas referentes à amostra pareada (P) apresentam as médias dos grupos de tratados e de controles após o pareamento e, portanto, usam pesos definidos pelo procedimento de *kernel matching*. Além disso, a amostra pareada exclui as observações fora do suporte comum.

UMA ABORDAGEM DE DECOMPOSIÇÃO QUANTÍLICA PARA DIFERENCIAIS REGIONAIS DE SALÁRIOS NO BRASIL METROPOLITANO

Matheus Menezes dos Santos[1]
Ana Maria Hermeto[2]

Este trabalho busca avançar na compreensão dos diferenciais regionais de salário no Brasil metropolitano. Isso é pretendido por meio da aplicação de uma metodologia de regressões quantílicas desenvolvida por Koenker e Basset (1978) para estimar uma curva de diferenciais e, a partir disso, entender como o hiato regional se comporta em diferentes pontos da distribuição dos salários. É também usada a metodologia de decomposição quantílica detalhada desenvolvida por Fortin, Lemieux e Firpo (2009) para encontrar os determinantes dos movimentos ao longo dessa curva. Como principal resultado, atesta-se que o hiato regional é mais forte entre os trabalhadores do topo da distribuição, e esse fenômeno é agravado por diferenças na composição dos trabalhadores em diferentes regiões.

Palavras-chave: diferenciais regionais de salário; desigualdade regional; regressão quantílica; decomposição quantílica.

A QUANTILE DECOMPOSITION APPROACH FOR REGIONAL WAGE DIFFERENTIALS IN METROPOLITAN BRAZIL

This study intends a better understanding of regional wage gap in Brazil. To accomplish that, a methodology of quantile regression first presented by Koenker and Basset (1978) is used to estimates a regional gap curve for a better image of regional gap behavior along the yield curve. Then, a methodology of quantile detailed decomposition developed by Fortin, Lemieux and Firpo (2009) is used to decompose the determinants of this behavior. As a result, the regional gap seems to be more effective for those workers in the top of the distribution, and this effect is positively affected by regional differences in labor force's composition.

Keywords: regional wage gap; regional inequality; quantile regression; quantile decomposition.

JEL: C21; J31; R23.

1 INTRODUÇÃO

A questão da desigualdade regional no Brasil sempre ocupou um papel protagonista na discussão econômica. A partir do século passado o tema ganhou relevância, com a formulação e a aplicação de diversas políticas públicas de desenvolvimento regional. Entre todos os diversos prismas pelos quais se pode encarar esse problema, o estudo da desigualdade regional por meio do mercado do trabalho vem ganhando espaço na literatura econômica nas últimas décadas. Os resultados encontrados por

1. Mestre em demografia pelo Centro de Desenvolvimento e Planejamento Regional (Cedeplar) da Universidade Federal de Minas Gerais (UFMG).
2. Professora associada do Departamento de Economia do Cedeplar da UFMG.

Barros e Mendonça (1995) já validavam essa abordagem, mostrando que o mercado de trabalho é realmente um fomentador da desigualdade de renda interpessoal, e essa análise pode ser estendida para o problema da desigualdade entre indivíduos de regiões diferentes.

Pessôa (2001) levanta uma importante questão: as disparidades regionais de salário são um problema social ou regional? Não é uma pergunta banal, uma vez que traz grandes implicações para políticas públicas. O autor em nenhum momento nega que diferenças nas estruturas produtivas das regiões têm implicações sobre os salários dessas regiões, apenas questiona se questões sociais, tal como a diferença entre os níveis educacionais, não seriam mais relevantes para explicar o diferencial regional. Porém, neste caso, a linha que divide o problema social do regional é um tanto quanto nebulosa. Por exemplo, Duarte, Ferreira e Salvato (2003) estimam curvas de densidades de salário para comparar os salários das regiões Sudeste e Nordeste, e dos estados de São Paulo e do Ceará, concluindo que a educação é o principal determinante dessa desigualdade regional de rendimentos. À primeira vista, pode-se concluir que uma política social é o mais recomendável para este problema, objetivando equalizar o nível educacional dos trabalhadores dessas regiões. Mas e se diferenças nas estruturas produtivas dessas regiões tenderem a segmentar os mercados de trabalhos entre atividades que exigem mais qualificação em uma região e atividades que exigem menos qualificação em outra região? Este não consistiria em um problema passível de ser tratado por meio de uma política de desenvolvimento regional?

Em uma tentativa de separar melhor os aspectos sociais e regionais da desigualdade salarial, muitos autores utilizaram-se dos métodos de decomposição de Oaxaca-Blinder e suas extensões. De natureza metodológica relativamente simples, esses modelos de decomposição mostraram-se uma ferramenta poderosa para examinar essa questão, ao permitir o quanto de um diferencial se deve a diferenças na composição dos atributos educacionais e o quanto se deve a diferenças no retorno sobre esses atributos. Ou seja, o chamado efeito-preço captado pelo modelo mostra como a região remunera os atributos de seus trabalhadores, o que está diretamente relacionado às características de sua estrutura produtiva. De fato, diversos trabalhos para o Brasil que usam essa metodologia ou variantes próximos dela (Reis e Barros, 1989; Savedoff, 1990; Freguglia, Menezes-Filho e Souza, 2007; entre outros) concluem que o efeito-preço é muito maior do que o efeito-composição para explicar disparidades regionais de trabalho.

Uma vez aceito que o efeito-preço é o maior componente dos diferenciais regionais de salário, a questão passa a ser sobre seus determinantes. Afinal, quais as características de uma região que a levam a remunerar melhor os atributos de um indivíduo em relação à outra?

É largamente aceito que o prêmio salarial urbano está diretamente correlacionado com a densidade econômica de uma região. A intuição é que economias geradas pela aglomeração produtiva elevariam a produtividade dos trabalhadores e, consequentemente, os seus salários, por meio de um processo de transbordamento (*spillover*). Simões e Freitas (2014) encontram evidências dessa teoria para o Brasil, Wheaton e Lewis (2002) para os Estados Unidos, Whalley e Xing (2010) para a China e Bernard *et al*. (2003) para o Reino Unido. Ainda para o Brasil, Menezes, Fernandez e Dedecca (2005) utilizam-se da decomposição de Oaxaca-Blinder para encontrar os determinantes da diferença de salários entre as regiões metropolitanas de São Paulo e Salvador, e mostram que o peso da diferença entre estruturas produtivas é maior que o peso da diferença entre atributos pessoais. Fontes, Simões e Oliveira (2010) utilizam-se da construção de um modelo hierárquico em dois níveis para determinar qual nível responde pela maior parte da variância dos salários, as características do indivíduo ou da região. Como conclusão, este trabalho mostra que grande parte dos diferenciais regionais de salário no Brasil deve-se a economias de aglomeração.

A partir disso, surge outra questão adicional: o efeito da localização do trabalhador sobre seu salário é constante ao longo da curva de distribuição de rendimentos? Os atributos de uma região impactam da mesma forma um trabalhador que ganha pouco e um trabalhador que ganha muito? E uma questão até anterior a esta: como o diferencial regional de salários varia ao longo da curva de distribuição? Quem compõe um grupo mais homogêneo, os trabalhadores mais qualificados ou os menos qualificados? Quem carrega o maior peso da desigualdade regional? Galego e Pereira (2013) frisam a importância de se estudar a desigualdade regional ao longo da distribuição, pois a estimação pela média pode distorcer o que estiver acontecendo nos extremos da distribuição. Diversos outros trabalhos sobre desigualdade apontam para essa recomendação, como Figueiredo e Silva (2012), que usam de decomposição quantílica para analisar a questão da desigualdade de oportunidades no Brasil, e Bartalotti e Leme (2007), que utilizam dessa metodologia para estudar a questão da discriminação por gênero e cor. É natural, então, se pensar que, embora haja poucos trabalhos para o Brasil sobre como a composição da população por gênero e raça impacta o efeito regional nos rendimentos, características não produtivas dos trabalhadores afetarão os diferenciais regionais de forma diferente para diferentes pontos da curva de distribuição dos salários. É uma hipótese que remete a um dos pontos principais levantados por Galego e Pereira (2013): o papel das covariadas na determinação do efeito regional ao longo da distribuição. Por isso, um dos grandes avanços pretendidos por este trabalho é conseguir estabelecer essas relações para o mercado de trabalho brasileiro.

Essas são questões sobre o hiato regional de salário ao longo da curva e seus determinantes que este trabalho pretende ajudar a responder. Primeiro, estimando regressões quantílicas para comparar o efeito da localização sobre os salários dos trabalhadores ao longo da curva de distribuição nas dez regiões metropolitanas originais do Brasil, e depois utilizando a metodologia desenvolvida por Fortin, Lemieux e Firpo (2009) para decompor esses resultados entre efeito-composição e efeito-preço de cada covariada.

Além desta seção introdutória, há mais quatro: a segunda faz uma breve revisão da bibliografia sobre diferenciais regionais de salário ao longo da curva de rendimentos e apresenta a metodologia sobre o tema; a terceira introduz e examina a base de dados utilizada; a quarta apresenta os principais resultados encontrados; e a última conclui e sugere implicações para políticas públicas.

2 UMA ABORDAGEM QUANTÍLICA DOS DIFERENCIAIS REGIONAIS DE RENDIMENTOS

Desde que Koenker e Basset (1978) desenvolveram um método específico para a regressão nos quantis, a chamada regressão quantílica tem sido utilizada para entender melhor o comportamento da variável de uma regressão ao longo de uma curva de distribuição. Diversos estudos já provaram o valor desse método de estimação ao ser usado em regressões mincerianas. A hipótese central que justifica o uso de regressões quantílicas é que a estimação de coeficientes pela média pode esconder características importantes do comportamento da variável-resposta. Para uma exposição completa do método, ver Koenker e Basset (1978; 2001) e Buchinsky (1998a; 1998b; 2001).

A aplicação original das regressões quantílicas tem sido a estimação dos retornos educacionais ao longo da curva de distribuição dos salários, mantendo-se assim dentro do arcabouço da teoria do capital humano. Importantes trabalhos usando regressões quantílicas têm sido produzidos no Brasil e no mundo para investigar diferenciais por gênero e por cor no mercado de trabalho do ponto de vista de diferentes retornos educacionais, tal qual Buchinsky (2001), Martins e Pereira (2004), Maciel, Campêlo e Raposo (2001), Bartalotti e Leme (2007) e Coelho, Veszteg e Soares (2010). Todos esses exercícios têm como resultado a demonstração de que a estimação da equação salarial pela média esconde informações preciosas sobre o efeito do gênero e da cor nos salários. Portanto, analisar os diferenciais regionais de salário sob a ótica da regressão quantílica mostra-se um exercício interessante para compreender melhor como esses diferenciais se comportam.

Há que se pontuar que o efeito das externalidades do capital humano sobre o prêmio salarial urbano pode não ocorrer de forma homogênea entre trabalhadores de diferentes pontos da curva de rendimentos. Se assumirmos a hipótese de que os retornos serão maiores para trabalhadores mais qualificados, os trabalhadores

menos qualificados (e com menores rendimentos) tendem a ganhar menos com as externalidades do capital humano. Logo, a disparidade salarial entre trabalhadores de duas regiões tende a ser menor entre os mais pobres do que entre os mais ricos. Autor, Katz e Kearney (2005) encontram evidências desse fenômeno para os Estados Unidos, chamando a atenção para um processo de polarização do mercado de trabalho americano entre ricos e pobres. Goos e Manning (2007) descrevem o mesmo fenômeno para o mercado de trabalho no Reino Unido. Matano e Naticchioni (2013) realizam um exercício interessante para o mercado de trabalho italiano. Aplicando a metodologia da regressão quantílica para os dados de um painel com migrantes de áreas economicamente menos densas para áreas mais densas, eles mostram que os salários dos migrantes subiram de forma desigual. Entre os menos qualificados, o salário subiu lentamente, indicando um processo de aprendizado informal. Já entre os mais qualificados, embora também tenham ocorrido ganhos salariais ao longo do tempo, o maior efeito foi o imediato, o que os autores atribuem ao efeito do *matching*.

A partir desses resultados, o caminho natural é tentar entendê-los melhor. Para tanto, era necessária uma metodologia que expusesse os componentes dos movimentos dos diferenciais ao longo da curva de rendimento. O método de decomposição de diferenciais por meio de uma análise contrafactual foi proposto originalmente por Oaxaca (1973) e Blinder (1973), levando-o a ser denominado de método de Oaxaca-Blinder. O método Oaxaca-Blinder usa, originalmente, um modelo de mínimos quadrados para decompor o diferencial da média dos salários por meio de uma análise contrafactual. Dado que w_a é o salário médio esperado na região a e w_b é o salário médio esperado na região b, com x_a e β_a e x_b e β_b como o vetor de variáveis explicativas e coeficientes para cada uma das regiões, respectivamente, tem-se as seguintes equações salariais:

$$w_a = x_a\beta_a. \tag{1}$$

$$w_b = x_b\beta_b. \tag{2}$$

O diferencial entre o salário dos dois grupos pode ser descrito na forma de:

$$w_a - w_b = (x_a - x_b)\beta_a + (\beta_a - \beta_b)x_b. \tag{3}$$

Assim, o primeiro termo da equação é interpretado como o diferencial devido à composição, e o segundo termo como o diferencial de retornos das variáveis explicativas para cada grupo. Como essa literatura concentrou-se nos diferenciais de rendimentos entre homens e mulheres ou brancos e negros, muitas vezes esse segundo termo é chamado de discriminação. Ao longo dos anos, o método Oaxaca-Blinder foi sendo aperfeiçoado, com controles por viés de seletividade

(Neuman e Oaxaca, 2004) e extensões para análise do efeito temporal (Juhn, Murphy e Pierce, 1993), por exemplo. A extensão do método tradicional de decomposição da média para regressões quantílicas era ineficaz em quantis incondicionais, tornando necessário o uso de métodos alternativos.

Porém, o método tradicional encontra um empecilho ao estimar o contrafactual incondicional a partir da distribuição dos grupos *a* e *b*:

$$Quant_{\theta}^{c} = x_{a,i}{}'\beta_{b,\theta}, \text{ pois } Quant_{\theta}(w) \neq E_{x}[Quant_{\theta}(w|x)], \tag{4}$$

ou seja, a esperança do quantil θ não produz a distribuição marginal de θ. Isso inviabiliza a aplicação direta do método de Oaxaca-Blinder por ferir a Lei das Expectativas Iteradas. A partir deste problema, surgiram alguns métodos para tentar contorná-lo. Fortin, Lemieux e Firpo (2011) analisam criticamente alguns deles, tais quais Juhn, Murphy e Pierce (1993) e DiNardo, Fortin e Lemieux (1996). Um método que se popularizou nos últimos anos é o desenvolvido por Machado e Mata (2005), com aplicação especial para o caso da regressão quantílica. Para contornar o problema da quebra da Lei das Expectativas Iteradas, o método estima toda a distribuição da variável dependente para todos os quantis, para então realizar o exercício contrafactual, ao substituir a distribuição condicional de um grupo pelo outro. Melly (2006) desenvolve um método de decomposição quantílica muito semelhante, diferenciando-se na forma como são feitas as simulações para se estimar o contrafactual. Contudo, essa abordagem não permite a realização de uma decomposição detalhada, pois não computa o efeito individual de cada covariada na distribuição incondicional (Fortin, Lemieux e Firpo, 2011).

O método de decomposição para além da média que mais se aproxima dos objetivos da metodologia original de Oaxaca-Blinder é o desenvolvido por Fortin, Lemieux e Firpo (2009), por permitir uma decomposição detalhada. O método baseia-se na abordagem da função de influência recentralizada (*recentered influece function*, RIF desse ponto em diante). Abordagem desenvolvida por Hampel (1974), a função de influência (*influence function,* IF) de uma determinada estatística $\upsilon(F_Y)$ é definida em Fortin, Lemieux e Firpo (2009) como a representação da influência de uma observação individual sobre aquela distribuição. Os autores chamam a atenção para o fato de o método ser aplicável para diversas estatísticas, como a variância e o índice de Gini. O que o método está propondo com essa abordagem é a linearização da distribuição marginal das variáveis, tornando possível atender à Lei das Expectativas Iteradas e, enfim, realizar a decomposição tal qual a metodologia de Oaxaca-Blinder. Ressalta-se que, justamente por se tratar de uma linearização da estatística original, a eficácia da técnica está diretamente relacionada à qualidade desta aproximação linear. A decomposição é desenhada como:

$$Q_{a,\theta}(Y) - Q_{b,\theta}(Y) = (X_a - X_b)\gamma_{a,\theta} + \left(\gamma_{a,\theta} - \gamma_{b,\theta}\right)X_b, \quad (5)$$

onde o primeiro termo representa o efeito-composição (ou quantidade) e o segundo representa o efeito-estrutura (ou preço). Ver Fortin, Lemieux e Firpo (2009; 2011) para um melhor detalhamento do método.

Há ainda alguns pontos importantes sobre o método que merecem especial atenção, como é apontado em Fortin, Lemieux e Firpo (2011). O primeiro é o pressuposto da ignorabilidade, para o qual a distribuição do erro condicional a X ($\varepsilon|X$) deve ser igual para os dois grupos da decomposição. Outro pressuposto importante para este método é o da sobreposição, segundo o qual deve haver um suporte comum para os dois grupos, ou seja, em ambos devem haver indivíduos com características semelhantes. Se esses dois pressupostos não forem respeitados, a interpretação do efeito-estrutura detalhado, ou seja, o retorno de cada covariada, torna-se enviesado, sendo recomendável a interpretação apenas por meio da decomposição agregada. Se não há garantias de que esses dois pressupostos estão sendo satisfeitos, o efeito-preço será viesado, pois estará captando também o efeito das características não observáveis.

3 BASE DE DADOS E ESTRATÉGIA DE ESTIMAÇÃO

Nesta seção, será discutida a estratégia utilizada para adaptar a metodologia escolhida à pergunta a ser respondida, bem como apresentar a base de dados escolhida.

Para responder à questão central do trabalho, é preciso estimar uma equação salarial que capte a variação marginal nos salários de uma região em comparação à outra, que pode ser interpretada aqui como o grupo de controle. Nessa analogia, o impacto que se quer medir no grupo de tratamento seria a localização dos indivíduos em determinada região. Assim, o modelo para se estimar as regressões quantílicas usado aqui é o mesmo comumente usado na estimação por mínimos quadrados:

$$lw = \beta_0 + \beta_1 X + \beta_2 W + \beta_3 Z + \beta_4 R + \mu. \quad (6)$$

Para essa equação, tem-se *lw* como o logaritmo do salário-hora, *X* como um vetor de características pessoais (sexo, cor, idade), *W* como um vetor com *dummies* para cada grau de escolaridade concluído pelo indivíduo, *Z* um vetor com características do posto de trabalho (se faz parte do setor formal, setor de atividade e categoria da ocupação) e *R* um vetor de *dummies* regionais. O coeficiente β_4, variável para cada quantil, fornecerá a resposta para a pergunta.

A base de dados utilizada para este trabalho é a Pesquisa Nacional por Amostra de Domicílios (Pnad) de 2013, realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Além de atender às necessidades básicas deste

trabalho, como dimensão geográfica desejada e variáveis de controle, a escolha dessa base de dados justifica-se por ser a mais recente disponível (IBGE, 2013).

A escolha de estudar apenas o Brasil metropolitano se dá para facilitar a comparação dos resultados, uma vez que se trata de um mercado de trabalho mais homogêneo. Se o estudo se estendesse a todo o Brasil urbano, por exemplo, seria necessária a inclusão de novas variáveis que captassem o tamanho da cidade, pois o mercado de trabalho em uma cidade de 500 mil habitantes tem característica bem distintas do mercado de trabalho em uma cidade de 50 mil habitantes. Contudo, esse tipo de análise, sobre o efeito de atributos urbanos sobre os salários, extrapola os objetivos deste trabalho. Além disso, o fato de haver regiões metropolitanas de todas as regiões do país já torna o estudo abrangente o bastante para fazer afirmações sobre a desigualdade regional no Brasil.

Os algoritmos para a decomposição de diferenciais, bem como os métodos teóricos de decomposição e análise contrafactual, foram desenvolvidos para a comparação de dois grupos bem distintos, como homens e mulheres ou brancos e negros, por exemplo. Para adaptar esses algoritmos ao problema deste trabalho, as bases de dados foram separadas por regiões metropolitanas. Para facilitar a comparação dos resultados, foi definida uma só região metropolitana (São Paulo) como base ou grupo de controle. Deste modo, o que os resultados apresentarão são os diferenciais salariais de cada região metropolitana em comparação a São Paulo. Logo, há nove bases de dados, cada uma agregando a região metropolitana de São Paulo e a região metropolitana que se deseja analisar (Rio de Janeiro, Brasília, Belo Horizonte, Porto Alegre, Curitiba, Belém, Fortaleza, Recife ou Salvador). Assim, para cada banco de dados foi possível usar uma *dummy* que define o grupo de tratamento (a região metropolitana analisada) e o grupo de controle (a região metropolitana de São Paulo).

A respeito da escolaridade dos trabalhadores, optou-se pelo uso de *dummies* construídas pela própria Pnad por maior grau concluído em vez da variável contínua de anos de estudo. O controle por *dummies* de grau concluído mostrou ter um melhor ajuste do que a variável contínua, pois capta melhor as características dos retornos educacionais no Brasil, onde o grau concluído tem muito mais impacto do que o ano adicional de estudo. Deste modo, tem-se níveis de maior escolaridade atingida: fundamental incompleto, médio incompleto, superior incompleto e superior completo, sendo a primeira a escolhida como variável omitida.

Como controle por características do posto de trabalho, foi usada a classificação do IBGE para setores de atividades (Classificação Nacional de Atividades Econômicas – CNAE) e grupos de ocupações (Classificação Brasileira de Ocupações-Domiciliar – CBO-Domiciliar), conforme descrito na tabela 1.

O setor industrial foi tomado como variável omitida para setor de atividade e o trabalhador da produção foi tomado como variável omitida para grupo ocupacional nas regressões. Optou-se por usar controles também para participação no setor público e/ou formal da economia. Além de serem características com já demonstrado impacto nos salários, elas podem ajudar a explicar diferenças regionais de rendimentos.

Fortin, Lemieux e Firpo (2011) ressaltam o problema que possa ser gerado a partir da escolha da variável omitida. Para cada categoria que se omite, os resultados das outras mudarão drasticamente, como o esperado. E para nenhum dos casos tratados por esse trabalho há estudos que sugerem que determinada variável seria uma melhor escolha como omitida, fazendo essa escolha ser meramente arbitrária. Embora esse fato não afete a qualidade dos resultados, ele deve ser levado em consideração na hora de avaliar os resultados. É preciso ter sempre em mente qual o grupo que está sendo comparado. Para este trabalho especificamente, esse problema é relativizado por não pretender analisar nenhuma das covariadas isoladamente, mas sim grupos de covariadas: características pessoais (gênero, raça e idade), escolaridade e características do posto de trabalho.

A base de dados foi reduzida apenas para trabalhadores ocupados na semana de referência, com idade entre 25 e 59 anos, salários não nulos, em atividades não rurais e residentes em uma das dez regiões metropolitanas brasileiras pela classificação original do IBGE. A categoria de cor foi dicotomizada entre brancos e negros (negros e pardos). A população amarela e indígena foi excluída da amostra por se tratar de grupos étnicos com características muito específicas, o que poderia enviesar os resultados caso fossem alocados em uma das duas categorias. A tabela 1 traz um resumo da base de dados.

Há algumas características interessantes sobre a tabela 1 que merecem destaque. Por exemplo, como a composição de cor dos indivíduos varia conforme a região metropolitana. Nas regiões metropolitanas da região Sul do Brasil, a proporção de brancos é muito maior do que nas regiões metropolitanas das regiões Norte e Nordeste. Em termos de gênero e idade, não parece haver grandes discrepâncias entre as regiões.

Quanto à escolaridade, é interessante notar que o que difere as regiões mais ricas das menos ricas é a proporção de trabalhadores nas categorias extremas. A proporção de trabalhadores com ensino médio ou superior incompleto é similar em todas as regiões, estando a maior parte das variações nas categorias de fundamental incompleto e superior completo. É um fator que pode ajudar a explicar resultados dos diferenciais regionais de salário, principalmente no caso do Distrito Federal, onde a composição educacional difere mais fortemente da média das regiões metropolitanas.

TABELA 1
Resumo da base de dados

	Distrito Federal	RM de Belém	RM de Belo Horizonte	RM de Curitiba	RM de Fortaleza	RM de Porto Alegre	RM do Recife	RM do Rio de Janeiro	RM de Salvador	RM de São Paulo	**Total**
Homem	0,52	0,56	0,53	0,55	0,55	0,53	0,55	0,55	0,52	0,54	**0,54**
Branco	0,45	0,24	0,38	0,76	0,27	0,82	0,33	0,48	0,16	0,59	**0,50**
Idade	39,27	40,03	39,74	39,75	39,44	40,60	39,78	40,74	39,65	39,98	**40,04**
Fundamental incompleto	0,19	0,24	0,27	0,20	0,32	0,24	0,25	0,21	0,25	0,20	**0,22**
Médio incompleto	0,11	0,18	0,16	0,17	0,18	0,17	0,13	0,15	0,13	0,15	**0,15**
Superior incompleto	0,38	0,45	0,38	0,39	0,38	0,39	0,47	0,42	0,46	0,42	**0,41**
Superior completo	0,32	0,13	0,20	0,23	0,12	0,19	0,14	0,22	0,16	0,24	**0,21**
Outras indústrias	0,01	0,01	0,02	0,00	0,00	0,01	0,00	0,01	0,01	0,00	**0,01**
Indústrias de transformação	0,04	0,07	0,13	0,17	0,18	0,17	0,09	0,09	0,08	0,17	**0,13**
Construção civil	0,07	0,12	0,11	0,10	0,10	0,08	0,11	0,10	0,11	0,08	**0,09**
Comércio	0,16	0,21	0,18	0,18	0,20	0,17	0,19	0,18	0,18	0,16	**0,17**
Serviços	0,09	0,15	0,12	0,14	0,11	0,11	0,14	0,14	0,14	0,13	**0,13**
Administração pública	0,16	0,08	0,06	0,05	0,05	0,06	0,06	0,07	0,06	0,03	**0,06**
Serviços sociais	0,27	0,26	0,25	0,22	0,24	0,26	0,28	0,28	0,28	0,26	**0,26**
Outras atividades	0,20	0,10	0,14	0,13	0,11	0,15	0,13	0,14	0,15	0,17	**0,15**
Dirigentes	0,08	0,03	0,07	0,09	0,04	0,07	0,03	0,06	0,06	0,08	**0,07**
Ciências e artes	0,18	0,09	0,13	0,13	0,09	0,12	0,10	0,14	0,10	0,14	**0,13**
Técnicos	0,11	0,08	0,09	0,09	0,07	0,10	0,08	0,09	0,09	0,09	**0,09**
Trabalhadores administrativos	0,13	0,11	0,11	0,12	0,09	0,13	0,13	0,12	0,11	0,12	**0,12**
Trabalhadores de serviços	0,31	0,42	0,32	0,28	0,36	0,30	0,39	0,34	0,39	0,32	**0,33**
Trabalhadores de produção	0,15	0,25	0,26	0,26	0,31	0,25	0,24	0,22	0,23	0,25	**0,24**
Outras ocupações	0,04	0,03	0,02	0,03	0,03	0,02	0,02	0,03	0,03	0,01	**0,02**
Setor público	0,25	0,14	0,12	0,12	0,12	0,14	0,13	0,13	0,11	0,10	**0,12**
Setor formal	0,78	0,59	0,79	0,79	0,61	0,82	0,71	0,72	0,69	0,77	**0,75**
Total de observações	**3.508**	**2.944**	**4.951**	**2.887**	**3.914**	**6.295**	**3.943**	**6.256**	**4.480**	**6.875**	**46.053**

Fonte: Pnad 2013/IBGE.
Elaboração dos autores.

Nesta classificação mais agregada construída pelo IBGE, os setores da economia parecem estar mais homogeneamente distribuídos entre as regiões, com exceção da indústria de transformação, que permanece mais forte nas regiões Sul e Sudeste do Brasil. Destaca-se ainda como o Distrito Federal apresenta características muito diferentes das demais regiões no que diz respeito ao seu mercado de trabalho: a proporção de trabalhadores no setor público, na administração pública e nas "ciências e artes" é muito maior do que nas outras regiões metropolitanas.

A tabela 2 mostra que há grandes diferenças nos níveis salariais das regiões metropolitanas brasileiras. O que mais chama a atenção nos dados aqui é como essas diferenças se comportam ao longo da distribuição. Tomando o caso do Distrito Federal, por exemplo: o salário dos trabalhadores do primeiro decil é 0,97 vez o salário dos trabalhadores do primeiro decil na região metropolitana de São Paulo, enquanto entre os trabalhadores do último decil essa diferença cresce para 1,6 vez.

TABELA 2
Análise dos salários regionais em pontos da distribuição
(Em R$)

	10%	25%	50%	75%	90%
Distrito Federal	679	900	1.700	4.000	8.000
RM de Belém	500	678	854	1.500	2.600
RM de Belo Horizonte	678	800	1.200	2.000	4.000
RM de Curitiba	800	1.000	1.500	2.500	4.800
RM de Fortaleza	400	678	800	1.200	2.300
RM de Porto Alegre	678	900	1.300	2.100	4.300
RM do Recife	500	678	820	1.400	2.500
RM do Rio de Janeiro	678	800	1.200	2.100	4.500
RM de Salvador	400	678	900	1.600	3.500
RM de São Paulo	700	1.000	1.400	2.500	5.000

Fonte: Pnad 2013/IBGE.
Elaboração dos autores.

Outro aspecto que chama a atenção na tabela é o fato de o salário dos trabalhadores do primeiro decil não atingir o salário mínimo nacional (R$ 678,00) nas regiões metropolitanas de Belém, de Fortaleza, do Recife e de Salvador. Ou seja, uma maior proporção de empregos informais nessas regiões contribui para distanciá-las das demais, em que o salário no primeiro decil é equalizado pelo cumprimento do salário mínimo.

4 RESULTADOS

4.1 Regressão quantílica

Nesta seção serão apresentados os resultados da regressão quantílica, objetivando quantificar o efeito controlado da região ao longo da curva e distribuição dos salários.

TABELA 3
Regressão quantílica
(Em logaritmo do salário-hora)

	10%	25%	50%	75%	90%	MQO
Homem	0,1616	0,1895	0,2283	0,2775	0,2976	0,2446
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0004)
Branco	0,0674	0,0843	0,1135	0,1549	0,2097	0,1456
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0003)
Idade	0,0208	0,0272	0,0326	0,0370	0,0381	0,0341
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0001)
Idade^2	-0,0002	-0,0003	-0,0003	-0,0003	-0,0003	-0,0003
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)
Fundamental incompleto			Omitido			
Médio incompleto	0,0859	0,0946	0,1170	0,1311	0,1557	0,1376
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0005)
Superior incompleto	0,1895	0,2090	0,2562	0,3274	0,4019	0,3088
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0004)
Superior completo	0,5699	0,6977	0,8545	0,9904	1,1000	0,8694
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0006)
Outras indústrias	0,0933	0,2091	0,2384	0,2994	0,3751	0,2915
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0019)
Indústrias de transformação			Omitido			
Construção civil	0,0551	0,0500	0,0208	0,0008	0,0192	0,0487
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0007)
Comércio	-0,0557	-0,0344	-0,0515	-0,0423	0,0524	-0,0281
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0006)
Serviços	0,0243	0,0492	0,0250	0,0201	0,0700	0,0468
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0006)
Administração pública	0,2661	0,3737	0,4439	0,4224	0,4846	0,4351
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0011)
Serviços sociais	0,0085	0,0142	0,0073	0,0204	0,0682	0,0240
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0006)
Outras atividades	0,0513	0,0538	0,0594	0,0503	0,0924	0,0913
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0006)
Dirigentes	0,3331	0,5028	0,6670	0,7784	0,8523	0,6096
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0008)

(Continua)

(Continuação)

	10%	25%	50%	75%	90%	MQO
Ciências e artes	0,2595	0,3777	0,4760	0,5638	0,6029	0,4484
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0008)
Técnicos	0,1106	0,1688	0,2331	0,3049	0,4032	0,2413
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0007)
Trabalhadores administrativos	-0,0415	-0,0580	-0,0712	-0,0515	-0,0582	-0,0631
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0007)
Trabalhadores de serviços	-0,1255	-0,1225	-0,1369	-0,1213	-0,1073	-0,1182
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0005)
Trabalhadores de produção			Omitido			
Outras ocupações	-0,1165	-0,0106	0,0727	0,1086	0,1126	0,0321
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0012)
Setor público	0,0523	0,0366	0,0124	0,0159	-0,0373	0,0013
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0007)*
Setor formal	0,3360	0,1739	0,0584	-0,0648	-0,1472	0,0941
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0004)
RM de Belém	-0,2488	-0,2592	-0,2463	-0,2483	-0,2502	-0,2516
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0009)
RM de Belo Horizonte	-0,0750	-0,0846	-0,0666	-0,0629	-0,0957	-0,0797
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0006)
Distrito Federal	0,0206	0,0468	0,1160	0,1871	0,1940	0,1436
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0008)
RM de Curitiba	0,0429	0,0480	0,0598	0,0540	0,0548	0,0519
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0007)
RM de Fortaleza	-0,3693	-0,3683	-0,3713	-0,3659	-0,3793	-0,3794
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0007)
RM de Porto Alegre	-0,0488	-0,0659	-0,0708	-0,0803	-0,1209	-0,0997
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0006)
RM do Recife	-0,2908	-0,3055	-0,3041	-0,3313	-0,3054	-0,3048
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0007)
RM do Rio de Janeiro	-0,0777	-0,0710	-0,0396	-0,0075	0,0874	0,0277
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0004)
RM de Salvador	-0,2424	-0,2209	-0,1993	-0,1325	-0,0299	-0,1424
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0007)
RM de São Paulo			Omitido			
Constante	1,9137	2,0568	2,1745	2,3451	2,5828	2,1110
	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0000)	(0,0030)
Pseudo R^2 (R^2 para o MQO)	0,1998	0,2358	0,2872	0,3132	0,2771	0,4047

Fonte: Pnad 2013/IBGE.

Obs.: 1. Erros-padrão entre parênteses.

2. * = significativo a 10% de confiança. Todos os demais resultados são significativos a 1% de confiança.

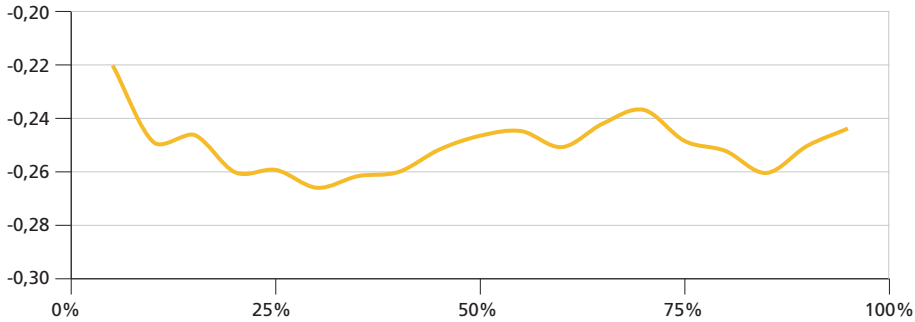
O comportamento dos retornos salariais para características pessoais, gênero e raça, adota a trajetória já estabelecida em trabalhos anteriores, como Bartalotti e Leme (2007). O crescente hiato salarial entre homens e mulheres e brancos e negros é condizente com a literatura que trata da questão, em um movimento nomeado de teto de vidro (*glass ceilling*). Os retornos educacionais também apresentam comportamento crescente ao longo da curva de distribuição, sinalizando que a educação tem um impacto maior sobre os trabalhadores mais bem remunerados. Não é um resultado surpreendente, uma vez que contribui para a visão de que os trabalhadores da base da distribuição são mais homogêneos.

Quanto às características do posto de trabalho, nota-se uma tendência geral por setor e por ocupação a aumentar o hiato salarial ao longo da distribuição. A única exceção parece ser o caso da construção civil, que apresenta um hiato favorável maior na base da distribuição do que no topo. Nota-se que a participação do trabalhador formal é um determinante importante do salário e apresenta um comportamento muito interessante: o hiato salarial entre trabalhadores dos setores formal e informal decresce ao longo da distribuição e inverte-se nas posições mais altas. Ou seja, para os trabalhadores da base da distribuição é vantajoso participar do mercado de trabalho formal, o que remete à discussão do papel do salário mínimo para equalizar o salário desse grupo de trabalhadores. Para os trabalhadores mais bem remunerados, é melhor não participar do mercado de trabalho formal, o que pode estar refletindo o peso dos tributos sobre o salário deles ou mesmo a opção por ser um trabalhador autônomo. Essa é uma conclusão próxima à feita por Tannuri-Pianto e Pianto (2002). Há ainda que se pontuar que o mesmo movimento é notado para a participação do trabalhador no setor público, semelhante aos resultados encontrados por Depalo, Giordano e Papapetrou (2015) para a zona do euro.

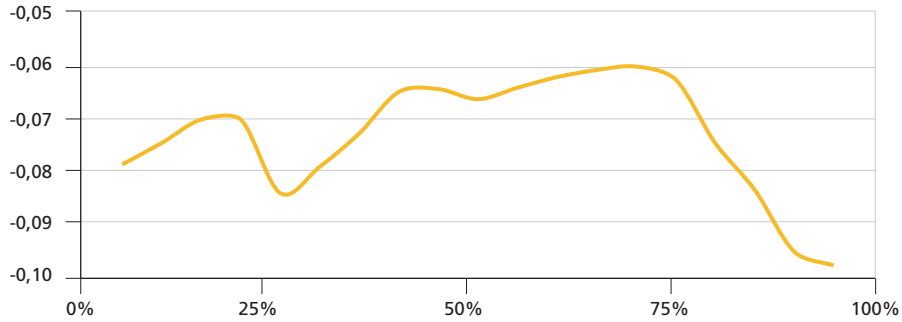
Quanto aos diferenciais regionais, o gráfico 1 traz o desenho da curva do efeito da localização do indivíduo sobre o salário em comparação à região metropolitana de São Paulo. O resultado é próximo ao obtido por Silveira-Neto e Campelo (2003). O primeiro ponto a se observar, ainda que já bem documentado pela literatura do tema, é a diferença de nível entre as curvas de diferenciais das regiões metropolitanas do Norte-Nordeste e do Sul-Sudeste do Brasil. As regiões metropolitanas de Belo Horizonte, de Porto Alegre e de Curitiba apresentam um diferencial máximo que nunca ultrapassa o valor de -0,2 logaritmo do salário-hora em relação a São Paulo, enquanto para as regiões metropolitanas do Recife, de Fortaleza e de Belém esse é o valor mínimo.

GRÁFICO 1
Resultados da regressão quantílica: diferenciais regionais salariais em relação à região metropolitana de São Paulo
(Em logaritmo do salário-hora)

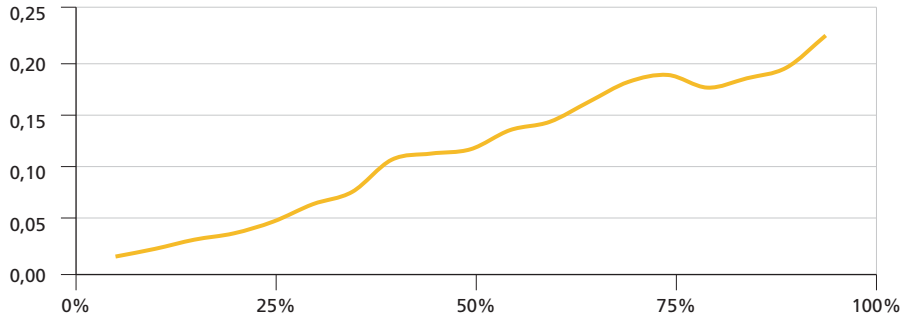
1A – RM de Belém



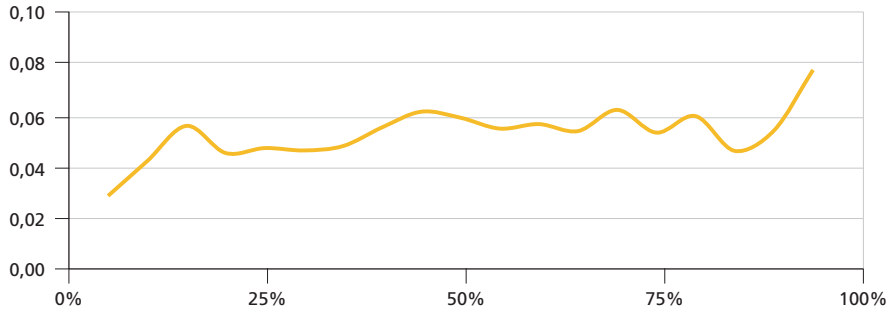
1B – RM de Belo Horizonte



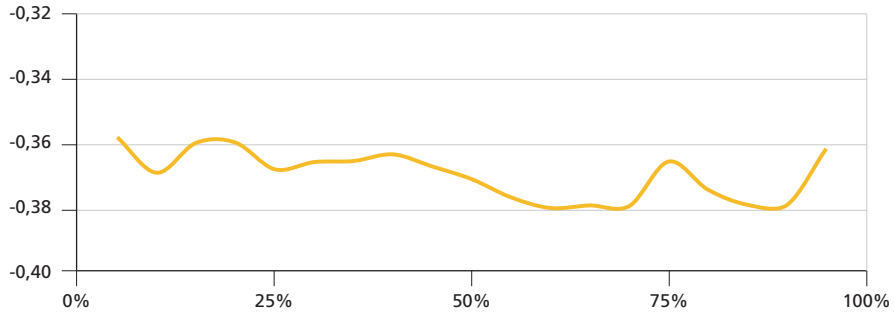
1C – Distrito Federal



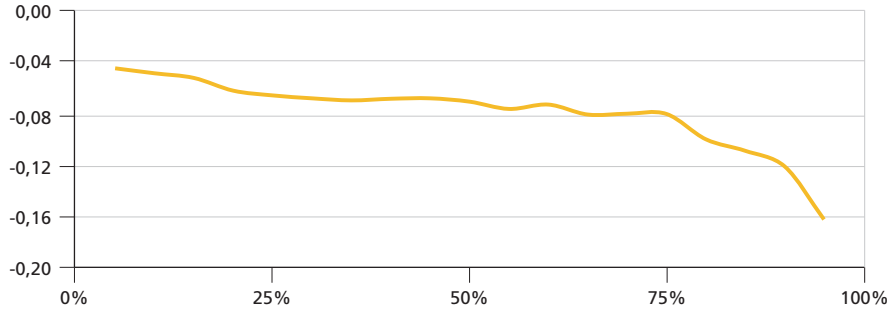
1D – RM de Curitiba



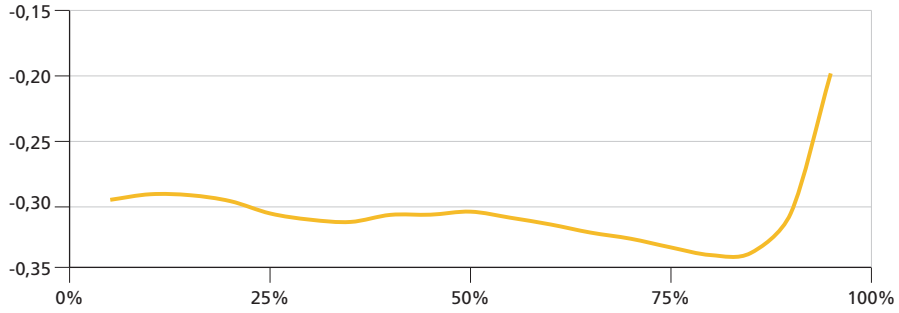
1E – RM de Fortaleza



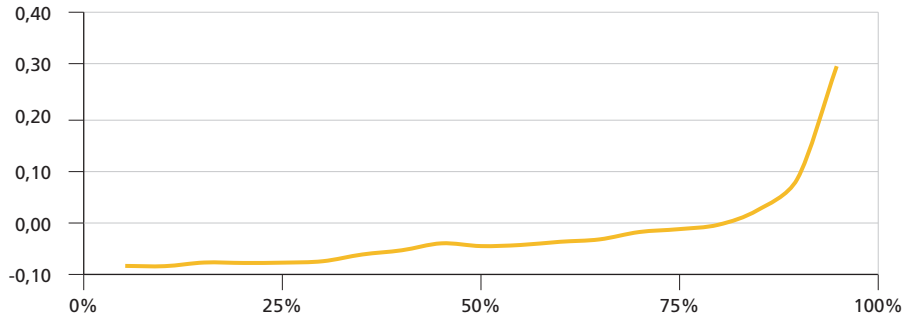
1F – RM de Porto Alegre



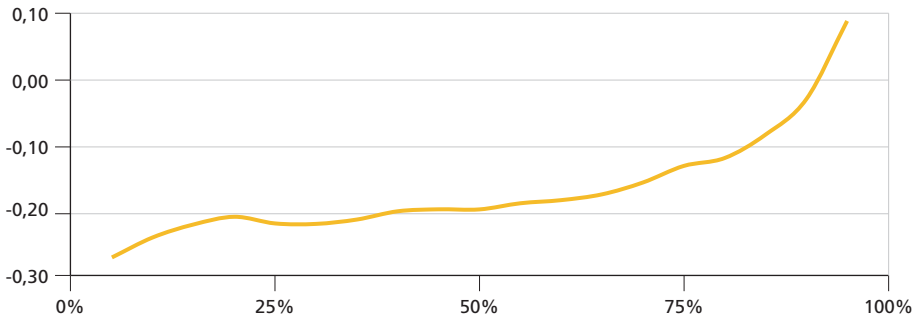
1G – RM do Recife



1H – RM do Rio de Janeiro



1I – RM de Salvador



Fonte: Pnad 2013/IBGE.
Elaboração dos autores.

Chama a atenção também o fato de a região metropolitana de Curitiba e o Distrito Federal apresentarem diferenciais salariais positivos em relação a São Paulo, ou seja, são regiões com salários maiores em todos os quantis apresentados. Há que se observar que, além de o Distrito Federal apresentar um nível de hiato salarial em relação a São Paulo significantemente superior ao da região metropolitana de Curitiba, o movimento da curva de diferenciais é bem distinto. Enquanto para Curitiba o diferencial salarial permanece relativamente constante, aumentando apenas nos últimos dois decis, para o Distrito Federal esse hiato é crescente por toda a distribuição.

Interessante se notar que, exceto para as regiões metropolitanas de Belo Horizonte e de Porto Alegre, nos últimos dois decis da distribuição há um forte ganho salarial, o que diminui o hiato salarial em regiões nas quais o nível dos salários é menor em relação a São Paulo (Belém, Fortaleza, Recife e Salvador) e aumenta em regiões em que o nível dos salários é maior ou semelhante ao de São Paulo (Brasília, Curitiba e Rio de Janeiro). Ou seja, não é um movimento de equiparação salarial em relação à região metropolitana de São Paulo, mas sim um aumento salarial generalizado no topo da distribuição. Por outro lado, nas regiões metropolitanas de Belo Horizonte e de Porto Alegre, que apresentam curvas de diferencias muito semelhantes, o hiato salarial aumenta em favor de São Paulo.

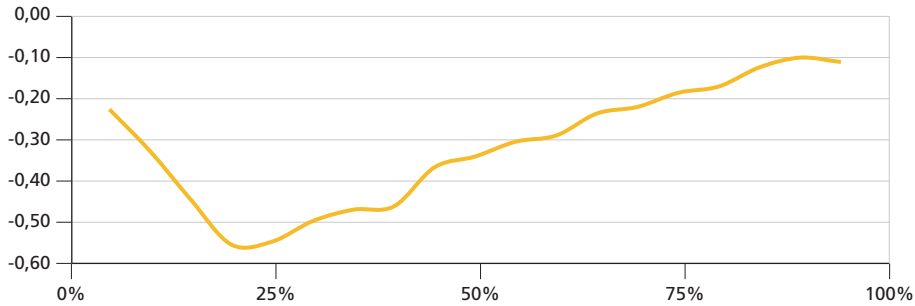
4.2 Decomposição quantílica

O próximo passo lógico é decompor essas curvas de diferenciais visando encontrar explicações para os resultados encontrados na seção anterior.

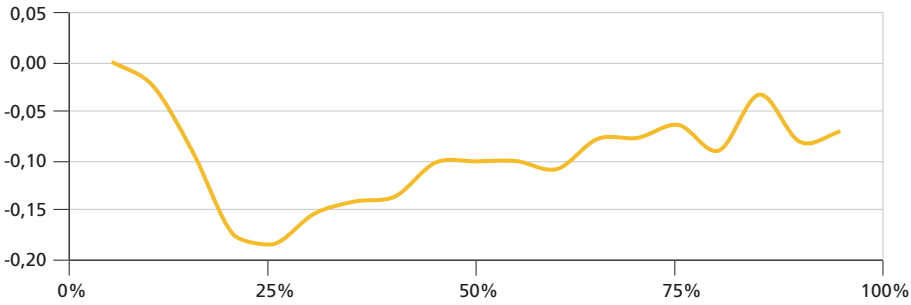
Seguindo a metodologia de Fortin, Lemieux e Firpo (2009), primeiro se estima as *recentered influence functions* (RIFs) para cada um dos diferenciais regionais de salários. O gráfico 2 traz o resultado. Todas as curvas de diferenciais assemelham-se quando estimadas por regressões quantílicas e por RIFs, com exceção da curva de diferenciais referente à região metropolitana de Curitiba, que aqui aparece negativamente inclinada nos quantis mais elevados. Pela estimação das RIFs, nos quantis mais altos os diferenciais regionais de Curitiba em relação a São Paulo não só diminuem, como mudam de sinal. Ou seja, nos quantis mais altos, os salários de Curitiba são menores que os de São Paulo, ao contrário do que ocorre nos quantis menores. O mais interessante de se pontuar aqui é que as curvas de diferenciais das regiões metropolitanas de Curitiba e de Porto Alegre têm um desenho muito semelhante, só se diferenciando quanto ao nível. Tendo em vista que são duas regiões com perfis econômicos muito próximos, levanta-se a dúvida de quão assertivo é esse método em relação às regressões quantílicas. Contudo, a rigor, a regressão quantílica traz resultados mais confiáveis, pois a RIF usa de aproximação linear. Logo, se o objetivo fosse apenas estimar as curvas de diferenciais, a regressão quantílica é a metodologia mais indicada, mas este trabalho pretende ir além e decompor essas curvas, o que é possível por meio da RIF. Os resultados apresentados aqui se referem à decomposição das RIFs.

GRÁFICO 2
RIF dos diferenciais regionais de rendimentos
(Em diferença do logaritmo do salário-hora em relação à região metropolitana de São Paulo)

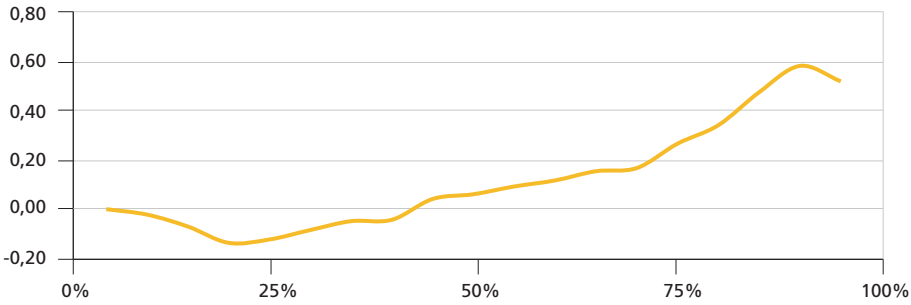
2A – RM de Belém



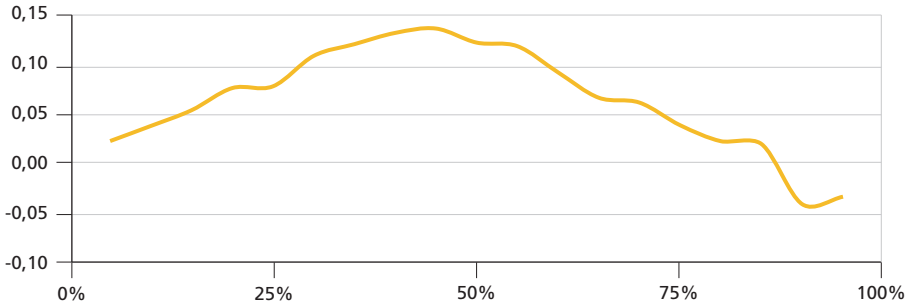
2B – RM de Belo Horizonte



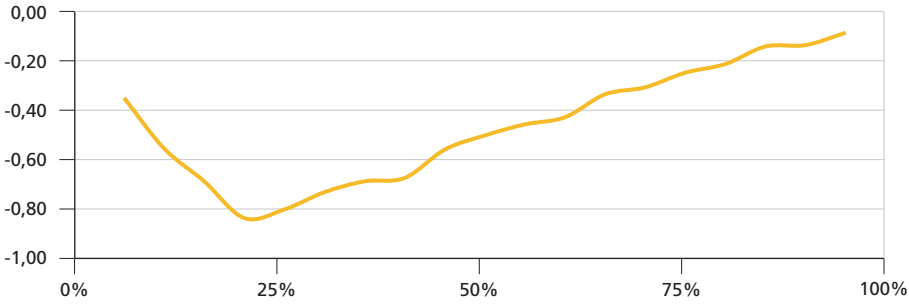
2C – Distrito Federal



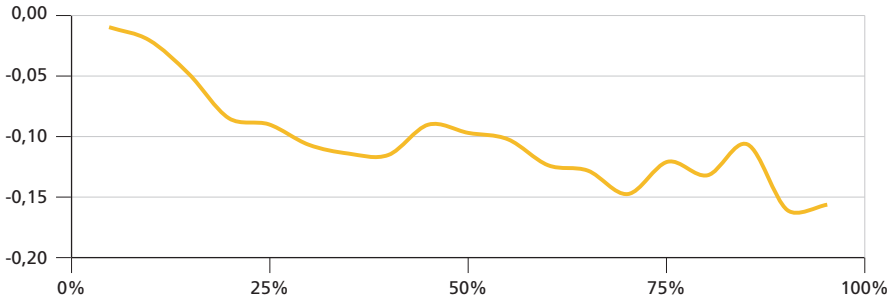
2D – RM de Curitiba



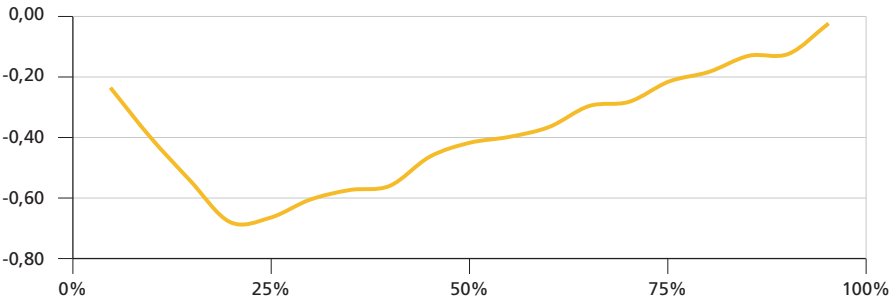
2E – RM de Fortaleza



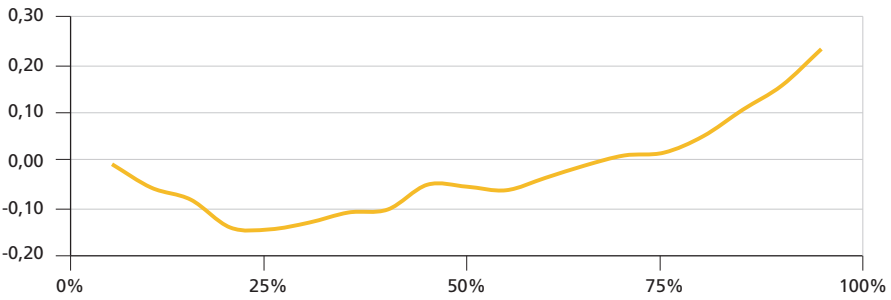
2F – RM de Porto Alegre



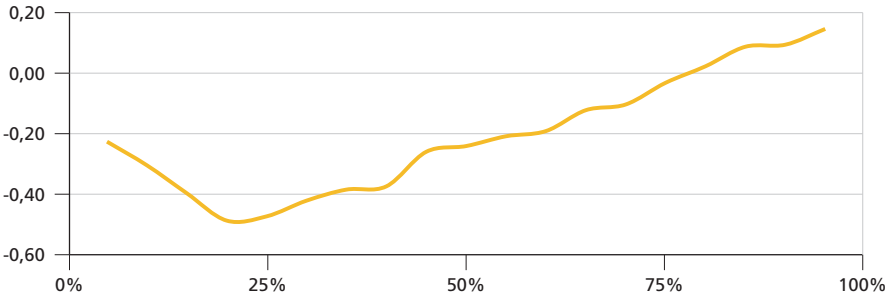
2G – RM do Recife



2H – RM do Rio de Janeiro



2I – RM de Salvador

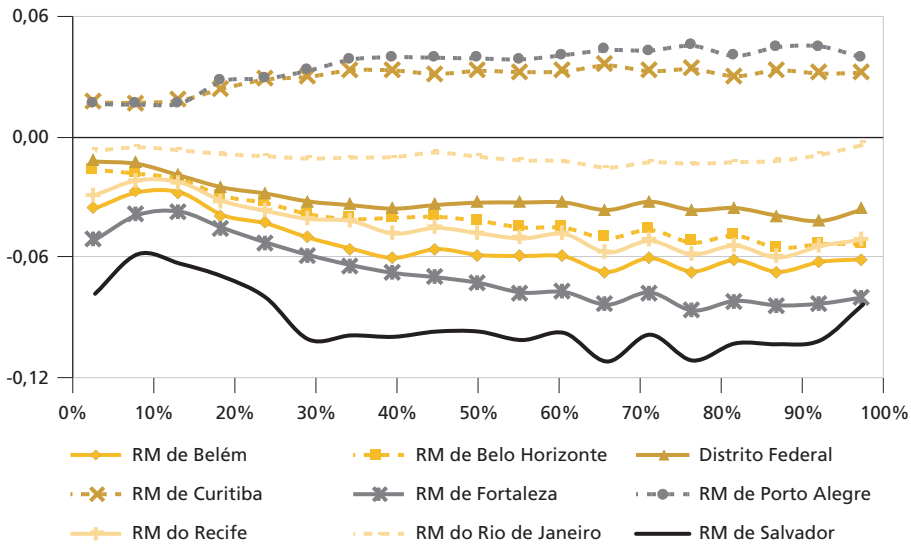


Fonte: Pnad 2013/IBGE.
Elaboração dos autores.

A decomposição é feita detalhadamente e agrupada em conjuntos de covariadas. São eles: características pessoais não produtivas, escolaridade, características do posto de trabalho, que juntas formam o efeito-composição, e o efeito-estrutura (ou efeito-preço), que fornece a parcela do diferencial explicada por diferenças nos retornos de cada atributo.

O gráfico 3 apresenta as curvas desenhadas a partir da estimação da decomposição dos diferenciais regionais em relação às características pessoais não produtivas. Conforme mostrado na tabela 1, a proporção de mulheres na amostra não tem grandes variações entre as regiões, o mesmo ocorrendo para as médias de idade. Ou seja, a hipótese mais forte aqui é que o efeito-composição das características pessoais dos trabalhadores está captando principalmente o papel do hiato salarial entre brancos e negros nos diferenciais regionais de salário. As regiões metropolitanas de Porto Alegre e de Curitiba são as únicas com a proporção de trabalhadores autodeclarados brancos maior do que a região de referência – 82% e 76%, respectivamente, enquanto na região metropolitana de São Paulo é de 60%. Isso explica o fato de o efeito-composição das características pessoais ser positivo apenas para essas duas regiões. Este fator pode ser um dos mais fortes determinantes para o nível salarial da região metropolitana de Curitiba ser maior que o da região metropolitana de São Paulo durante grande parte da distribuição. De modo análogo, o raciocínio pode ser aplicado a regiões com maior participação de negros no mercado de trabalho, sugerindo que a questão racial ajuda a explicar as diferenças de nível das curvas de diferenciais regionais de rendimentos.

GRÁFICO 3
Comparativo do efeito-composição: características pessoais
(Em diferença do logaritmo do salário-hora em relação à região metropolitana de São Paulo)



Fonte: Pnad 2013/IBGE.
Elaboração dos autores.

Com todas as outras regiões negativamente posicionadas em relação a São Paulo, espera-se que outros determinantes atuarão para elevar o nível salarial de Brasília e compensar o efeito-composição das características pessoais. O mesmo ocorrerá com a região metropolitana de Porto Alegre em sentido oposto: outros determinantes derrubarão o nível salarial em um valor que compensará o hiato positivo do efeito-composição das características pessoais.

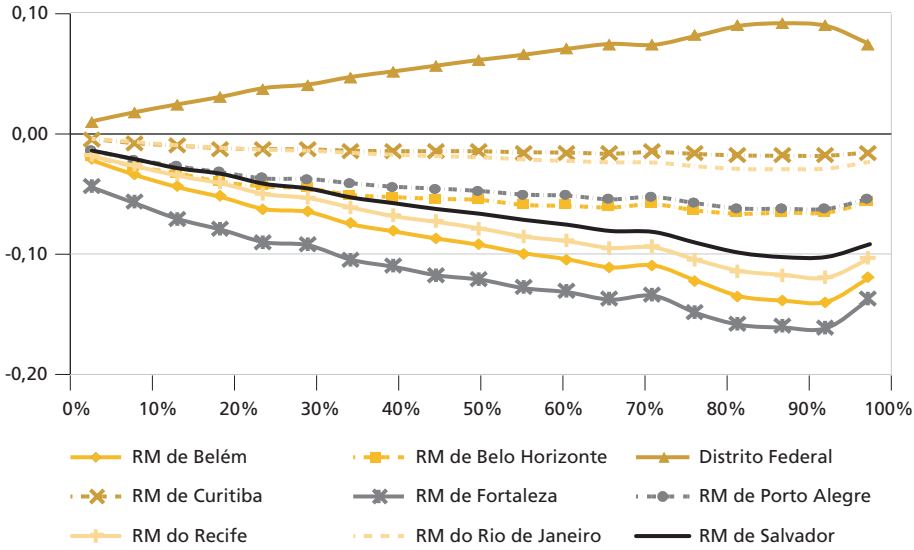
Há de se pontuar também que o efeito da composição das características pessoais sobre as disparidades de rendimentos entre as regiões metropolitanas cresce substancialmente no primeiro decil da distribuição, depois se estabilizando em um patamar mais alto. Ou seja, a questão do hiato salarial racial é menor entre os mais pobres.

O gráfico 4 apresenta as curvas desenhadas a partir da estimação da decomposição dos diferenciais regionais em relação à escolaridade. O efeito-composição da escolaridade sobre os diferenciais regionais de salário é um dos mais interessantes justamente por ocupar lugar de destaque na literatura sobre o tema. Para todas as regiões, há um aumento dos diferenciais regionais ocasionados por diferenças no nível educacional, sendo o Distrito Federal o único caso com hiato positivo em relação à região metropolitana de São Paulo. Vale frisar que a escolaridade não parece ser o principal determinante para a diferença de rendimentos entre São Paulo e Brasília, apesar de ter uma forte participação. Porém, para as regiões metropolitanas do Norte-Nordeste, a escolaridade aparenta ser um determinante muito forte para o hiato salarial regional.

Aqui, novamente as estatísticas descritivas apresentadas na tabela 1 explicam as diferenças em nível aqui apresentadas. Regiões com níveis de escolaridade próximos mostram-se agrupadas no gráfico, como Belo Horizonte e Porto Alegre, e Rio de Janeiro e Curitiba.

É interessante o movimento crescente do hiato regional atribuído à composição da escolaridade. Todas as regiões aparentam apresentar um diferencial crescente ao longo da distribuição, e esse crescimento é igual para todas as regiões metropolitanas, respeitando as diferenças de níveis entre elas. Nota-se também que o ponto em que o hiato é maior fica entre o 80º e o 90º centis em todas as regiões, sugerindo que o peso da escolaridade sobre as diferenças regionais é menor entre os 10% mais ricos do que entre o grupo imediatamente anterior.

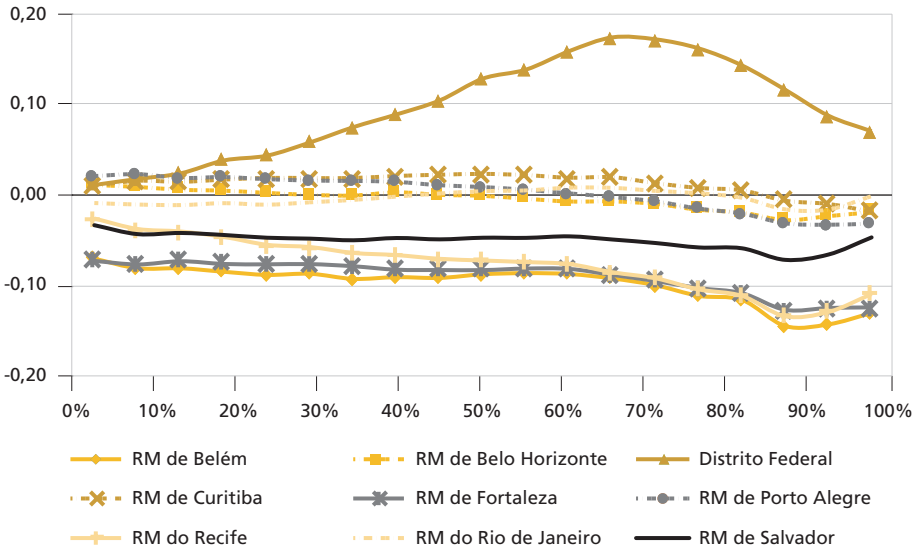
GRÁFICO 4
Comparativo do efeito-composição: escolaridade
(Em diferença do logaritmo do salário-hora em relação à região metropolitana de São Paulo)



Fonte: Pnad 2013/IBGE.
Elaboração dos autores.

O gráfico 5 apresenta as curvas desenhadas a partir da estimação da decomposição dos diferenciais regionais em relação às características do posto de trabalho do indivíduo. De modo geral, as características do posto de trabalho não aparentam ser determinantes significativos das diferenças regionais de salário para as regiões metropolitanas do Sul-Sudeste. Para as regiões metropolitanas do Norte-Nordeste, parecem ser determinantes mais fortes. Isso pode ser motivado pelo fato de a região de comparação, São Paulo, apresentar um mercado de trabalho com características mais próximas ao mercado de trabalho das regiões metropolitanas do Sul-Sudeste do que das do Norte-Nordeste. Esse resultado reforça o argumento de que uma parte significativa dos diferenciais regionais de salário entre o Sul-Sudeste e o Norte-Nordeste pode ser explicada por diferenças na demanda por trabalho.

GRÁFICO 5
Comparativo do efeito-composição: características do posto de trabalho
(Em diferença do logaritmo do salário-hora em relação à região metropolitana de São Paulo)



Fonte: Pnad 2013/IBGE.
Elaboração dos autores.

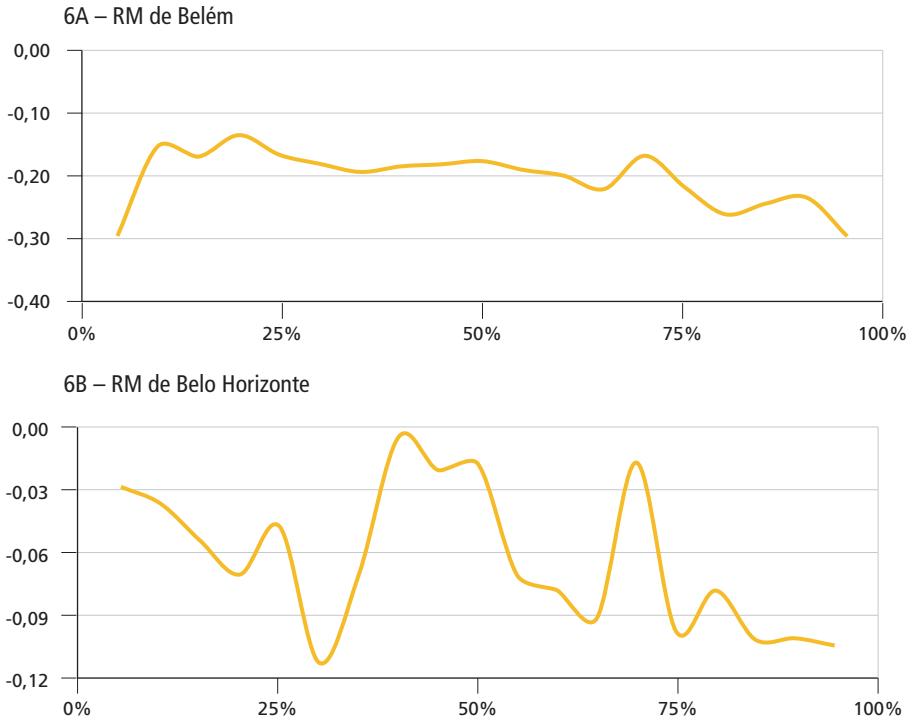
Como acontece com o efeito-composição da escolaridade sobre os diferenciais regionais, o efeito-composição das características do posto de trabalho contribuem para o aumento do hiato salarial ao longo da distribuição. As evidências aqui apontam que há um movimento de equiparação salarial no último decil da distribuição em relação a São Paulo, sugerindo que, em termos de posto de trabalho, os trabalhadores mais bem remunerados formam um grupo mais homogêneo entre as regiões. Isso explica por que os diferenciais diminuem em todas as regiões comparadas, inclusive para o Distrito Federal.

Já o Distrito Federal mostra-se um caso especial a ser analisado justamente em função do seu papel de centro administrativo do país e, consequentemente, com a forte participação do setor público no mercado de trabalho local. As características do posto de trabalho, em especial a participação no setor público e no setor de administração pública, são determinantes mais fortes do que a escolaridade para explicar o hiato salarial entre Brasília e São Paulo. É uma conclusão que reforça ainda mais o argumento do impacto de aspectos da demanda por trabalho sobre os salários. Esse movimento pode ser atribuído às próprias características de participação no setor público: o setor privado tende a ter salários piores no começo da distribuição, enquanto remunera melhor o topo da distribuição. Pode-se supor que trabalhadores mal remunerados tenham características do posto de trabalho semelhantes entre as regiões, o que explicaria o diferencial mais baixo nos primeiros quantis.

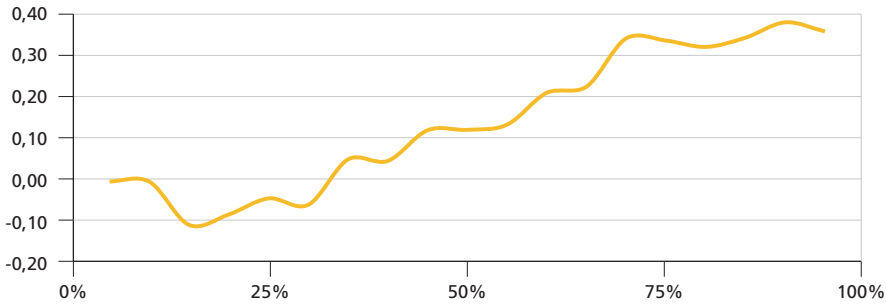
Sobre a parcela do efeito-preço nos diferenciais regionais de rendimentos, ou seja, nas diferenças causadas por discrepância nos retornos sobre as variáveis do modelo, o primeiro ponto a se destacar é que parece ser o mais forte determinante do hiato salarial das regiões analisadas em relação a São Paulo. Esse resultado já tinha sido apresentado em diversos trabalhos, tais como Menezes, Fernandez e Dedecca (2005) e Reis e Barros (1990). O gráfico 6 traz o desenho das curvas do efeito-preço. Pode-se notar como o grupo formado pelas regiões metropolitanas do Norte-Nordeste tem um diferencial atribuído ao efeito-preço consideravelmente maior do que o grupo das regiões metropolitanas do Sul-Sudeste. Mais uma vez, o Distrito Federal aparece com um diferencial fortemente positivo em relação a São Paulo.

Alguns pontos são interessantes sobre o comportamento do efeito-preço ao longo da curva de distribuição dos salários. Para todas as regiões, há um aumento do diferencial do efeito-preço no topo da distribuição, chamando a atenção para o fato de que, para o Distrito Federal e as regiões metropolitanas de Salvador e do Rio de Janeiro, esse diferencial é positivo. Ou seja, para os trabalhadores no topo da distribuição, o retorno das variáveis em cada região comparadas a São Paulo parece ser maior.

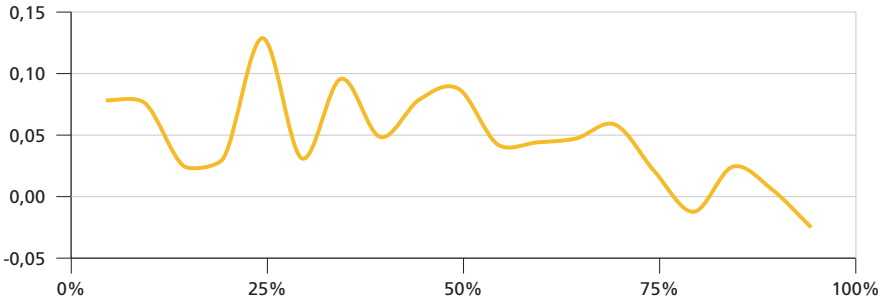
GRÁFICO 6
Comparativo do efeito-estrutura salarial
(Em diferença do logaritmo do salário-hora em relação à região metropolitana de São Paulo)



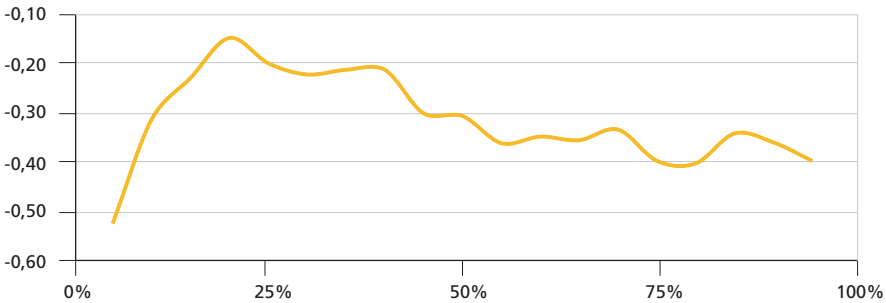
6C – Distrito Federal



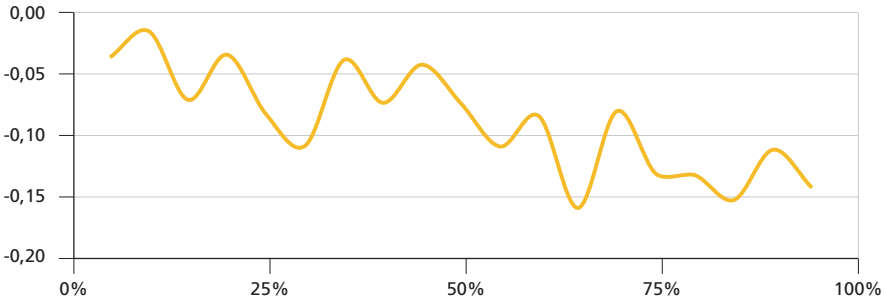
6D – RM de Curitiba

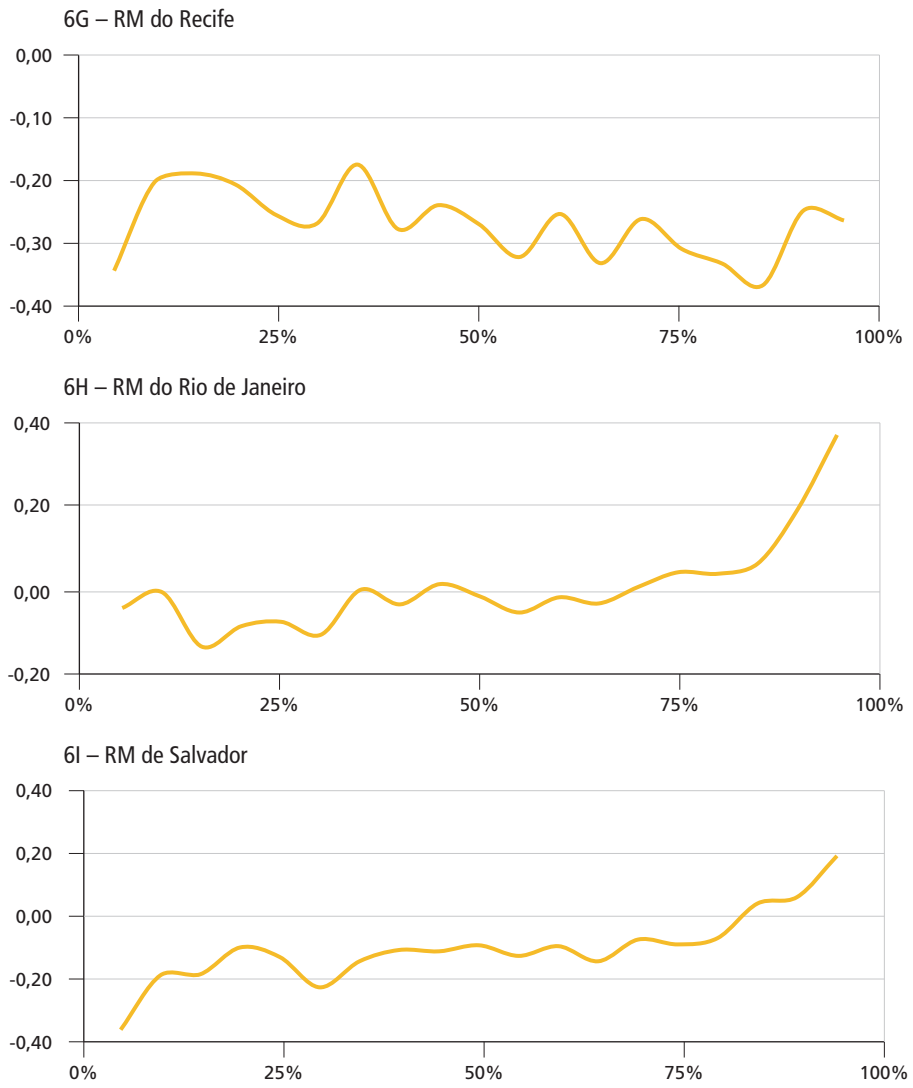


6E – RM de Fortaleza



6F – RM de Porto Alegre





Fonte: Pnad 2013/IBGE.
Elaboração dos autores.

Para as regiões metropolitanas de Belo Horizonte, de Curitiba e de Porto Alegre, nota-se curvas bastante similares. O efeito-preço das disparidades regionais parece crescer a taxas constantes ao longo da distribuição, sendo praticamente nulo (ou até mesmo positivo, no caso de Curitiba) nos primeiros quantis. A interpretação seria que, para indivíduos com as mesmas características (produtivas e não produtivas), o salário seria o mesmo morando em uma dessas regiões ou em São Paulo para os trabalhadores com menores salários, mas isso não se manteria para trabalhadores

com salários maiores. Nas regiões metropolitanas de Belém, de Fortaleza e do Recife, a interpretação é similar, diferindo que, para os trabalhadores do primeiro decil, o efeito-preço é muito forte. Uma boa explicação seria o efeito do mercado formal de trabalho: como é visto na tabela 3, nas regiões metropolitanas de Belém, de Fortaleza, do Recife e de Salvador os trabalhadores dentro do primeiro decil da distribuição ganham menos do que um salário mínimo, ou seja, participam do mercado informal de trabalho. Como as outras regiões apresentam salários acima ou iguais ao salário mínimo no primeiro decil, há uma equalização de salários por motivos externos ao funcionamento do mercado de trabalho, o que resulta em uma anulação dos diferenciais regionais.

As regiões metropolitanas do Rio de Janeiro e de Salvador apresentam um comportamento muito particular: o efeito-preço dos diferenciais regionais é negativo ao longo da maior parte da distribuição, para então se tornar fortemente positivo. Uma hipótese para explicar esse comportamento é a própria singularidade da atividade econômica nessas duas regiões. O Rio de Janeiro assume um papel de complementaridade econômica em relação a São Paulo, formando a única megalópole do hemisfério sul. De modo semelhante, Salvador tem um papel particular na economia brasileira por ser o centro econômico da região Nordeste do país, um posto que não existe para a região Sul, por exemplo, em que esse papel ainda cabe a São Paulo. Ou seja, centros econômicos diferenciados atraem uma mão de obra especializada no topo da distribuição, o que explica o desenho de suas curvas de diferenciais. Isso não necessariamente significa que essas regiões respondem diferentemente ao efeito-preço, apenas que a comparação com a região metropolitana de São Paulo pode maquiar esse efeito.

O resultado geral é que o diferencial entre os retornos das variáveis é maior para os trabalhadores do topo da distribuição do que para os da base. Isso mostra que, semelhantemente ao que acontece com o efeito-composição, os trabalhadores da base da distribuição são mais homogêneos entre as regiões. A hipótese para que isso ocorra é que a remuneração entre esses trabalhadores é semelhante porque há pouca variabilidade de empregos entre regiões na base da distribuição. Seria como se ocorresse um efeito *sorting* inverso: como as posições na base da distribuição não exigem qualificação extra, a disparidade regional dos salários desse grupo será muito menor. As diferenças de nível das curvas de diferenciais, assim, poderiam ser atribuídas às economias de aglomeração, que elevariam os níveis dos salários de toda a distribuição. Esses são resultados parecidos aos encontrados em trabalhos aplicados a outros países (Autor, Katz e Kearney, 2005; Goos e Manning, 2007; Matano e Naticchoni, 2013; Galego e Pereira, 2013).

5 CONCLUSÃO

Este trabalho teve como objetivo entender melhor como as disparidades regionais de rendimentos se comportam ao longo da curva de distribuição de salários e os determinantes desse comportamento. A primeira conclusão a que se chega é confirmar que realmente há um hiato salarial considerável entre as regiões metropolitanas das regiões Sul-Sudeste e Norte-Nordeste, e ele é persistente por toda a distribuição.

O que este trabalho traz de novo são respostas sobre como os determinantes se comportam em diferentes pontos da distribuição. Sobre o efeito-composição, nota-se que ele é consideravelmente menor na base da distribuição para todas as regiões, ou seja, os trabalhadores com menores rendimentos formam um grupo mais homogêneo. O efeito da composição da força de trabalho sobre os diferenciais regionais cresce ao longo da distribuição, o que por si só já fornece muitas informações sobre desigualdades regionais no Brasil. Entre as variáveis escolhidas, como esperado, a escolaridade mostra-se a principal componente do efeito-composição, reforçando a ideia de que qualquer política pública que objetive diminuir o hiato salarial entre regiões deve dar especial atenção à questão. Ainda chama a atenção para como o diferencial por cor também contribui para as disparidades regionais, um ponto muito interessante e que merece mais estudos para ser melhor compreendido.

O efeito-preço, a diferença entre os retornos das covariadas, mostra-se como o principal determinante das disparidades regionais, conclusão a qual outros trabalhos já haviam chegado (Menezes, Fernandez e Dedecca, 2005; Reis e Barros, 1990). O que este trabalho traz de novo é mostrar que o efeito-preço também é crescente ao longo da distribuição: os trabalhadores mais bem remunerados são os mais afetados pelo efeito regional, conclusão semelhante à de trabalhos aplicados a outros países (Autor, Katz e Kearney, 2005; Goos e Manning, 2007; Matano e Naticchoni, 2013; Galego e Pereira, 2013). Se à primeira vista pode parecer muito forte chamar o efeito-preço de efeito regional, um bom argumento a favor dessa interpretação é o comportamento das curvas de efeito-preço para cada região. É possível distinguir-se claramente que regiões com dinâmicas econômicas semelhantes apresentam curvas de efeito-preço semelhantes, embora haja pontos em comuns entre todas. Para as regiões do Norte-Nordeste, por exemplo, nota-se que o efeito regional é forte também para os 10% mais pobres, e que isso pode ser atribuído a um maior grau de informalidade nessas regiões metropolitanas, como argumentado na seção anterior. Pontua-se aqui que o incentivo à maior formalização do mercado de trabalho contribuiria para diminuir o hiato salarial entre Sul-Sudeste e Norte-Nordeste, principalmente entre os mais pobres.

Este é um resultado importante, pois uma das maiores limitações da metodologia escolhida é a impossibilidade de distinguir efeito-preço e efeito das variáveis não

observáveis. O fato de haver uma forte correlação entre o desenho das curvas e a dinâmica econômica da região depõe a favor da interpretação do efeito-preço como efeito regional. Assim, há duas explicações possíveis para esse fenômeno: ou os trabalhadores de regiões distintas não apresentam diferenças significantes de suas características não observáveis, ou essas características são tão fortemente associadas à região que é captada como um componente do efeito regional. Os resultados encontrados por Freguglia, Menezes-Filho e Souza (2007) usando a região metropolitana de Belo Horizonte como região de comparação indicam que a primeira explicação é a mais plausível. Ou seja, não há evidências que sustentam o argumento da heterogeneidade dos trabalhadores de regiões distintas como explicação para os diferenciais salariais. O diferencial de produtividade é creditado em sua maior parte ao efeito regional, mais especificamente aos efeitos das economias de aglomeração.

É sempre importante ressaltar que este trabalho foca os indivíduos para explicar os diferenciais regionais. Ou seja, o conjunto dos diferenciais entre os trabalhadores constrói o hiato regional. O que não é a explicação completa para a questão, uma vez que as características da região e dos indivíduos interagem. Para ficar em apenas alguns exemplos, Combes, Duranton e Gobillon (2008) destacam o fato de regiões economicamente mais dinâmicas atraírem trabalhadores mais qualificados, e Glaeser e Maré (2001) e Moretti (2004) discutem como as externalidades do capital humano impactam a produtividade do trabalho. São explicações importantes para explicar o diferencial do efeito-preço entre as regiões metropolitanas no Brasil e devem ser alvos de mais trabalhos para elucidar a questão.

De um modo geral, este artigo conclui que a dinâmica da desigualdade regional de salários no Brasil tem estrutura semelhante à de outros países, em que o efeito regional é maior para os trabalhadores do topo da distribuição. Só que o caso brasileiro é agravado pela alta desigualdade na composição da força de trabalho, e este é um resultado que diz muito sobre o problema da desigualdade regional no Brasil.

REFERÊNCIAS

AUTOR, D.; KATZ, L.; KEARNEY, M. **Rising wage inequality**: the role of composition and prices. Massachusetts: NBER, 2005. (Working Paper, n. 11628). Disponível em: https://goo.gl/uR5WRo. Acesso em: 19 nov. 2014.

BARROS, R. P.; MENDONÇA, R. **Os determinantes da desigualdade no Brasil**. (Texto para Discussão, n. 377). Rio de Janeiro: Ipea, 1995. Disponível em: https://goo.gl/iXGzVY. Acesso em: 19 nov. 2014.

BARTALOTTI, O.; LEME, M. C. S. Discriminação salarial além da média: uma abordagem de decomposição contrafactual utilizando regressões quantílicas.

In: ENCONTRO NACIONAL DE ECONOMIA, 35., 2007, Recife. **Anais** [...]. Recife: Anpec, 2007. Disponível em: https://goo.gl/KQwCcQ. Acesso em: 19 nov. 2014.

BERNARD, A. *et al.* Relative wage variation and industry location in the United Kingdom. **Oxford Bulletin of Economics and Statistics**, v. 70, n. 4, p. 431-459, 2003.

BLINDER, A. Wage discrimination: reduced form and structural estimates. **Journal of Human Resources**, v. 8, n. 4, p. 436-455, 1973.

BUCHINSKY, M. Recents advances in quantile regression models: a practical guideline for empirical research. **Journal of Human Resources**, v. 33, p. 88-126, 1998a.

_____. The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. **Journal of Applied Econometrics**, n. 13, p. 1-30, 1998b.

_____. Quantile regression with sample selection: estimating women's return to education in the U.S. **Empirical Economics**, n. 26, p. 87-113, 2001.

COELHO, D.; VESZTEG, R.; SOARES, F. V. Regressão quantílica com correção para seletividade amostral: estimativa dos retornos educacionais e diferenciais raciais na distribuição de salários das mulheres no Brasil. **Pesquisa e Planejamento Econômico**, Brasília, v. 40, n. 1, p. 85-102, 2010.

COMBES, P.; DURANTON, G.; GOBILLON, L. Spatial wage disparities: sorting matters! **Journal of Urban Economics**, v. 63, n. 2, p. 723-742, 2008.

DEPALO, D.; GIORDANO, R.; PAPAPETROU, E. **Public-private wage differentials in euro area countries**: evidence from quantile decomposition analysis. Rome: Bank of Italy, 2015. (Economic Working Papers, n. 907).

DINARDO, J.; FORTIN, N. M.; LEMIEUX, T. Labor market institutions and the distribution of wages, 1973-1992: a semiparametric approach. **Econometrica**, v. 64, n. 5, p. 1001-1044, 1996.

DUARTE, A.; FERREIRA, P.; SALVATO, M. Disparidades regionais ou educacionais? *In:* ENCONTRO NACIONAL DE ECONOMIA, 31., 2003, Porto Seguro. **Anais** [...]. Porto Seguro: Anpec, 2003. Disponível em: https://goo.gl/sG8V72. Acesso em: 19 nov. 2014.

FIGUEIREDO, E. A.; SILVA, C. R. F. Desigualdade de oportunidades no Brasil: uma decomposição quantílica contrafactual. **Pesquisa e Planejamento Econômico**, Rio de Janeiro, v. 42, n. 1, p. 41-60, 2012.

FONTES, G. G.; SIMÕES, R. F.; OLIVEIRA, A. M. H. C. Urban attributes and wage disparities in Brazil: a multilevel hierarchical model. **Regional Studies**, v. 44, p. 595-607, 2010.

FORTIN, N.; LEMIEUX, T.; FIRPO, S. Unconditional quantile regressions. **Econometrica**, v. 77, n. 3, p. 953-973, 2009.

_____; _____; _____. Decomposition methods in economics. *In:* ACEMOGLU, D.; AUTOR, D. (Eds.). **Handbook of Labor Economics**. Amsterdam: Elsevier, 2011.

FREGUGLIA, R. S.; MENEZES-FILHO, N.; SOUZA, D. B. Diferenciais salariais inter-regionais, interindustriais e efeitos fixos individuais: uma análise a partir de Minas Gerais. **Estudos Econômicos**, v. 37, n. 1, p. 129-150, 2007.

GALEGO, A.; PEREIRA, J. **Decomposition of regional wage differences along the wage distribution in Portugal**: the importance of the covariates. Portugal: University of Évora, 2013. (Cefage-UE Working Papers, n. 2013-16).

GLAESER, E.; MARÉ, D. Cities and skills. **Journal of Labor Economics**, v. 19, n. 2, p. 316-342, 2001.

GOOS, M.; MANNING, A. Lousy and lovely jobs: the rising polarization of work in Britain. **The Review of Economics and Statistics**, v. 89, n. 1, p. 118-133, 2007.

HAMPEL, F. R. The influence curve and its role in robust estimation. **Journal of the American Statistical Association**, v. 60, p. 383-393, 1974.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Pesquisa Nacional por Amostra de Domicílios**. Rio de Janeiro: IBGE, 2013.

JUHN, C.; MURPHY, K.; PIERCE, B. Wage inequality and the rise in returns to skill. **Journal of Political Economy**, v. 101, n. 3, p. 410-442, 1993.

KOENKER, R.; BASSET, G. Regression quantiles. **Econometrica**, v. 46, n. 1, p. 33-50, 1978.

_____; _____. Quantile regression. **Journal of Economics Perspectives**, v. 15, n. 4, p. 143-156, 2001.

MACHADO, J.; MATA, J. Counterfactual decomposition of changes in wage distributions using quantile regression. **Journal of Applied Econometrics**, v. 20, p. 445-465, 2005.

MACIEL, M. C.; CAMPÊLO, A. K.; RAPOSO, M. C. F. A dinâmica das mudanças na distribuição salarial e no retorno em educação para mulheres: uma aplicação de regressão quantílica. *In:* ENCONTRO NACIONAL DE ECONOMIA, 29., 2001, Salvador. **Anais** [...]. Salvador: Anpec, 2001. Disponível em: https://goo.gl/39wVqq. Acesso em: 19 nov. 2014.

MARTINS, P.; PEREIRA, P. Does education reduce wage inequality? Quantile regressions evidence from 16 countries. **Labour Economics**, v. 11, n. 3, p. 355-371, 2004.

MATANO, A.; NATICCHIONI, P. **What drives the urban wage premium?** Evidence along the distribution. Bonn: IZA, 2013. (Working Paper, n. 7811).

MELLY, B. **Estimation of counterfactual distributions using quantile regression**. St. Gallen: University of St. Gallen, 2006. Mimeo.

MENEZES, W.; FERNANDEZ, J. C.; DEDECCA, C. Diferenciações regionais de rendimentos do trabalho: uma análise das regiões metropolitanas de São Paulo e Salvador. **Estudos Econômicos**, São Paulo, v. 35, n. 2, p. 271-296, 2005.

MORETTI, E. Human capital externalities in cities. *In:* HENDERSON, J. V.; THISSE, J. F. (Eds.). **Handbook of Regional and Urban Economics**. Amsterdam: Elsevier, 2004.

NEUMAN, S.; OAXACA, R. Wage decompositions with selectivity-corrected wage equations: a methodological note. **Journal of Economic Inequality**, v. 2, n. 1, p. 3-10, 2004.

OAXACA, R. Male female wage differentials in urban labor markets. **International Economic Review**, v. 14, n. 3, p. 693-709, 1973.

PESSÔA, S. Economia regional, crescimento econômico e desigualdade regional de renda. *In:* ENCONTRO NACIONAL DE ECONOMIA, 29., 2001, Salvador. **Anais** [...]. Salvador: Anpec, 2001. Disponível em: https://goo.gl/PEydxy. Acesso em: 19 nov. 2014.

REIS, J. G. A.; BARROS, R. P. Desigualdade salarial e distribuição de educação: a evolução das diferenças regionais no Brasil. **Pesquisa e Planejamento Econômico**, Rio de Janeiro, v. 20, n. 3, p. 415-478, 1990.

SAVEDOFF, W. Os diferenciais regionais de salário no Brasil: segmentação versus dinamismo da demanda. **Pesquisa e Planejamento Econômico**, Rio de Janeiro, v. 20, n. 3, p. 521-556, 1990.

SILVEIRA-NETO, R. M.; CAMPELO, A. K. O perfil das disparidades regionais de renda no Brasil: evidências a partir de regressões quantílicas para os anos de 1992 e 2001. *In:* ENCONTRO NACIONAL DE ECONOMIA, 31., Porto Seguro, 2003. **Anais** [...]. Porto Seguro: Anpec, 2003. Disponível em: https://goo.gl/bhWQFa. Acesso em: 19 nov. 2014.

SIMÕES, R. F.; FREITAS, E. E. Urban attributes and regional differences in productivity: evidence from the external economics of Brazilian micro-regions from 2000-2010. **International Jounal of Economics**, v. 1, n. 2, p. 30-44, 2014.

TANNURI-PIANTO, M. E.; PIANTO, D. Informal employment in Brazil: a choice at the top and segmentation at the bottom: a quantile regression approach.

In: ENCONTRO BRASILEIRO DE ECONOMETRIA, 24., 2002, Nova Friburgo. **Anais** [...]. Nova Friburgo: SBE, 2002.

WHALLEY, J.; XING, C. **The regional distribution of skill premia in urban China**. Massachusetts: NBER, 2010. (Working Paper, n. 16575). Disponível em: https://goo.gl/qQyEjA. Acesso em: 19 nov. 2014.

WHEATON, W.; LEWIS, M. Urban wages and labor Market agglomeration. **Journal of Urban Economics**, Cambridge, v. 51, p. 542-562, 2002.

BIBLIOGRAFIA COMPLEMENTAR

CAMPANTE, F.; CRESPO, A.; LEITE, P. Desigualdade salarial entre raças no mercado de trabalho urbano brasileiro: aspectos regionais. **Revista Brasileira de Economia**, v. 58, n. 2, p. 185-210, 2004.

CIRINO, J. F.; DALBERTO, C. R. Trabalhadores formais versus informais: diferenças de rendimento para a região metropolitana de Belo Horizonte. *In:* SEMINÁRIO SOBRE A ECONOMIA MINEIRA, 14., 2014, Diamantina. **Anais** [...]. Diamantina: UFMJ, 2014.

DURANTON, G.; MONASTIRIOTIS, V. Mind the gaps: the evolution of regional earnings inequalities in the U.K., 1982-1997. **Journal of Regional Science**, v. 42, n. 2, p. 219-256, 2002.

FERNANDES, R.; MENEZES-FILHO, N. A evolução da desigualdade no Brasil metropolitano entre 1983 e 1997. **Estudos Econômicos**, v. 30, n. 4, p. 549-569, 2000.

GHINETTI, P.; LUCIFORA, C. Public-private wage gaps and skill levels: evidence from French, British and Italian micro data. **International Journal of Manpower**, v. 34, n. 5, p. 429-446, 2013.

HALFDANARSON, B.; HEUERMANN, D.; SÜDEKUM, J. **Human capital externalities and the urban wage premium**: two literatures and their interrelations. Bonn: IZA, 2008. (Working Paper, n. 3493).

MORETTI, E. Local labor markets. *In:* ACEMOGLU, D.; AUTOR, D. (Eds.). **Handbook of Labor Economics**. Amsterdam: Elsevier, 2011.

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

Lauro Nogueira[1]
Erik Figueiredo[2]

Este estudo investiga duas importantes questões referentes ao desempenho na avaliação do Programa Internacional de Avaliação de Estudantes (Pisa) 2012. A priori, estima-se a transmissão intergeracional da educação. Em seguida, faz-se uma análise contrafatual derivada de um aumento no nível socioeconômico parental dos estudantes que prestaram o exame. Os principais resultados apontam para uma incipiente transmissão educacional parental nos países da América do Sul. Especificamente no Brasil, apura-se um valor aproximadamente sete vezes inferior ao encontrado na República Tcheca e 20% do total encontrado na França e no Japão. Além disso, observa-se um *gap* médio para educação parental em torno de 8%, em que 2% desse diferencial é explicado pelo nível educacional parental da distribuição.

Palavras-chave: transmissão intergeracional da educação; igualdade de oportunidades educacionais; efeito-tratamento.

LIKE FATHER LIKE SON? AN ANALYSIS OF FACTORING EFFECTS OF CIRCUMSTANCES ON THE PERFORMANCE OF STUDENTS IN THE PISA EVALUATION 2012

This study investigates two important questions regarding performance in the evaluation of the International Student Assessment Program (Pisa) 2012. Initially, it is estimated the intergenerational transmission of education. Then, a counterfatual analysis is made, derived from an increase in the parental socioeconomic level of the students who took the exam. The main results point to an insipient parental educational transmission in the countries of South America. Specifically in Brazil, it is estimated to be approximately seven times lower than that found in the Czech Republic and 20% of the total found in France and Japan. In addition, there is an average gap for parental education of around 8%, where 2% of this difference is explained by parental educational level of distribution.

Keywords: intergenerational transmission of education; equal educational opportunities; treatment effect.

JEL: I20; I21; D63.

1. Professor do Departamento de Ciências Sociais Aplicadas da Universidade Federal Rural do Semiárido (UFERSA) e do PPGE da Universidade do Estado do Rio Grande do Norte (UERN).

2. Professor do Programa de Pós-Graduação em Economia (PPGE) da Universidade Federal da Paraíba (UFPB). Bolsista de Produtividade 1D do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

1 INTRODUÇÃO

Em geral, economistas e cientistas sociais divergem sobre a maior parte das teorias econômicas. Contudo, educação de qualidade é consensualmente vista como forte indicador de bem-estar. Nesse contexto, Heckman (2011) relata que o diferencial econômico e intelectual dos agentes tem origem já a partir da primeira infância – 0 a 6 anos –, e, portanto, a desigualdade de oportunidades educacionais acarreta importantes diferenciais na habilidade cognitiva e não cognitiva. Além disso, o estudo relata que sociedades que possuem baixa qualidade educacional tendem a apresentar baixo grau de produtividade individual, e, portanto, menores níveis de renda *per capita*.

Nessa linha de pensamento, Daude (2011) advoga que o acesso à educação de qualidade constitui uma ferramenta poderosa para incrementar o sucesso individual, e que a provisão social educacional para todos é uma necessidade insubstituível no processo para equalização de oportunidades, notadamente dos menos favorecidos.[3] Nesse sentido, devem ser garantidas determinadas condições para que isso ocorra. Por exemplo, um sistema educacional de qualidade homogênea deve ser acessível a todos os indivíduos, independentemente do seu conjunto de oportunidades. Além disso, a sociedade deve avalizar a supremacia do mérito, de forma que o talento e as habilidades individuais prevaleçam, pois sem essas garantias, os retornos dos investimentos educacionais são insatisfatórios, especialmente para aqueles mais vulneráveis[4] da sociedade. Dessa forma, a persistência educacional irá prevalecer, e, consequentemente, a mobilidade intergeracional da renda e educação é reduzida.

Em outras palavras, a educação é vista como elemento-chave para o desenvolvimento econômico e social de qualquer sociedade. Tanto no que se refere à educação como característica essencial para o bom funcionamento da democracia, permitindo às pessoas exercer plenamente os seus direitos e as suas responsabilidades como cidadãos, quanto o que ela propicia a uma qualificação da força de trabalho, elevando a produtividade das atividades econômicas.

Entretanto, o relatório da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco), versão 2005, enfatiza que, apesar do crescente consenso em torno da importância da qualidade educacional, o nível de ações que se referem a esse conceito é bem menor na realidade. Segundo o relatório, dois princípios distinguem geralmente as formas de definir a qualidade da educação: o primeiro estabelece o desenvolvimento cognitivo dos alunos como principal objetivo explícito de qualquer sistema educacional, reconhecendo o sucesso alcançado por estes últimos como

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

4. Indivíduos com *background* familiar e social baixo. Por exemplo, indivíduos com nível de renda e grau educacional familiar baixos.

indicativo de sua qualidade; e o segundo destaca o papel da educação na promoção de valores compartilhados e no desenvolvimento criativo e emocional (Unesco, 2005). Nesse caso, a consecução desses objetivos é bem mais complexa de analisar.

Focando esses conceitos, recentemente cientistas sociais têm estudado a transmissão intergeracional da educação. Por exemplo, Solon (1999), Ferreira e Veloso (2003), Heineck e Riphahn (2009), Niknami *et al.* (2010), Black e Devereux (2011). Entretanto, os estudos ainda são bastante esparsos e inconclusivos. Além disso, a maior parte utiliza nível educacional – anos de estudo – ao invés de desempenho educacional. Os resultados encontrados nesses estudos sugerem baixa transmissão educacional, em especial nos países em desenvolvimento. Esses achados reforçam os elevados níveis de persistência intergeracional da educação observados e contribuem para a manutenção da desigualdade de oportunidades educacionais.

Contudo, podem-se considerar diversos fatores responsáveis pela baixa transmissão educacional, como supor baixa qualidade educacional nesses países, pois, embora nos últimos anos tenha crescido o grau de escolaridade nessas economias, tal fenômeno não é condição suficiente para equivalência de oportunidades, uma vez que anos de estudo não expressam a qualidade educacional de um sistema, essencialmente devido à heterogeneidade existente entre as escolas em suas respectivas regiões, países e continentes.[5]

Diante disso, compreender os mecanismos subjacentes por trás dessa relação – transmissão, persistência e mobilidade intergeracional da educação – é de fundamental importância para analisarmos esse processo. Nesse sentido, o presente estudo pretende investigar duas questões fundamentais. Primeiro, o processo de transmissão intergeracional da educação no Programa Internacional de Avaliação de Estudantes (Pisa) associado à teoria de igualdade de oportunidades educacionais.[6] Para tanto, emprega-se a função de produção educacional proposta por Hanushek (1970; 1979), com algumas peculiaridades.[7] Segundo, construir uma análise contrafatual proveniente de um aumento no nível socioeconômico parental dos estudantes que prestaram o exame do Pisa 2012. Para isso, aplica-se a metodologia de inferências contrafactuais desenvolvida por Chernozhukov, Fernández-Val e Melly (2013).

Importantes diferenciais destacam-se neste estudo. Um deles, por exemplo, é verificar o padrão de transmissão intergeracional da educação entre as economias participantes do Pisa 2012. Outro diz respeito a utilizar dados de uma avaliação educacional como variável de resultado, o que permite captar mais claramente a habilidade educacional apresentada pelo estudante. E, por fim, simular o efeito contrafatual de possíveis políticas públicas educacionais.

5. Por exemplo, Ferreira e Gignoux, (2008).
6. Por exemplo, Ferreira e Gignoux (2011).
7. Foram empregadas diversas variáveis de controle. Em especial, atributos que expressam o esforço individual.

Posto isso, o restante do trabalho está dividido em cinco seções, incluindo esta introdução. Na segunda seção é apresentada uma síntese conceitual sobre transmissão, mobilidade e persistência intergeracional da educação. Em seguida, são mostrados alguns resultados contundentes do problema em questão. A terceira seção é reservada a detalhar a metodologia utilizada, incluindo a descrição e a análise dos dados. A quarta é destinada à apresentação dos resultados. E a última seção é dedicada a algumas considerações extraídas dos resultados obtidos.

2 REVISÃO DA LITERATURA

Esta seção é formada por duas subseções. Apresenta-se, na primeira, um quadro conceitual sobre transmissão, mobilidade e persistência intergeracional da educação. Na segunda, serão reportados alguns importantes resultados da literatura em questão.

2.1 Transmissão, mobilidade e persistência intergeracional da educação

Seguindo a tradicional literatura sobre mobilidade educacional, o papel da educação dos pais na educação dos filhos é expresso por:

$$Y_i^F = \alpha + \beta Y_i^P + \gamma X_i^F + \varepsilon_i, \tag{1}$$

onde Y_i^F e Y_i^P relatam, respectivamente, o resultado educacional dos filhos e dos pais. Por sua vez, X_i^F representa um vetor de características familiares do filho. Por exemplo: idade, número de irmãos, região de nascimento, se o indivíduo é migrante ou não, gênero, entre outros atributos individuais. Por fim, ε_i denota o termo de erro, de maneira que o $\hat{\beta}$ estimado é o foco central:

$$\hat{\beta} = \rho_{FP} \frac{\sigma_F}{\sigma_P}, \tag{2}$$

sendo ρ_{FP} o coeficiente de correlação entre a educação dos filhos e dos pais. E, respectivamente, σ_F e σ_P representam os desvios-padrão do termo de erro referente à educação dos filhos e dos pais. Assim, um incremento (redução) em $\hat{\beta}$ pode ser decorrente de um incremento (redução) na mobilidade intergeracional da educação ρ_{FP} ou na razão observada entre os desvios educacionais existentes em cada grupo (filhos e pais) $\frac{\sigma_F}{\sigma_P}$. Portanto, $\hat{\beta}$ fornece um grau absoluto da transmissão intergeracional da educação, dado que não explica uma mudança na variância da educação entre as gerações.

Associando os conceitos *transmissão intergeracional da educação* e *função de produção de educação*, doravante FPE, propostos por Hanushek (1970; 1979), com algumas especificidades, tem-se:

$$ln(Y_i^F) = g(F_i^t, P_i^t, I_i S_i^t), \quad (3)$$

em que Y_i^F é o vetor educacional realizado pelo estudante *i*; F_i^t é o vetor de características individuais e familiares do estudante acumulado no período *t*; P_i^t é o vetor de influências do corpo discente (influências de pares), ou seja, variáveis socioeconômicas e *background* familiar de outros estudantes na escola acumulados no período *t*; I_i é o vetor de dotações iniciais do *i*-ésimo indivíduo; e S_i^t é o vetor de insumos escolares relevantes para o estudante acumulados no período *t*. Na verdade, a FPE analisa a forma como os diversos insumos do processo educacional podem afetar os resultados educacionais dos indivíduos. Acrescentando variáveis de controle que expressem as circunstâncias e os esforços individuais, torna-se idêntica à equação (1):

$$ln\,(Y_i^F) = \alpha + \beta log\,(Y_i^P) + \gamma X_i^F + \varepsilon_i, \quad (4)$$

onde $ln\,(Y_i^F)$ representa o *log* do desempenho educacional do filho da família *i*.

E o termo de erro estocástico é expresso na equação (5):

$$E(\varepsilon_i) = 0, E(\varepsilon_i Y_i^P) = 0 \; e \; E(\varepsilon_i^2) = \sigma_\varepsilon^2. \quad (5)$$

De acordo com Ferreira e Veloso (2003), β mensura o nível de persistência intergeracional da educação. Isto é, o valor β representa a elasticidade intergeracional da educação. Por exemplo, caso o $\beta = 0{,}6$ então o filho de um pai cujo nível educacional exceda em três anos a média – educação dos pais – terá uma expectativa educacional de 1,8 ano acima da média. Nesse contexto, $1 - \beta_1$ fornece o grau de regressão à média, denominado medida de mobilidade intergeracional da educação. Há outros métodos utilizados para mensurar a mobilidade intergeracional. Mais detalhes em Solo (1999), Pastore e Silva (1999), Black e Devereux (2011).

Em geral, trabalhos destinados a explicar mobilidade, persistência e transmissão intergeracional da educação utilizam como nível educacional os anos de estudo individual como variável de resultado. Entretanto, alguns estudos têm sinalizado que, não necessariamente, anos de estudo refletem o conhecimento educacional obtido, especialmente quando se pensa em qualidade da educação.

Diante disso, este estudo propõe como alternativa utilizar uma medida de desempenho educacional – nota do exame Pisa 2012 – como variável de resultado educacional. Essa caracterização parece ser uma estratégia bem mais adequada para apurar tais fenômenos, pois o desempenho educacional em uma avaliação de conhecimentos gerais certamente reflete melhor o grau de educação alcançado, isto é, o conhecimento adquirido. Ainda mais quando se considera a habilidade observada para solucionar problemas. Porém, é importante destacar uma importante

limitação nesse caso, pois é prudente admitir que possa haver falta de interesse na resolução da avaliação por alguns indivíduos, uma vez que a avaliação Pisa não oferece punição aos envolvidos (escola, alunos, professores, entre outros).

2.2 Resultados importantes da literatura

Inicia-se destacando o estudo de Black e Devereux (2011), o qual investiga quais fatores são essenciais para que pais com níveis educacionais mais elevados tenham crianças mais educadas. Segundo os resultados obtidos, há várias explicações possíveis; entretanto, duas são preponderantes. A primeira retrata a seleção pura ou os efeitos indiretos, em que pais mais educados ganham salários mais elevados, e, portanto, oferecem demandas importantes que irão contribuir substancialmente para a educação dos filhos. Por exemplo, matricular os filhos nas melhores escolas, comprar os melhores livros, assim como o investimento em mecanismos que auxiliam o aprendizado. A segunda característica é conhecida como efeito causalidade ou direto. Nesse caso, pressupõe-se que níveis educacionais mais elevados fazem do indivíduo um tipo de mãe/pai especial, e, assim, induzem seus filhos a resultados educacionais mais elevados. Essa relação direta de causalidade ocorre por meio de fatores potencialmente não observáveis, tais como ambiente compartilhado e herdabilidade genética.

Nessa conjectura, Roemer (1998) argumenta que dois conceitos de igualdade de oportunidades prevalecem nas democracias ocidentais: *i)* todo indivíduo com potencial relevante no seu período de formação deve ser admitido como possível candidato a competir por posições na sociedade; *ii)* denominado princípio não discriminatório, estabelece que na competição por posições na sociedade, todo indivíduo que possui atributos relevantes para exercer uma determinada função deve ser incluído como candidato elegível e que somente será julgado por características relevantes.

Guiados por esta linha teórica, Lefranc, Pistolesi e Trannoy (2009) defendem que a igualdade de oportunidades constitui um princípio básico para atenuar desigualdades entre indivíduos, de modo que os resultados dependem de um conjunto de fatores determinísticos e aleatórios relevantes para o sucesso ou o fracasso do agente econômico. Nesse processo, apenas será possível identificar as variáveis que estão sob a responsabilidade dos indivíduos e as que não estão.

Todavia, destaca-se a existência de diversos estudos sobre os determinantes no desempenho educacional, inclusive utilizando dados de logro educacional; porém, a maior parte destes não passa de meros registros descritivos. Por exemplo, em Aguirreche (2012) investiga-se o modo como o grau de desigualdade de oportunidades de uma nação afeta o desempenho médio dos seus estudantes. Para tanto, é utilizado um modelo estrutural baseado em Fleurbaey e Schokkaert (2009).

Os resultados sinalizam elevado grau de desigualdade de oportunidades. Isto é, mais de 30% de desigualdade injusta. Além disso, apura-se uma relação negativa de -0,69 entre desigualdade de oportunidades e desempenho educacional.

Em outro estudo, envolvendo dados do Pisa 2006-2009, Gamboa e Waltenberg (2012) avaliam a desigualdade de oportunidades educacionais em seis países da América Latina. Nesse ensejo, adotam uma abordagem não paramétrica proposta por Checchi e Peragine (2010), a fim de decompor um índice de desigualdade. Os resultados indicam um grau de desigualdade de oportunidade educacional variando entre 1% e 25%. Esse resultado denota uma considerável heterogeneidade entre os seis países.

De modo semelhante, Ferreira e Gignoux (2008), empregando dados do Pisa 2000, calculam a parcela da desigualdade observada no desempenho escolar, que é determinada pelo conjunto de oportunidades. Em suma, constata-se que os maiores níveis de desigualdade de oportunidades educacionais são registrados em países em desenvolvimento – América Latina –, embora haja considerável heterogeneidade entre os países da Organização para a Cooperação e Desenvolvimento Econômico (OCDE). O peso das variáveis de circunstância, na desigualdade educacional total, varia de 9% a 30% nos conteúdos de matemática e de 14% a 33% nos referentes à linguagem.

Em outro estudo, Ferreira e Gignoux (2011) propõem dois métodos de inferir a desigualdade educacional. O primeiro é voltado para o desempenho educacional – variância ou desvio-padrão –, enquanto o segundo para a oportunidade educacional – parcela da variância que explica a influência das circunstâncias. Os resultados reportam que 35% de todas as disparidades no desempenho educacional dos 57 países participantes do Pisa 2006 ocorrem em função da desigualdade de oportunidades.

Por sua vez, em Heineck e Riphahn (2009), foram utilizados dados de corte de nascimento – na Alemanha, entre 1929 e 1978 – a fim de investigar a influência de atributos parentais sobre os resultados educacionais – mobilidade intergeracional da educação – obtidos ao longo do tempo. O estudo informa que, embora o sistema educacional na Alemanha tenha passado por diversas reformas objetivando reduzir a desigualdade de oportunidades educacionais, não há nenhuma evidência empírica da redução do papel socioeconômico parental no resultado educacional obtido pelos filhos nas últimas décadas.

Em linhas gerais, verifica-se que o conjunto de oportunidades é decisivo para o desempenho educacional dos indivíduos. Tais aspectos são mais incisivos em países em desenvolvimento, em especial em países da América do Sul, da Europa Oriental e da Ásia. Ao contrário, os menores índices de desigualdade de oportunidades são constatados, em grande maioria, em países da América do Norte, da Europa Ocidental e da Oceania, nos quais a educação dos pais é a característica de circunstância mais importante no resultado obtido.

3 ESTRATÉGIA EMPÍRICA

A priori, estimamos as equações (2) e (4) a fim de obter a transmissão intergeracional da educação. Em seguida, pretende-se apurar em que medida um incremento no *background* social – circunstâncias – afeta o desempenho na avaliação, visto que tais fatores mostram-se decisivos nos resultados individuais.[8] Para tanto, emprega-se a metodologia de inferências em distribuições contrafatuais desenvolvida por Chernozhukov, Fernández-Val e Melly (2013), doravante denominado IDC. A escolha pelo IDC deve-se em função de ele basear-se em diversas abordagens principais para estimar funções quantílicas condicionais e funções de distribuição condicional. Outra vantagem é seu emprego tanto para analisar o efeito de simples intervenções – mudança unitária em uma característica – quanto também em alterações complexas – mudanças gerais na distribuição das características.

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

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

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

$$Y^0 = Q_W(U^0|X^0), \text{ onde } U^0 \sim U(1,0), \text{ independentemente de } X^0 \sim F_X^0, \quad (6)$$

em que (5) destaca que o resultado é função das covariadas e do termo de erro U^0. Sobretudo, em modelos de regressões clássicos, o termo de erro é separado das variáveis independentes, como nos modelos de regressão pontual, mas, em geral, não é necessário. Este método abrange ambos os casos.

8. Ver, por exemplo, Barros (2009).

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

$$Y^C = Q_W(U^C|X^C), \tag{7}$$

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

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

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

Este tipo de intervenção pode ser utilizado para estimar, por exemplo, qual o efeito sobre as despesas de alimentos resultantes de uma alteração no Imposto de Renda. Ou qual a influência sobre os preços dos imóveis resultantes da limpeza de resíduos perigosos no entorno. Além do mais, para fazer inferências do efeito total no resultado decorrente da intervenção, é preciso identificar as funções distribuição e quantílica do resultado antes e após a política. A função distribuição condicional associada à função quantílica $Q_Y(u|x)$ é expressa por:

$$F_Y(y|x) = \int_0^1 1\{Q_Y(u|x) \leq y\}\, du. \tag{8}$$

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

$$F_{Y^j}(y) := Pr\{Y^j \leq y\} = \int_x F_Y(y|x) dF_X^j(x). \quad (9)$$

Com funções quantil marginal correspondente:

$$Q_{Y^j}(u) = inf\{y: F_{Y^j}(Y) \geq u\}, \quad (10)$$

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

$$QTE_Y(u) = Q_{Y_C}(u) - Q_{Y_0}(u). \quad (11)$$

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

$$DE_Y(y) = F_{Y_C}(y) - F_{Y_0}(y). \quad (12)$$

Por fim, o IDC permite uma decomposição semelhante à proposta por Juhn, Murphy e Pierce (1993) e baseada em trabalho de DiNardo, Fortin e Lemieux (1996), ressalvadas algumas peculiaridades. A principal vantagem do IDC é o emprego de regressão quantílica contrafatual em alternativa a mínimos quadrados ordinários (MQO), uma vez que a decomposição quantílica mensura o diferencial entre grupos, assumindo que tais diferenças variam ao longo da distribuição.

Em suma, o IDC apura o diferencial total entre os grupos – controle e tratamento – e decompõe em três partes. A primeira representa o efeito das características individuais observáveis dos estudantes, ou seja, refere-se a todas as covariadas – fatores familiares e individuais – da análise. A segunda componente representa o efeito médio dos coeficientes de toda a distribuição. Isto é, os coeficientes médios associados a cada uma das covariadas. Por fim, a última componente está associada às características não observadas; em outras palavras, representa o efeito residual dos fatores não observados que influenciam o desempenho.

Em termos práticos, para apurar o diferencial entre os grupos analisados, a priori, exige-se encontrar quantis incondicionais com a finalidade de desenhar a distribuição contrafatual que retrata exatamente o grupo de controle, caso ele tivesse passado pelo tratamento em questão. Em síntese, o IDC adota um procedimento em quatro etapas:

a) estima a função inversa condicional referente a cada quantil;

b) obtém a função de distribuição incondicional;[9]

9. Este procedimento não é necessariamente trivial, dado que a função em questão não obrigatoriamente é monotônica, e, portanto, não pode ser invertida usualmente. Mais detalhes em Melly (2006).

c) integra a função de distribuição condicional em relação às covariadas;

d) encontra a função de distribuição inversa incondicional.

De posse da função de distribuição inversa incondicional é possível estimar cada quartil contrafatual e obter a decomposição do tratamento em relação aos quartis dos grupos de tratamento e controle estimados.

De acordo com Azevedo e Foguel (2007), uma das maiores vantagens obtidas por essa decomposição é permitir um diagnóstico amplo da análise, pois se encontra a direção, a dimensão e o *timing* das forças que influenciam as variações dos resultados individuais obtidos.

3.1 Descrição e análise dos dados

Os resultados apresentados nesta pesquisa serão possibilitados pela utilização dos microdados do Pisa, versão 2012 sendo a quinta edição do programa, e avalia as competências dos alunos em leitura, matemática e ciências com ênfase em matemática. O exame Pisa é uma pesquisa internacional trienal que tem como objetivo avaliar os sistemas de educação em todo o mundo. Para tanto, busca testar as habilidades e o conhecimento de estudantes de 15 anos de idade. O programa é desenvolvido e coordenado pela OCDE, composto por uma coordenação nacional em cada país participante.

A adoção desta base ocorre em virtude da riqueza apresentada, em que cerca de 510 mil estudantes de 65 economias participaram desta edição, representando algo em torno de 28 milhões de jovens a nível mundial. Além disso, ele é considerado único, por desenvolver atividades que não estão inteiramente ligadas ao currículo escolar. Na verdade, as provas são projetadas para avaliar em que medida os alunos, no final do ensino fundamental, podem aplicar seu conhecimento a situações cotidianas da vida real, e como estão capacitados para a plena participação na sociedade. Além disso, as informações coletadas por meio de questionários fornecem elementos auxiliares na interpretação de resultados.

Já em relação às análises preliminares dos dados reportados na tabela 1, verifica-se um elevado grau de desigualdade no desempenho médio em todas as áreas do conhecimento, especialmente em relação a países da América do Sul.

Por exemplo, o Brasil, que, apesar de possuir uma das dez maiores economias do planeta, segundo o Fundo Monetário Internacional (FMI), apresenta um resultado extremamente modesto. No entanto, destaca-se que os resultados não se modificam significativamente em relação ao método e aos relatórios apresentados pelo programa, pois, de acordo com o Pisa, o Brasil ocupa a 58ª posição no *ranking* dos 65 países analisados.

TABELA 1
Desempenho médio: avaliação Pisa 2012

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

Fonte: Pisa 2012.
Elaboração dos autores.

Por outro lado, cabe destacar as excelentes *performances* da maioria dos países asiáticos que obtiveram médias acima dos países da OCDE. Para se ter uma ideia, conforme denotam os resultados apresentados, os cinco melhores resultados expostos pertencem a esse grupo. Além do mais, impressiona a diferença de desempenho no exame de matemática, que é o foco do Pisa 2012. Especificamente, Xangai obteve em torno de 90 pontos acima do Japão, considerado na análise membro da OCDE.

Em relação aos países da América do Sul, os asiáticos obtêm um desempenho aproximadamente 30% superior. Por exemplo, comparando o Chile – melhor desempenho entre os sul-americanos –, Singapura apresenta um diferencial positivo próximo a 28% em favor dos asiáticos. Embora nesses últimos haja uma maior variância nas notas médias, este fato ocorre em virtude do baixo desempenho apresentado notadamente por Malásia e Indonésia.

Posto isto, ressalta-se que os microdados utilizados são divididos em: *i)* variáveis relacionadas aos estudantes; *ii)* variáveis relacionadas aos pais (características dos indivíduos e de sua família); *iii)* variáveis relacionadas às escolas (características específicas). A junção destas bases possibilita compilar informações a respeito das variáveis de circunstâncias, esforços e resultados individuais.

Entretanto, cabe ressaltar que foram realizados quatro testes de especificação do modelo com a finalidade de elencar o melhor grupo de covariadas.[10] Os testes realizados foram: *ovtest, stepwise backward, forward* e *hierarchical*. O *ovtest* objetiva detectar problemas de variáveis omitidas. Os outros três testes servem para identificar a inclusão ou remoção de variáveis no modelo.[11] Além da observância dos resultados dos testes, procurou-se optar também por variáveis clássicas utilizadas na literatura.[12] Guiado por isto, na primeira abordagem regrediu-se o *log* da média individual no exame Pisa em função do conjunto de variáveis circunstanciais descritas a seguir. Na segunda abordagem, acrescentam-se variáveis que expressam o esforço individual.

Especificamente, as variáveis utilizadas neste estudo são sumarizadas no quadro 1. Conforme reportado anteriormente, no início a abordagem utiliza somente o vetor de variáveis circunstanciais e corresponde a informações de 336.286 estudantes, de um total 422.413, de 58 países.

Em um segundo momento, emprega-se nas estimativas o vetor de variáveis que expressam o esforço, em que são utilizados dados de 225.629 estudantes de 58 países. A razão pela qual os dados utilizados não correspondem ao total disponível, bem como a exclusão de alguns países, ocorre em função de informações incompletas. Além disso, existe uma consensual limitação na natureza dos dados empregados, pois o resultado obtido no Pisa não origina qualquer consequência direta para alunos, professores, diretores, escolas, entre outros. Portanto, é possível que a habilidade dos alunos esteja sendo subestimada.[13]

QUADRO 1
Descrição das variáveis

Variáveis	Descrição
Resultados individuais	*Log* da média aritmética do resultado: linguagens; matemática e ciências.
Variáveis de circunstâncias	
Educação parental	*Log* do maior grau de instrução parental – anos de estudo – do pai ou da mãe.
Diferença Educacional	Diferença, em anos de estudo, do pai e da mãe.
Tipo de escola	*Dummy* para escola. 0 para escola pública e 1 para privada.
Sexo	*Dummy* para gênero. 0 para mulheres e 1 para homens.
Atendimento pré-escolar	*Dummy* para pré-escolar. 0 para não atendidos e 1 para atendidos.
Localização da escola	*Dummy* localização da escola. 0 para zona rural e 1 para urbana.
Estrutura familiar	*Dummy* estrutura familiar. 0 para famílias monoparental e 1 para biparental.
Siblings – irmãos	*Dummy* irmãos. 0 caso o indivíduo tenha irmãos e 1 caso contrário.

(Continua)

10. Destaca-se que, embora o *ovtest* tenha indicado a existência de variáveis omitidas – admitindo, por exemplo, habilidade e motivação –, ele também indica o modelo *log* linearizado como melhor.
11. Mais detalhes em Chatterjee e Hadi (2015).
12. Por exemplo, Firmo (2008) e Bauer e Riphahn (2007).
13. Mais detalhes em Quintano, Castellano e Longobardi (2008).

(Continuação)

Variáveis	Descrição
Variáveis de esforço individual	
Repetência de ano	*Dummy* repetência. 0 caso o aluno já tenha repetido o ano e 1 caso contrário.
Migrante[1]	*Dummy* migração. 0 caso o aluno seja migrante e 1 caso seja nativo.
Perseverança	Índice baseado na disponibilidade do aluno de resolver problemas difíceis.
Esforço real	Índice de esforço de autorrelato.
Esforço potencial	Taxa de não resposta da avaliação.

Fonte: Pisa 2012.
Elaboração dos autores.
Nota: [1] A razão por considerar a variável migrante como esforço e não circunstância, em especial, ocorre necessariamente por ser uma escolha. Contudo, é possível argumentar que um jovem de 15 anos não migra por vontade própria. Entretanto, ser migrante ou não migrante é uma característica controlável, e, portanto, não é imutável. Ao contrário, por exemplo, de nascer branco, ao qual foge totalmente do controle individual.

Dito isto, faz-se agora um breve relato a respeito da amostra do Pisa. Os dados denotam que praticamente 50% dos que prestaram o exame em 2012 são do sexo feminino; 80% são oriundos de escola pública; 86% frequentaram no mínimo um ano de pré-escolar; 61,5% são de países-membros da OCDE; e 8,9% da América do Sul.

TABELA 2
Análise descritiva
(Em %)

Sexo		Membros – exceto outros		
Masculino	Feminino	OCDE	América do Sul	Asiáticos
49,42	50,48	61,50	8,90	9,12
Nível educacional dos pais – pai/mãe				
Primária	Básica	Secundária	Superior	
3,42	21,28	32,66	35,94	
3,89	21,42	33,48	37,41	
Tipo de escola		Atendimento pré-escolar		
Pública	Privada	Atendido	Não atendido	
80,58	19,42	86,90	13,10	

Fonte: Pisa 2012.
Obs.: Porcentagens diferentes de 100% (*missing*).

Além do mais, constata-se que em torno de 37% dos pais possuem no mínimo educação superior, e apenas um número próximo de 4% dos alunos é composto por filhos de indivíduos que, no máximo, concluíram o primeiro ciclo do ensino fundamental.

4 RESULTADOS

Nesta seção serão apresentados os resultados do estudo. Primeiramente, na subseção 4.1, apresentam-se os fatores determinantes do resultado educacional – desigualdade de oportunidades educacionais –, individual, bem como a medida absoluta da transmissão intergeracional da educação. Por fim, na subseção 4.2 infere-se o efeito-tratamento de um incremento socioeducacional parental.

4.1 Primeiro bloco

A priori, a tabela 3 reproduz as estimativas da equação (4)[14] para a amostra geral. Como esperado, os resultados sugerem elevado grau de correlação entre as circunstâncias e os resultados individuais observados. Isto é, há evidências que permitem supor acentuada participação do conjunto de oportunidades nas disparidades existentes no resultado do exame.

TABELA 3
Determinantes do resultado individual: média geral

Variáveis	OLS	QREG
Educação parental	0,1449***	0,1610***
Dif. educação	-0,0024***	-0,0021***
Tipo escola	0,0468***	0,0477***
Sexo	-0,0172***	-0,0121***
Pré-escolar	0,1150***	0,1181***
Local escola	0,0440***	0,0416***
Estrutura familiar	0,0543***	0,0551***
Irmãos	0,0226***	0,0261***
Intercepto	5,6000***	5,5704***
N	336.286	336.286
Adj. R^2	0,138	-

Fonte: Pisa 2012.
Elaboração dos autores.
Obs.: *, ** e *** = p<0,10, p<0,05 e p<0,01, respectivamente.

Embora os coeficientes apresentados na tabela 3 não possam ser tratados como causalidade, ou seja, como fatores determinantes do desempenho alcançado pelo estudante, eles vão de encontro aos principais resultados da literatura. Por exemplo, em Ferreira e Gignoux (2011), o *background* social – circunstâncias – tem papel crucial no desempenho educacional individual. Outra importante observação refere-se à relação negativa evidenciada entre a diferença educacional dos pais e o

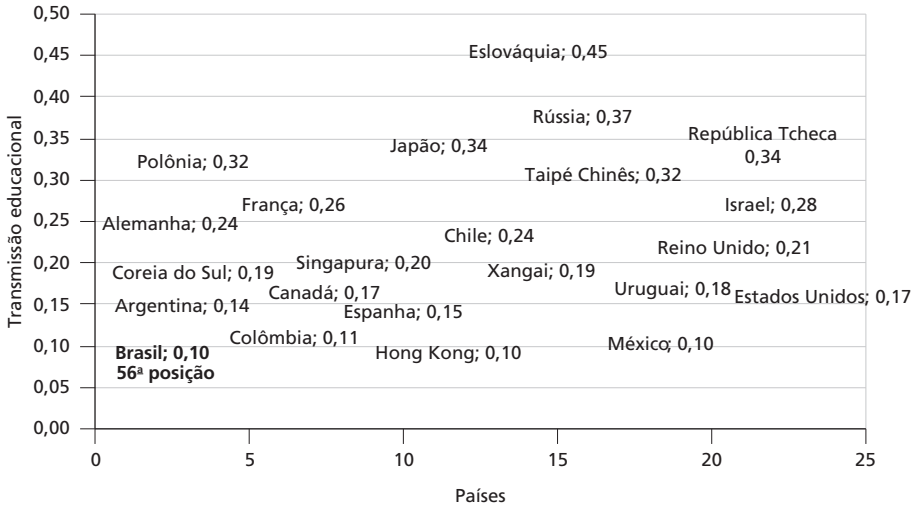
14. Estimou-se a equação (4) por mínimos quadrados ordinários (OLS), regressão quantílica (QREG) (para o quantil 0,50). Os resultados mantiveram-se para média em ciências, línguas e matemática.

resultado obtido. Isto é, quanto maior a assimetria – disparidade – educacional entre pai e mãe, menor a nota do estudante. Este resultado também é coerente com a literatura. Mais detalhes em Bourguignon, Ferreira e Menéndez (2007).

Entretanto, há uma limitação nos achados da tabela 3, pois tais resultados referem-se a diversos países e continentes, e, portanto, deve-se levar em conta a heterogeneidade existente, pois sistemas e políticas públicas educacionais existentes em cada país originam efeitos bem peculiares. Uma possível solução para o problema seria incluir uma variável instrumental nas estimações. Entretanto, encontrar um instrumento capaz de superar tal problema não é uma tarefa trivial.

Por sua vez, de acordo com o exposto no gráfico 1 e na tabela A.2 do apêndice, obtidos a partir da equação (2), verifica-se que o Brasil ocupa a 56ª posição entre os 58 países analisados em termos de transferência intergeracional de educação. Para se ter uma ideia, Argentina e Brasil, juntos, assinalam uma influência educacional – transmissão – parental 40% inferior à observada em relação a japoneses e 87,5% menor que os eslovacos.

GRÁFICO 1
Transmissão intergeracional da educação: Pisa 2012 – média geral



Fonte: Pisa 2012.
Elaboração dos autores.

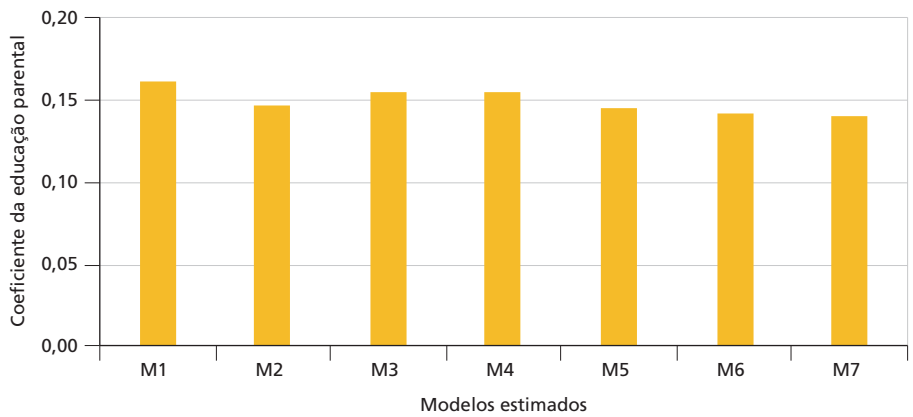
Na América do Sul, especificamente, os estudantes brasileiros foram os que apresentaram a menor transmissão entre as gerações – pai e filho –, ficando, em geral, à frente apenas da Tunísia, da Indonésia e de Macau. Corroborando a direção desses resultados, Black e Devereux (2011) assumem que a escolha educacional dos jovens é fruto de alguns fatores, em especial custos educacionais, retornos da educação e renda familiar. Essa última, especialmente nos casos em que há restrições de acesso a crédito.

Todavia, a literatura mostra que, em geral, os retornos educacionais são maiores para indivíduos mais habilidosos e também para aqueles que têm pais mais educados. Estas hipóteses demonstram que indivíduos que têm pais mais educados tendem a buscar maior nível educacional devido a dois efeitos: *i)* efeitos diretos – pais mais educados –, também interpretados como canal causal; e *ii)* efeitos indiretos, ou seja, possuir maior habilidade – herdabilidade – evidenciada pela transmissão intergeracional da educação.

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

Outras questões interessantes podem ser visualizadas na tabela A.1. Por exemplo, atributos como tipo de escola, atendimento pré-escolar e localização da escola exibem forte correlação com o desempenho educacional obtido na avaliação, especialmente nas economias em desenvolvimento. Nos países desenvolvidos, esses fatores parecem afetar bem menos os resultados. Nomeadamente, na Argentina e no Brasil, o simples fato de se estudar em escola privada pode estar influenciando a nota média do Pisa em torno de 14% e 17%, respectivamente. No entanto, em países como Alemanha e Japão, os coeficientes de correlação mostram-se bem insignificantes.

GRÁFICO 2
Influência educacional parental



Fonte: Pisa 2012.
Elaboração dos autores.

Neste rol de investigação, acrescentaram-se nas estimativas algumas variáveis que expressam o esforço individual, embora, a princípio, ainda se assume a hipótese postulada na equação (4). Entretanto, antes de discutirmos os resultados reproduzidos na tabela 4, o gráfico 2 mostra sete especificações distintas que consideram *dummies* continentais. Claramente, os resultados estimados são bem semelhantes.[15] Todavia, os testes de especificação informam que o modelo geral é mais ajustado sem as *dummies* continentais.

TABELA 4
Determinantes do resultado individual: média geral

Variáveis	OLS	QREG
Educação parental	0,1361***	0,1489***
Dif. educação	-0,0012***	-0,0012***
Tipo de escola	0,0503***	0,0515***
Sexo	-0,0056***	-0,0019**
Pré-escolar	0,0996***	0,1029***
Local escolar	0,0430***	0,0391***
Migrante	0,0204***	0,0175***
Repetente	0,1124***	0,1192***
Perseverança	0,0160***	0,0145***
Esforço real	0,0038***	0,0047***
Esforço potencial	-0,0009***	-0,0010***
Intercepto	5,5540***	5,5228***
N	225.629	225.629
Adj. R^2	0,190	-

Fonte: Pisa 2012.
Elaboração dos autores.
Obs.: *, ** e *** = p<0,10, p<0,05 e p<0,01, respectivamente.

Nitidamente, os coeficientes estimados – efeitos das covariadas –, apesar de menores, situam-se bem próximos, sugerindo, assim, o bom ajuste do modelo. Ademais, os testes de identificação anteriormente citados ratificam estes resultados.

Contudo, dois fatos merecem destaque entre as covariadas que representam o esforço individual nos resultados. Primeiro, o grau de correlação existente entre o desempenho individual observado e o fato de o aluno já ter repetido o ano. Segundo, o baixo efeito atribuído ao esforço real e ao esforço potencial, com destaque especial para o sinal deste último fator. Esse resultado evidencia o problema de limitação

15. M1: não há *dummies* continentais; M2: *dummy* para OCDE; M3: igual a M2 mais *dummy* (países asiáticos); M4: M3 acrescido da *dummy* (países nórdicos); M5: modelo geral incluindo o vetor de esforços individuais. Nos demais modelos acrescentam-se, uma a uma, as *dummies* continentais.

em dados desta natureza, uma vez que o desempenho na prova não exerce nenhum reflexo na vida de alunos, professores e escolas. Nesse sentido, observa-se elevada taxa de não respostas, especialmente a questões mais complexas. Nessa mesma linha de raciocínio, percebe-se que o coeficiente da variável perseverança parece confirmar essas evidências.

Em relação às variáveis alusivas à escola, como tipo, localização e pré-escola, estas apresentam substancial relação positiva com o desempenho obtido na avaliação. O fator pré-escola, especificamente, aparenta ser o mais importante entre elas. Esse fato torna-se ainda mais expressivo quando consideramos os resultados dos países asiáticos, nos quais foram apuradas as melhores médias do Pisa 2012. Isto é, em média, nessas economias, a relação obtida entre desempenho educacional e pré-escola é quase quatro vezes maior do que o observado nos países da OCDE e países sul-americanos, e, aproximadamente, três vezes mais do que nos países nórdicos. Vale ressaltar que, embora os resultados gerais referentes ao tipo de escola – pública ou privada – informam uma menor influência desta no resultado do exame, quando avaliamos os resultados dos países da América do Sul, verifica-se que a influência chega a ser cinco vezes mais efetiva do que nos países-membros da OCDE, aproximadamente sete vezes ao efeito atribuído aos asiáticos e infinitamente superior aos países nórdicos.

4.2 Segundo bloco

Com o intuito de propiciar respostas para questões como "quais fatores estão associados a essa desigualdade de oportunidades educacionais?", ou "o que poderia ser feito para erradicar ou amenizar a desigualdade de oportunidades educacionais observada?", esta subseção estimará o efeito-tratamento, expresso pela equação (11). Primeiramente, serão abordados os resultados da estimativa do *log* da média geral das notas individuais em função de uma *dummy* para educação do pai e da mãe, como também para estrutura familiar, tipo de escola, atendimento pré-escolar e localização da escola.

O efeito de tratamento médio quantílico (ETMQ) apurado para todas as variáveis de circunstâncias – exceto sexo e irmãos – está presente na tabela 5. Os resultados mostram um *gap* médio para educação parental em torno de 8%, em que cerca de 2% é explicado por diferenças no nível educacional dos pais, e os outros 6% são devido a diferenças nos coeficientes médios entre filhos de pais com, no mínimo, nível superior, e filhos de pais com, no máximo, nível médio. Quanto ao tipo de escola – pública ou privada –, a diferença média observada no resultado do Pisa situa-se em 6%, em que 1,5% deste diferencial advém de se estudar em escola privada. Destaca-se que a parte explicada pelos resíduos é negligenciável para todos os tratamentos propostos.

TABELA 5
ETMQ: circunstâncias

Desempenho individual – *log* da média geral – quantílico									
Quantil	0,10	0,50	0,90	0,10	0,50	0,90	0,10	0,50	0,90
Tratamento	Educação da mãe			Educação do pai			Estrutura familiar		
Características	0,025	0,020	0,013	0,025	0,020	0,012	0,024	0,014	0,007
Coeficientes médios	0,065	0,063	0,061	0,059	0,058	0,057	0,037	0,036	0,036
Diferencial total	0,081	0,086	0,064	0,071	0,080	0,063	0,060	0,051	0,031
Tratamento	Tipo de escola			Pré-escola			Localização da escola		
Características	0,020	0,014	0,009	0,0282	0,0290	0,023	0,0314	0,027	0,017
Coeficientes médios	0,058	0,050	0,045	0,1040	0,1069	0,110	0,0467	0,040	0,037
Diferencial total	0,084	0,065	0,032	0,1383	0,1381	0,100	0,0832	0,066	0,057

Elaboração dos autores.
Obs.: 1. Os tratamentos utilizados foram significativos.
2. *** = p<0,01.
3. As estimações estão decompostas em três componentes para cada quantil e representam, respectivamente: efeito das características individuais (observáveis); efeito dos coeficientes médios das covariadas; e diferencial total (tratamento médio). Os valores residuais foram ignorados em função de serem bastante incipientes.

Por sua vez, o tratamento médio referente à pré-escola ocorre próximo a 14%. Entretanto, apenas 21% desse efeito ocorre em função de o aluno ter frequentado a pré-escola. Contudo, na base analisada, não há informações sobre a idade de ingresso na pré-escola, o que limita nossa análise, pois, segundo Spinath *et al.* (2003), a primeira infância – 0 a 6 anos – é substancialmente importante no desenvolvimento da habilidade cognitiva geral.[16] No que se refere às demais variáveis – estrutura familiar e localização das escolas –, respectivamente, observou-se um ETMQ variando em torno de 5% e 6%, em que o efeito de ter uma família monoparental e estudar em escola localizada na zona urbana afeta o desempenho individual em torno de 1,5% a 2%, nessa ordem.

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

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

De forma análoga, observa-se que, nos países sul-americanos, o diferencial médio entre os grupos de tratamento e controle – escola privada e pública – é infinitamente maior que nos países asiáticos, e, respectivamente, duas a três vezes maior que o observado nos membros da OCDE e países nórdicos. Esse efeito no Brasil é quase sete vezes maior do que na França e vinte vezes superior ao apresentado nos Estados Unidos. Contudo, somente cerca de um terço dele é explicado pela origem escolar do aluno. Além disso, tais efeitos são bem semelhantes aos encontrados em alguns países da OCDE, asiáticos e nórdicos. Esses resultados parecem sugerir novamente baixa qualidade educacional.

Contudo, o efeito-tratamento da pré-escola é bem mais expressivo nos países que obtiveram os melhores desempenhos no exame Pisa. Por exemplo, na França e no Japão a influência dessa característica situa-se em 10%, contra 4% do que se observa na Argentina e no Brasil. Em linhas gerais, todos os resultados apontam que o baixo desempenho obtido, em especial dos países da América do Sul em relação aos demais, é determinado crucialmente pelo conjunto de oportunidades, pois olhando superficialmente para a tabela A.1, no apêndice, percebe-se claramente que a desigualdade observada, explicada pelo conjunto de características individuais – soma dos efeitos individuais –, é de aproximadamente 0,35 nesta região, contra 0,11 na OCDE, 0,15 nos países asiáticos e 0,13 nos países nórdicos. Isto é, corresponde a 91% do total verificado nos outros três continentes juntos.

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

5 CONSIDERAÇÕES FINAIS

Neste trabalho buscou-se compreender os mecanismos subjacentes que estão por trás do desempenho educacional. Elencaram-se duas metas principais: verificar o grau de transmissão e correlação intergeracional da educação; e construir uma análise contrafatual originada pela equalização de circunstâncias socioeconômicas dos estudantes que prestaram o exame do Pisa 2012. Os resultados encontrados possibilitam destacar ao menos duas importantes deduções: *i)* percebe-se uma baixa transmissão da educação parental nos países em desenvolvimento, em especial no

continente sul-americano; *ii)* o efeito-tratamento é substancialmente maior nos países com essas características. Por exemplo, o fato de ser filho de mães com maior nível educacional – tratamento educação da mãe – na América do Sul gera um ganho de desempenho 7,4 vezes superior ao observado em jovens estudantes com as mesmas características nos países da OCDE.

Ademais, olhando para os resultados reportados nas tabelas A.1 e A.3 percebe-se que os fatores associados à escola, exceto a pré-escola, revelam ser incisivos apenas nos países em desenvolvimento, especificamente nos países sul-americanos. Por outro lado, com exceção do Chile, o efeito-tratamento do atendimento pré-escolar nessas economias é geralmente inferior ao apurado em países que obtiveram os melhores desempenhos. Esses resultados sugerem que a pré-escola, na realidade, é importante; no entanto, novamente a qualidade é enfatizada (ver, por exemplo, Foguel e Veloso, 2014), pois, se olharmos para o Chile, país sul-americano com melhor desempenho no Pisa, percebemos que este é o fator escolar (associado ao ambiente socioeconômico individual) mais importante, o que reforça a criação de políticas públicas nessa perspectiva. Já para os outros países deste continente – em geral que apresentaram os piores desempenhos –, averiguamos que este fator não se apresenta com tal importância. Ou seja, diferentemente da maioria, em que se observam os melhores resultados no Pisa, a pré-escola exerce pouca influência nesses países. Isso reforça ainda mais a desconfiança sobre a qualidade da educação nesses países.

REFERÊNCIAS

AGUIRRECHE, Agurtzane Lecuona. **Inequality of opportunity in education**: empirical applications and policy. Biscay: Universidad Del Pais Vasco, 2012.

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

AZEVEDO, João Pedro; FOGUEL, Miguel Nathan. Uma decomposição da desigualdade de rendimentos do trabalho no Brasil, 1995-2005. *In:* BARROS, Ricardo Paes; FOGUEL, Miguel Nathan; ULYSSEA, Gabriel (orgs.). **Desigualdade de renda no Brasil**: uma análise da queda recente. Brasília: Ipea, 2007. v. 2.

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

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

BLACK, Sandra E.; DEVEREUX, Paul J. Recent developments in intergenerational mobility. *In:* ASHENFELTER, Orley; CARD, David (eds.). **Handbook of Labor Economics**. Amsterdam: Elsevier, 2011. v. 4.

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

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

CHATTERJEE, Samprit; HADI, Ali S. **Regression analysis by example**. New Jersey: John Wiley & Sons, 2015.

CHECCHI, Daniele; PERAGINE, Vito. Inequality of opportunity in Italy. **Journal of Economic Inequality**, v. 8, p. 429-450, 2010.

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

DAUDE, Christian. **Ascendance by descendants?** On intergenerational education mobility in Latin America. Paris: OECD Publishing, 2011.

DINARDO, John; FORTIN, Nicole M.; LEMIEUX, Thomas. Labor market institutions and the distribution of wages, 1973-1992: a semiparametric approach. **Econometrica**, v. 64, n. 5, p. 1001-1044, 1996.

FERREIRA, Francisco; GIGNOUX, Jérémie. **The measurement of inequality of opportunity**: theory and an application to Latin America. Washington: World Bank, 2008. (World Bank Policy Research Working Paper Series).

_____; _____. The measurement of inequality of opportunity: theory and an application to Latin America. **Review of Income and Wealth**, v. 57, n. 4, p. 622-657, 2011.

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

FIRMO, Marcio Gold. **Uma análise da transmissão intergeracional de capital humano no Brasil**. 2008. Tese (Doutorado) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008.

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

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

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

HANUSHEK, Eric. The production of education, teacher quality and efficiency. *In:* HANUSHEK, Eric (ed.). **Do teachers make a difference?** Washington: U.S. Office of Education, 1970.

_____. Conceptual and empirical issues in the estimation of educational production functions. **Journal of Human Resources**, v. 14, n. 3, p. 351-388, 1979.

HECKMAN, James J. A economia da desigualdade: o valor da educação infantil. **Americano Educador**, v 35, n. 1, p. 31, 2011.

HEINECK, Guido; RIPHAHN, Regina T. Intergenerational transmission of educational attainment in Germany: the last five decades. **Jahrbücher für Nationalökonomie und Statistik**, v. 229, n. 1, p. 36-60, 2009.

JUHN, Chinhui; MURPHY, Kevin M.; PIERCE, Brooks. Wage inequality and the rise in returns to skill. **Journal of Political Economy**, v. 101, n. 3, p. 410-442, 1993.

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

MELLY, Blaise. Estimation of counterfatual distributions using quantile regression. **Review of Labour Economics**, v. 68, n. 4, p. 543-572, 2006.

NIKNAMI, Susan *et al*. **Transmissão intergeracional da educação entre imigrantes mães e suas filhas na Suécia**. Estocolmo: University of Estocolmo, 2010.

PASTORE, José; SILVA, Nelson do Valle. **Mobilidade social no Brasil**. São Paulo: Makron Books, 1999

QUINTANO, Claudio; CASTELLANO, Rosalia; LONGOBARDI, Sergio. Implementation and evaluation of imputation strategies to improve the data accuracy. The case of Italian students data from the Programme for International Student Assessment (Pisa 2003). *In:* Q2008, EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS, 2008, Rome. **Annals** […]. Rome: Istat, 2008.

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

SOLON, Gary. Intergenerational mobility in the labor market. *In:* ASHENFELTER, Orley; CARD, David (eds.). **Handbook of Labor Economics**. Amsterdam: North-Holland, 1999. v. 3.

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

UNESCO – ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA. **Década das Nações Unidas da educação para o desenvolvimento sustentável (2005-2014)**. Brasília: Unesco, 2005.

BIBLIOGRAFIA COMPLEMENTAR

HERTZ, Tom *et al*. The inheritance of educational inequality: international comparisons and fifty-year trends. **The BE Journal of Economic Analysis & Policy**, v. 7, n. 2, p. 1-48, 2007.

APÊNDICE A

RESULTADOS ADICIONAIS

TABELA A.1
Efeitos das circunstâncias no Pisa: média geral

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

Fonte: Pisa 2012.
Elaboração dos autores.
Obs.: * = Não significativos a 10%.

TABELA A.2
Transmissão intergeracional da educação

Países	σ_{filhos}	σ_{pais}	$\beta_{1=\rho\frac{\sigma_{filhos}}{\sigma_{pais}}}$ Transmissão intergeracional	ρ_{fp}
República Eslovaca	0,22	0,18	0,38	0,45
Rússia	0,17	0,12	0,26	0,37
Polônia	0,16	0,17	0,38	0,36
República Tcheca	0,17	0,15	0,29	0,34
Japão	0,18	0,14	0,27	0,34
Taipei Chinês	0,19	0,20	0,33	0,32
Austrália	0,20	0,19	0,27	0,30
Bulgária	0,23	0,26	0,35	0,30
Nova Zelândia	0,20	0,18	0,27	0,30
Israel	0,22	0,21	0,27	0,28
Hungria	0,18	0,24	0,37	0,28
França	0,20	0,20	0,27	0,26
Irlanda	0,17	0,20	0,29	0,25
Lituânia	0,17	0,19	0,27	0,25
Chile	0,18	0,34	0,47	0,24
Áustria	0,18	0,20	0,27	0,24
Alemanha	0,18	0,23	0,30	0,24
Dinamarca	0,18	0,23	0,30	0,24
Noruega	0,19	0,16	0,19	0,23
Bélgica	0,20	0,24	0,27	0,23
Montenegro	0,19	0,20	0,24	0,23
Eslovênia	0,19	0,19	0,23	0,23
Finlândia	0,19	0,24	0,28	0,22
Grécia	0,19	0,25	0,28	0,21
Reino Unido	0,19	0,16	0,18	0,21
Suécia	0,21	0,22	0,21	0,20
Singapura	0,19	0,28	0,29	0,20
Letônia	0,15	0,19	0,24	0,19
Islândia	0,20	0,17	0,17	0,19
Coreia do Sul	0,17	0,18	0,20	0,19

(Continua)

(Continuação)

Países	σ_{filhos}	σ_{pais}	$\beta_{1=\rho\frac{\sigma_{filhos}}{\sigma_{pais}}}$ Transmissão intergeracional	ρ_{fp}
Xangai	0,15	0,28	0,36	0,19
Suíça	0,17	0,26	0,27	0,18
Uruguai	0,21	0,39	0,34	0,18
Tailândia	0,18	0,34	0,34	0,18
Estados Unidos	0,18	0,25	0,24	0,17
Canadá	0,17	0,19	0,19	0,17
Croácia	0,17	0,20	0,20	0,17
Luxemburgo	0,20	0,40	0,32	0,16
Qatar	0,26	0,30	0,19	0,16
Peru	0,21	0,45	0,35	0,16
Sérvia	0,19	0,22	0,19	0,16
Romênia	0,17	0,22	0,21	0,16
Holanda	0,18	0,23	0,19	0,15
Espanha	0,17	0,35	0,29	0,15
Argentina	0,21	0,42	0,29	0,14
Estônia	0,14	0,15	0,14	0,14
Portugal	0,19	0,44	0,32	0,13
Itália	0,18	0,28	0,20	0,13
Jordânia	0,21	0,27	0,15	0,12
Liechtenstein	0,16	0,24	0,17	0,11
Turquia	0,17	0,45	0,28	0,11
Colômbia	0,18	0,47	0,28	0,11
Hong Kong	0,16	0,32	0,21	0,10
México	0,16	0,43	0,28	0,10
Brasil	0,19	0,53	0,29	0,10
Tunísia	0,19	0,50	0,22	0,08
Indonésia	0,17	0,41	0,16	0,06
Macau-China	0,16	0,34	0,09	0,04

Fonte: Pisa 2012.
Elaboração dos autores.

TABELA A.3
Efeito de tratamento médio: circunstâncias – por países

Continentes/países	Educ. mãe	Educ. pai	Estrut. fam.	Tipo escola	Pré-escolar	Local escola
América do Sul	*0,098093*	*0,058824*	*0,024489*	*0,051932*	*0,056411*	*0,064881*
	0,118730	*0,115888*	*0,032333*	*0,175486*	*0,103262*	*0,153443*
Argentina	0,048386	0,044594	0,026608	0,038053	0,044710	0,075749
	0,104354	0,082832	0,031527	0,146395	0,142076	0,084823
Brasil	0,058463	0,071051	0,020118	0,054305	0,036645	0,057033
	0,098093	0,105993	0,048574	0,200697	0,076164	0,132065
Chile	0,039796	0,045090	0,021034	0,042969	0,072531	0,051453
	0,153724	0,151598	0,034362	0,165054	0,112729	0,147919
Países da OCDE	*0,013222*	*0,012544*	*0,007323*	*0,016521*	*0,041449*	*0,016772*
	0,088401	*0,083916*	*0,033916*	*0,054301*	*0,088213*	*0,047812*
Alemanha	0,026368	0,024343	-0,000638	0,033427	0,065402	-0,043289
	0,097151	0,096465	0,001665	0,069008	0,124583	0,090472
França	0,046355	0,037920	0,028868	0,001680	0,100873	0,093763
	0,112805	0,090926	0,041832	0,032044	0,181397	0,164834
Japão	0,001870	0,000880	0,003176	0,020460	0,095392	-
	0,094489	0,085052	0,049396	0,006075	0,091853	-
Estados Unidos	0,024951	0,024790	0,017103	0,022829	-0,024881	0,013143
	0,104685	0,106881	0,065594	0,010004	-0,029285	0,014046
Países asiáticos	*0,021931*	*0,021905*	*0,013826*	*0,002749*	*0,099789*	*-0,006460*
	0,095099	*0,099773*	*0,041630*	*0,001243*	*0,175045*	*0,099658*
Singapura	0,004782	0,004448	0,015869	0,037399	0,058787	-
	0,112488	0,116527	0,039765	-0,020264	0,100485	-
Xangai	0,021960	0,022025	0,007872	0,023597	0,089232	-
	0,085086	0,087289	0,014803	0,046244	0,199583	-
Taipei Chinês	0,022646	0,021297	0,023363	-0,014602	0,026113	0,010898
	0,100599	0,099438	0,059754	-0,104437	0,090202	0,104399
Hong Kong	0,008349	0,009217	0,011071	-0,013261	0,013131	-
	0,074560	0,077834	0,024750	-0,056790	0,090307	-
Países nórdicos	*0,032741*	*0,026820*	*0,010518*	*0,001404*	*0,053805*	*0,001146*
	0,085580	*0,072986*	*0,037608*	*0,024242*	*0,080825*	*0,040688*
Suécia	0,025901	0,018805	0,026544	0,013286	0,061552	-0,005258
	0,060257	0,054815	0,050230	0,035795	0,100419	0,049242
Dinamarca	0,038928	0,029816	0,023696	0,014031	0,063090	-0,017105
	0,100514	0,084131	0,045692	0,047543	0,101739	0,007524
Finlândia	0,036946	0,025590	0,013247	-0,007247	0,071405	-0,002194
	0,079705	0,059861	0,037939	0,011107	0,084542	0,012848

Fonte: Pisa 2012.
Elaboração dos autores.

INSERÇÃO PRODUTIVA INTERNACIONAL E INTER-REGIONAL: UMA ANÁLISE A PARTIR DAS CADEIAS GLOBAIS DE VALOR[1]

Fernando Salgueiro Perobelli[2]
Inácio Fernandes de Araújo[3]
Vinícius de Almeida Vale[4]
Mônica de Moura Pires[5]

Este artigo analisa a inserção produtiva no contexto da fragmentação internacional e inter-regional da produção a partir dos fluxos de comércio do estado da Bahia e do restante do Brasil. Para tal, utiliza-se a matriz de insumo-produto inter-regional da Bahia e o restante do Brasil, calibrada para o ano de 2009, com abertura setorial para 27 atividades econômicas. Os resultados permitiram identificar o valor adicionado doméstico nas exportações das duas regiões e compreender a integração econômica no contexto da reestruturação produtiva e da especialização vertical.

Palavras-chave: cadeias globais de valor; especialização vertical; insumo-produto; integração produtiva; comércio internacional.

INTERNATIONAL AND INTER-REGIONAL PRODUCTIVE INSERTION: AN ANALYSIS FROM THE GLOBAL VALUE CHAINS

This study analyzes the productive integration in the context of international and inter-regional fragmentation of production from trade flows of Bahia and the rest of Brazil. We have used an inter-regional input-output table for Bahia and the rest of Brazil, calibrated to 2009, considering 27 economic activities. The results allowed us to identify the domestic value added on exports of both regions and to understand economic integration in the context of the production restructuring and vertical specialization.

Keywords: global value chains; vertical specialization; input-output; productive integration; international trade.

JEL: F10; C67; R15; D57.

1. Os autores agradecem o suporte financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig). Além disso, estendem os agradecimentos à Superintendência de Estudos Econômicos e Sociais da Bahia (SEI) pela disponibilização das tabelas de recursos e usos de bens e de serviços (TRUs) e da tabela de componentes do valor adicionado da Bahia.

2. Professor do Departamento de Economia da Universidade Federal de Juiz de Fora (UFJF). Pesquisador do Laboratório de Análises Territoriais e Setoriais (Lates) da UFJF.

3. Doutor em economia pela UFJF. Pesquisador do Lates da UFJF.

4. Professor do Departamento de Economia da Universidade Federal do Paraná (UFPR). Pesquisador do Núcleo de Estudos em Desenvolvimento Urbano e Regional (Nedur), do Núcleo de Economia Internacional e Desenvolvimento Econômico (Neide) da UFPR e do do Lates da UFJF.

5. Professora do Departamento de Ciências Econômicas da Universidade Estadual de Santa Cruz (UESC), pesquisadora do Grupo de Economia Aplicada (Labor) da UESC.

1 INTRODUÇÃO

O comércio internacional tem aumentado desde da década de 1980 devido à maior liberalização comercial e redução nos custos de transações, por meio dos custos de transporte e das barreiras tarifárias e não tarifárias. A intensificação nas transações de bens e serviços entre os países também tem sido impulsionada pelas estratégias de terceirização dos estágios produtivos, cada vez mais dispersos entre as economias mundiais (Jones, 2000; Yi, 2003; Koopman, Wang e Wei, 2014). Além disso, a crescente verticalização e interconexão dos processos produtivos tem intensificado a especialização dos países ou regiões em fases específicas da cadeia produtiva (Hummels, Ishii e Yi, 2001).

A fragmentação do processo produtivo em estágios de produção localizados em diferentes países tem gerado uma tendência de aumento do comércio de insumos intermediários (Timmer *et al.*, 2014). Os países importam bens, adicionam valor em algum estágio de produção e, em seguida, os exportam para serem usados em outros estágios da cadeia produtiva. Assim, os bens transacionados no comércio internacional podem atravessar diversas vezes as mesmas fronteiras nacionais (Koopman, Wang e Wei, 2014).

Portanto, diante do cenário de intensificação da fragmentação produtiva internacional, o valor exportado deixou de ser gerado apenas na economia local. Desse modo, as estatísticas tradicionais de comércio, medidas em termos brutos, implicam dupla contagem do valor dos bens intermediários que atravessam as fronteiras internacionais mais de uma vez (Koopman, Wang e Wei, 2012; 2014). Nesse contexto é necessário, portanto, traçar estratégias para separar o conteúdo adicionado por meio da atividade doméstica ou estrangeira nas estatísticas de comércio, e, então, medir adequadamente a participação de cada país nas cadeias globais de valor (Tukker e Dietzenbacher, 2013).

Para contornar as limitações em mensurar a dimensão real das exportações no desempenho da economia de uma região, a metodologia de insumo-produto (IP) tem sido utilizada no contexto das cadeias globais de valor (Dietzenbacher *et al.*, 2013). Em análises empíricas alguns autores, como Hummels, Ishii e Yi (2001), Koopman, Wang e Wei (2012), Chen *et al.* (2012), Guilhoto e Imori (2014), Rébora e Vivanco (2015) e Fujii-Gambero e Cervantes-Martínez (2017), utilizaram a metodologia de IP no contexto nacional (matrizes regionais de IP). No contexto multirregional (matrizes inter-regionais de IP), Bems, Johnson e Yi (2011), Johnson e Noguera (2012), Dietzenbacher, Guilhoto e Imori (2013), Koopman, Wang e Wei (2014), Timmer *et al.* (2014), Los, Timmer e De Vries (2015; 2016), Fujita e Hamaguchi (2016), Prete e Rungi (2017) e Miller e Temurshoev (2017) são alguns dos autores que contribuem com a literatura.

Nesse contexto, vale ressaltar também o esforço conjunto da Organisation for Economic Co-operation and Development (OECD) e da World Trade Organization (WTO) na construção da base de dados *trade in value-added* (TiVA). A TiVA, construída a partir das matrizes globais de insumo-produto da OECD, apresenta estatísticas detalhas das cadeias globais de produção.

As cadeias de produção podem ser fragmentadas tanto em nível global quanto regional. Entretanto, as discussões na literatura em relação à fragmentação da produção regional ainda são escassas – principalmente dentro das fronteiras de um território nacional. Essa é uma questão importante e afeta diretamente a formulação de políticas, visto que as estratégias de inserção das cadeias de valor fragmentadas dentro do território nacional são distintas daquelas adotadas para cadeias de valor mais conectadas ao mercado internacional.

Dentro dessa temática, este trabalho analisa a inserção produtiva no contexto da fragmentação internacional e inter-regional da produção a partir dos fluxos de comércio do estado da Bahia e do restante do Brasil. A fragmentação produtiva é identificada no espaço nacional (a partir das interações entre o estado da Bahia e o restante do Brasil) e internacional (a partir das interações do estado da Bahia e do Brasil com o restante do mundo). Para tal, utiliza-se a matriz de insumo-produto inter-regional do estado da Bahia e restante do Brasil (MIP-IR BA-RB), calibrada para o ano de 2009, com abertura setorial para 27 atividades econômicas. A partir dessa matriz, identifica-se a participação do conteúdo doméstico e do valor adicionado doméstico nas exportações. Avalia-se, assim, a inserção produtiva, tomando por base as discussões a respeito das cadeias globais de valor, o que permite compreender a inserção dessas duas regiões no contexto da reestruturação produtiva e da especialização vertical.

A reorganização da produção nacional tem ocorrido por diversos fatores, em especial pela guerra fiscal[6] e pelas diferenças salariais entre as Unidades da Federação. Essas questões têm levado as indústrias, em especial as têxteis, automobilísticas e calçadistas, a deslocarem parte de suas estruturas produtivas das regiões Sul e Sudeste para a região Nordeste do país. Essa relocalização da estrutura produtiva também está relacionada à intensificação do fluxo de comércio entre as regiões brasileiras (Haddad e Perobelli, 2002; Perobelli e Haddad, 2006a; 2006b; Perobelli, Haddad e Domingues, 2009; Perobelli *et al.*, 2010), impulsionada pela redução nos custos de transportes inter-regionais (Haddad *et al.*, 2011). Para compreender esse fenômeno, escolheu-se, entre os estados do Nordeste, analisar a economia baiana, que ocupou a sétima posição na participação do produto interno bruto (PIB) nacional em 2015 (IBGE, 2017) e pela sua importância no contexto dessa

6. Para mais detalhes sobre a relocalização dos investimentos entre os estados brasileiros, condicionada à política tributária, consultar Domingues e Haddad (2003).

macrorregião. O PIB do estado correspondia a cerca de 30,0% do PIB do Nordeste em 2010, e a sua participação no total das exportações passou de 56%, em 2010, para 60%, em 2012, sobretudo pelos preços favoráveis das *commodities* (petróleo e derivados, soja, algodão e metais preciosos), consolidando a sua posição de liderança regional (Bahia, 2014). Para atingir essa posição de destaque, o estado recebeu diversos investimentos, o que tornou possível adentrar-se em etapas específicas da produção de indústrias conectadas a outras regiões, especialmente nas atividades de refino de petróleo, produtos químicos e produção automobilística.

Do exposto extrai-se que a contribuição deste estudo permite compreender a respeito da inserção regional e internacional no contexto da especialização vertical. Os resultados, por meio da mensuração do conteúdo local nas exportações, também auxiliam a compreender a importância real do comércio inter-regional e internacional sobre a economia. Além disso, com a utilização da MIP-IR BA-RB, tem-se uma mensuração dessa inserção com razoável nível de desagregação setorial.

Cabe destacar que as evidências sobre a inserção da economia brasileira nas cadeias globais de valor ainda são escassas. Entre os estudos já realizados utilizando a abordagem de insumo-produto, Dietzenbacher, Guilhoto e Imori (2013), Guilhoto e Imori (2014), Ferraz, Gutierre e Cabral (2016) e Araújo, Perobelli e Faria (2016) analisaram as relações comerciais do Brasil com outras economias no contexto da especialização vertical; e Imori (2016) investigou os aspectos econômicos e ambientais da participação dos estados brasileiros nas cadeias globais de valor. A inserção brasileira nas cadeias de valor também foi analisada em seus aspectos relacionados à política comercial e industrial por Veiga e Rios (2017), e, com enfoque nas atividades de serviços, por Oliveira, Reis e Bloch (2017).

Ademais, o panorama de mudanças no contexto internacional faz com que a ciência regional no Brasil depare-se com o seguinte questionamento: como promover o crescimento regional de forma a diminuir as disparidades ainda existentes no país e, ao mesmo tempo, inserir a economia brasileira na economia mundial? A maior inserção da economia brasileira no contexto internacional depende do aumento de produtividade capaz de gerar maior competitividade, redução de custos e reestruturação produtiva. No entanto, esse processo pode reforçar desequilíbrios regionais e criar dificuldades de desenvolvimento em áreas consideradas pouco dinâmicas. Portanto, tomando por base a ideia de aumento dos fluxos de comércio como propulsor do crescimento regional, e reconhecendo a heterogeneidade espacial do desenvolvimento brasileiro, pode-se, a partir de instrumentos analíticos eficazes, avaliar, dentro do contexto das cadeias globais de valor, a inserção das economias subnacionais no comércio inter-regional e internacional.

Além desta introdução, o trabalho está organizado da seguinte forma. Na segunda seção são discutidos os principais pontos teóricos acerca da inserção regional

no comércio exterior e nas cadeias globais de produção. Na terceira seção aborda-se a estratégia empírica do estudo, detalhando a estrutura de insumo-produto usada para medir o conteúdo importado e o valor adicionado estrangeiro nas exportações. Na quarta seção apresenta-se a base de dados e fatos estilizados da integração ao comércio internacional do estado da Bahia e do restante do Brasil. Na quinta seção são apresentados e discutidos os resultados. Por fim, na sexta seção apresentam-se as considerações finais.

2 ABORDAGEM TEÓRICA

A compreensão acerca do comércio internacional e da sua importância para o crescimento regional parte principalmente das formulações do teorema de Heckscher-Ohlin[7] e do modelo da base de exportação.[8] No entanto, essas formulações teóricas, apesar de importantes para compreensão do comércio internacional, não explicam uma parte essencial do padrão de comércio.[9] Brown e Anderson (1999) destacam que: a) o comércio acontece mesmo sem existir diferenças de dotação de fatores entre os países – em geral, observa-se que a maior parte do comércio ocorre entre os países desenvolvidos, apesar de eles possuírem dotação de fatores similares quando comparada aos países em desenvolvimento; e b) elevada participação do comércio mundial consiste em fluxos bidirecionais de produtos similares (comércio intraindústria). Além dessas críticas aos modelos tradicionais que explicam a inserção regional no comércio exterior, a tendência de intensificação da fragmentação internacional da produção impulsionou debates em torno das cadeias globais de valor[10] (Baldwin, 2006).

A intensificação da fragmentação internacional da produção apenas foi possível devido à coordenação dos processos produtivos por meio da governança das cadeias de valor. Gereffi, Humphrey e Sturgeon (2005) ressaltam que essa governança baseia-se na: *i)* complexidade dos custos de transação; *ii)* capacidade de codificar as transações; e *iii)* capacidade tecnológica e de aprendizagem no nível das empresas. Portanto, a compreensão das cadeias de produção está centrada em dois componentes principais: a) fragmentação de processos produtivos em diferentes firmas e países (Hummels, Ishii e Yi, 2001); e b) processo de governança das cadeias de valor por uma empresa líder (Gereffi, Humphrey e Sturgeon, 2005).

7. Consultar Flam e Flanders (1991) para mais detalhes sobre o teorema de Heckscher-Ohlin.

8. A hipótese central do modelo de base de exportação é determinação da renda regional pelas exportações, ou seja, uma determinada região deve aumentar a entrada de fluxos monetários para crescer. A única maneira efetiva disso ocorrer é aumentando as exportações. Assim, parte-se da ideia de que as atividades exportadoras são determinantes para o crescimento regional (Richardson, 1969).

9. As formulações do teorema de Heckscher-Ohlin coexistam às formulações a respeito das cadeias globais de produção – para mais detalhes, consultar Ito, Rotunno e Vézina (2017).

10. O processo de fragmentada internacional da produção é comumente referido como "especialização vertical", "*offshoring*", "cadeias globais de produção" e "cadeias globais de valor".

Nesse contexto, uma maior parcela de valor é adicionada ao longo das cadeias globais de produção distribuídas em diferentes países (Hummels, Ishii e Yi, 2001). No entanto, cabe ressaltar que essas cadeias de produção do século 21 são diferentes daquelas que existiam nas economias mais desenvolvidas no século passado (Baldwin e Evenett, 2015). Atualmente, as cadeias de produção são mais do que o comércio extra de peças e componentes. O investimento direto estrangeiro é uma parte integrante das redes globais de suprimento, e as relações comerciais de longo prazo também levam a fluxos transfronteiriços formais de *know-how*, como propriedade intelectual (tecnologia) e formas mais tácitas de transmissão de conhecimento, a exemplo de experiência de gestão e *marketing* (Baldwin, 2011). Desse modo, o desempenho das indústrias, das regiões e dos países estão cada vez mais dependentes da integração às cadeias globais de valor.

A fragmentação internacional da produção tem atraído a atenção de diversos pesquisadores, tanto em formulações teóricas – Grossman e Rossi-Hansberg (2008) e Costinot, Vogel e Wang (2013) – quanto em abordagens empíricas – Hummels, Ishii e Yi (2001), Bems, Johnson e Yi (2011), Johnson e Noguera (2012), Koopman, Wang e Wei (2012; 2014), Timmer *et al.* (2014), Los, Timmer e De Vries (2015; 2016), Fujita e Hamaguchi (2016), Prete e Rungi (2017), Miller e Temurshoev (2017) e Fujii-Gambero e Cervantes-Martínez (2017). Esses estudos focam principalmente na posição que os países ocupam nas redes globais de produção e na estimativa do conteúdo importado e do valor adicionado doméstico nas exportações.

Vale ressaltar que a amplitude das cadeias de valor pode ser descrita como local, uma vez que pode incluir países geograficamente próximos, ou global, envolvendo países de diferentes continentes (Los, Timmer e De Vries, 2015). Além disso, as cadeias de valor podem assumir característica essencialmente nacional, quando a produção está dividida em diversos estágios nas regiões dentro do território nacional.

3 ESTRATÉGIA EMPÍRICA

Com o intuito de analisar a inserção produtiva no contexto da fragmentação internacional e inter-regional da produção, este trabalho utiliza a abordagem proposta por Hummels, Ishii e Yi (2001) e especificada, explicitamente, na metodologia de insumo-produto por Chen *et al.* (2004), e posteriormente aperfeiçoada por Koopman, Wang e Wei (2012). Essa abordagem permite compreender a relação entre o conceito de conteúdo importado e de conteúdo nacional.

Os insumos intermediários importados ou produzidos pela economia doméstica podem ser contabilizados separadamente a partir da estrutura de insumo-produto. Em termos matriciais, pode-se representar os encadeamentos produtivos nas matrizes de insumo-produto da seguinte forma:

$$A^D X + Y^D = X, \quad (1)$$

$$A^F X + Y^F = I, \quad (2)$$

$$uA^D + uA^F + A_v = u, \quad (3)$$

em que $A^D = [a_{ij}]$ é uma matriz n x n de coeficientes técnicos de produtos domésticos; $A^F = [a_{ij}^F]$ é uma matriz n x n de conteúdo importado (internacional ou inter-regional) usado na produção doméstica, sendo que os subscritos i e j indicam os setores; Y^D é um vetor n x 1 de demanda final por bens produzidos domesticamente, incluindo exportações, consumo das famílias e do governo e formação bruta de capital fixo; Y^F é um vetor n x 1 de demanda final por produtos importados, incluindo famílias, governo e formação bruta de capital fixo; X é um vetor n x 1 de produto total; I é um vetor n x 1 de importações com origem $F = \{M, R\}$ que identifica a possibilidade de o insumo importado ter origem internacional (M) ou inter-regional (R); $A_v = [a_j^v]$ é um vetor 1 x n da divisão entre o valor adicionado (VA_i) e o produto total (X_i) para cada setor j; e u é um vetor unidade 1 x n.

Os elementos de A^D representam os insumos domésticos provenientes de um setor i que são necessários para a produção de uma unidade do setor j. Cada elemento a_{ij}^F de A^F representa os insumos importados (internacionais ou inter-regionais) do setor i usados na produção do setor j.

As equações (1) e (2) definem as duas condições de equilíbrio horizontais, para os bens produzidos internamente e importados, respectivamente. A equação (1) define que a produção doméstica total k deve ser igual à soma das vendas de k para todos os demandantes intermediários e finais na economia (as vendas finais incluem consumo interno, formação de capital fixo e exportações). A equação (2) especifica que o total das importações (internacionais ou inter-regionais) do setor h deve ser igual à soma das vendas do produto h para todos os usuários na economia, incluindo insumos intermediários para todos os setores e o consumo final doméstico e a formação de capital fixo. A equação (3), por sua vez, é uma condição de equilíbrio vertical e uma restrição de soma dos coeficientes de insumo-produto. Isso implica que a produção total (X) em qualquer setor k deve ser igual à soma do valor adicionado por k e o custo dos insumos intermediários de todos os bens produzidos internamente e importados, aplicando-se esses conceitos de contas separadas para insumos intermediários e importados (internacionais ou inter-regionais).

A partir da equação (1) tem-se o modelo convencional de insumo-produto:

$$X = (I - A)^{-1}Y, \tag{4}$$

em que $(I - A)^{-1}$ é a inversa de Leontief, que representa os requerimentos totais (diretos e indiretos) por insumos domésticos para produzir uma unidade de demanda final.

As matrizes de insumo-produto permitem calcular o valor dos insumos importados, por meio do comércio inter-regional ou internacional, e usados indiretamente na produção de um bem exportado. Assim, os insumos importados podem ser utilizados em um setor cujas vendas são destinadas para um segundo setor, e, eventualmente, incorporados em um bem exportado (requisitos indiretos). Desse modo, o termo L^F inclui todos os insumos importados diretos e indiretos:

$$L^F = A_0^F(I - A)^{-1}, \tag{5}$$

em que $A_0^F = \sum_{i=1}^{F} a_{ij}^F$ é um vetor 1 x n do total de coeficientes importados pelo setor j, com $F = \{M, R\}$.

A equação (5) indica que o conteúdo importado de um setor corresponde à participação das importações de cada setor ponderado pelo coeficiente de requerimento (direto e indireto) de produção doméstica. O modelo assume que as exportações e as vendas no mercado interno são produzidas pela mesma tecnologia. Assim, a equação (5) apresenta a participação do conteúdo importado na produção doméstica e nas exportações de cada setor.

Uma vez que a participação do conteúdo de importações tenha sido determinada, é possível calcular o conteúdo de importados nas exportações totais (λ):

$$\lambda = A_0^F(I - A)^{-1}e^X, \tag{6}$$

em que e^X é um vetor n x 1 de exportações (para outros países ou para outras regiões dentro do contexto inter-regional nacional). Neste caso, o termo $(I - A)^{-1}$ capta o insumo importado incluído nas exportações de forma direta e indireta, por meio dos *linkages* para trás. Essa equação possui a mesma formulação usada para calcular a especialização vertical por Hummels, Ishii e Yi (2001). Assim, os conceitos de especialização vertical e de conteúdo importado nas exportações totais são idênticos.

Similarmente, é possível estimar a contribuição do valor adicionado doméstico e estrangeiro nas exportações. Para tal, muda-se o vetor do total de importações (A_0^F) por um vetor com a razão entre o valor adicionado e a produção ($A_v = VA/X$). A participação do valor adicionado doméstico, $\eta = \{\eta_j\}$, é definida como um vetor

1 x n, com o valor adicionado gerado na economia local por uma unidade adicional de demanda final pelos produtos nacionais:

$$\eta = A_v(I - A)^{-1}. \quad (7)$$

A participação do valor adicionado estrangeiro na demanda final por bens produzidos domesticamente, por sua vez, é definida por $\varepsilon = 1 - \eta$. Consequentemente, a contribuição do valor adicionado estrangeiro no total das exportações ($VA\varepsilon$) é dada por:

$$VA\varepsilon = 1 - A_v(I - A)^{-1}e^X. \quad (8)$$

A equação (8) é uma medida para as exportações de valor adicionado, ou seja, o valor adicionado produzido no país/região de origem e absorvido no país/região de destino.

4 BASE DE DADOS

Este estudo usou a matriz MIP-IR BA-RB para o ano de 2009 estimada no Laboratório de Análises Territoriais e Setoriais (Lates) da Faculdade de Economia da Universidade Federal de Juiz de Fora (UFJF).[11] Para a estimação da matriz inter-regional foi tomada como fonte inicial dos dados a matriz regional de insumo-produto da Bahia, estimada por Perobelli *et al.* (2015), e a matriz nacional de insumo-produto do Brasil, estimada pelo Núcleo de Economia Regional e Urbana da Universidade de São Paulo (Nereus). Tanto a MIP regional da Bahia quanto a MIP nacional do Brasil foram estimadas de acordo com a metodologia proposta por Guilhoto e Sesso Filho (2005) e testada a posteriori por Guilhoto e Sesso Filho (2010). A MIP Bahia foi estimada a partir das tabelas de recursos e usos (TRUs) de bens e de serviços e da tabela de componentes do valor adicionado da Bahia para o ano de 2009, disponibilizadas pela Superintendência de Estudos Econômicos e Sociais da Bahia (SEI), que seguem a metodologia do Sistema de Contas Regionais do Brasil (IBGE, 2014) e possuem abertura setorial para 27 atividades econômicas (Bahia, 2013). A MIP nacional do Brasil foi estimada a partir das TRUs do Sistema de Contas Nacionais do Brasil (IBGE, 2011), para 2009 e com abertura setorial para 56 atividades econômicas.[12]

Na estimação da matriz inter-regional foram seguidas as estratégias metodológicas de Haddad *et al.* (2002), Haddad e Domingues (2003), Porsse, Haddad e Ribeiro (2003) e Brene (2013), fundamentadas em Hulu e Hewings (1993) e Miller e

11. A matriz de insumo-produto inter-regional da Bahia e restante do Brasil está disponível no endereço eletrônico: http://www.ufjf.br/lates/. Acesso em: 7 mar. 2019.

12. Foi realizada a compatibilização setorial prevalecendo a menor abertura setorial identificada na MIP da Bahia.

Blair (2009). Para os fluxos intrarregionais, $Z_{ij}^{BA,BA}$, utilizou-se a matriz da Bahia. O fluxo de comércio inter-regional com origem na Bahia e destino o restante do Brasil, $Z_{ij}^{BA,RB}$, foi estimado a partir do quociente locacional (QL), que identifica o quanto cada setor está concentrado em cada região. Desse modo, o $QL_i^r > 0{,}8$ indica que o setor *i* é concentrado na região *r*, sugerindo a capacidade desse setor em suprir a demanda local. Nesse caso, assume-se que os coeficientes técnicos de insumo nacional para o setor *i*, a_{ij}^n, aplicam-se à região *r*, e o excedente regional produzido por *i* será exportado para o restante do país. Por outro lado, se o setor *i* é menos concentrado na Bahia do que no restante do país, $QL_i^r < 0{,}8$, considera-se que a produção na Bahia não apresenta capacidade para suprir a demanda local; assim, os coeficientes técnicos de insumo inter-regionais são criados por meio dos coeficientes nacionais, a_{ij}^n, multiplicados pelo valor obtido no cálculo do QL_i^r. Os insumos intermediários com origem no restante do Brasil e destino a Bahia, $Z_{ij}^{RB,BA}$, foram obtidos a partir da ponderação do vetor de importações do restante do Brasil pela Bahia pela participação criada a partir da matriz de produção da Bahia. As estimativas obtidas por meio do QL foram ajustadas, aplicando-se o método RAS. Por fim, os fluxos de insumos intermediários entre o restante do Brasil, $Z_{ij}^{RB,RB}$, foram estimados pela subtração dos fluxos de comércio da MIP do Brasil pela MIP da Bahia. Os componentes inter-regionais da demanda final foram obtidos com o uso do mesmo procedimento. O quadro A.1, no apêndice, apresenta as transações inter-regionais do estado da Bahia com o restante do Brasil.

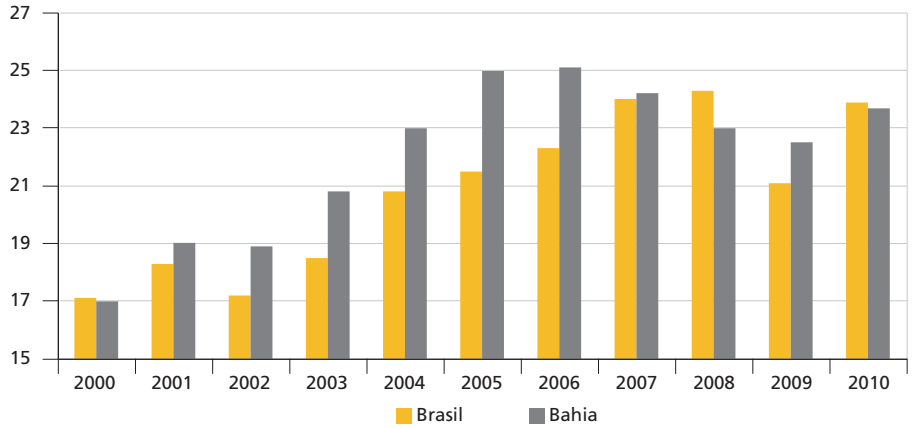
Assim, a partir da matriz de insumo-produto inter-regional da Bahia e restante do Brasil, foi possível mapear de forma explícita todos os fluxos de comércio entre as duas regiões (fluxos intermediários e fluxos inerentes à demanda final). Nessa matriz são decompostos o consumo intermediário, os elementos da demanda final e os componentes do valor adicionado com origem na Bahia e no restante do Brasil, o que aumenta o poder de explicação do modelo.

4.1 Fatos estilizados

A economia brasileira possui pouca inserção nos acordos internacionais de comércio (Thorstensen e Ferraz, 2014), apesar da importância da integração comercial como estratégia para o crescimento econômico. Cabe ressaltar que o Brasil tem tentado participar de acordos específicos, embora, por vezes, não tenha conseguido concluir as negociações. Enquanto isso, ocorre uma tendência de formação de novos acordos de comércio e ampliação dos acordos já existentes (North American Free Trade Agreement – Nafta, União Europeia, Mercado Comum do Sul – Mercosul, Acordo de Associação Transpacífico) –, embora também seja observada uma tendência de formação de acordos bilaterais em oposição à formação de acordos multilaterais.

O distanciamento da economia brasileira do mercado internacional pode ser verificado, por exemplo, pelo grau de abertura ao comércio exterior, medido pela soma das exportações e das importações em relação ao PIB, que não ultrapassou 25% ao longo da primeira década dos anos 2000 (gráfico 1). Enquanto para a América Latina, segundo dados da Comissão Econômica para a América Latina e o Caribe (Cepal), no início da década de 1990, esse índice de abertura comercial já era 20%, passando, no início da década de 2010, para 43%. No entanto, esses percentuais, segundo dados do Banco Mundial, ainda se encontram abaixo da média mundial, 45% e 73%, respectivamente.

GRÁFICO 1
Grau de abertura ao comércio exterior do Brasil e estado da Bahia (2000-2010)
(Em %)



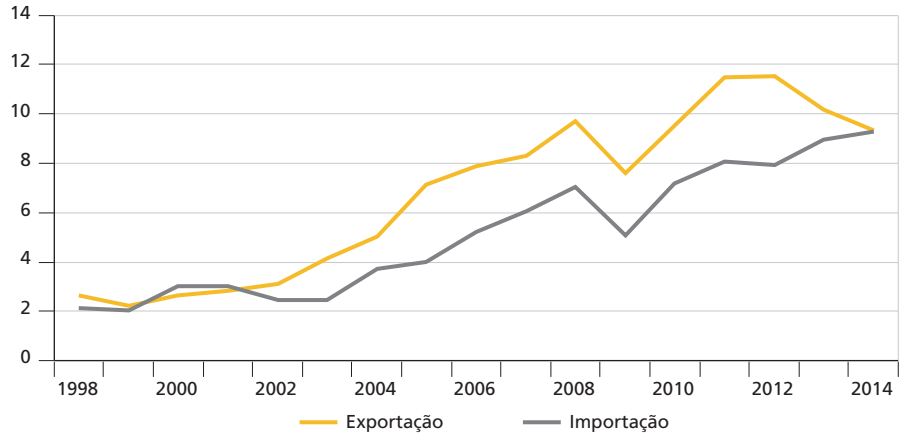
Fonte: Bacen (2013).
Elaboração dos autores.

O grau de abertura ao comércio exterior da Bahia está próximo à média nacional, o que demonstra uma economia ainda muito fechada ao mercado exterior. Mesmo assim, percebe-se que o volume exportado e importado ao longo do período 1998-2014 tem aumentado (gráfico 2), embora, em termos relativos, a participação da Bahia no comércio exterior no Brasil tenha se reduzido (gráfico 3). A contribuição das exportações e das importações baianas em relação ao total do Brasil, após um período de aumento, decresceu a partir de 2005, mantendo-se em torno de 4,2% em 2014 (gráfico 3).

No período 1998-2014, os principais parceiros comerciais da Bahia eram os países do Nafta, da União Europeia, dos BRICS e do Mercosul. As exportações com destino para a União Europeia e os BRICS atingiram, respectivamente, 24,9% e 19,5% das exportações totais do estado em 2014. O Nafta, que havia sido o principal destino das exportações estaduais em 2003, 41,9%, perdeu participação,

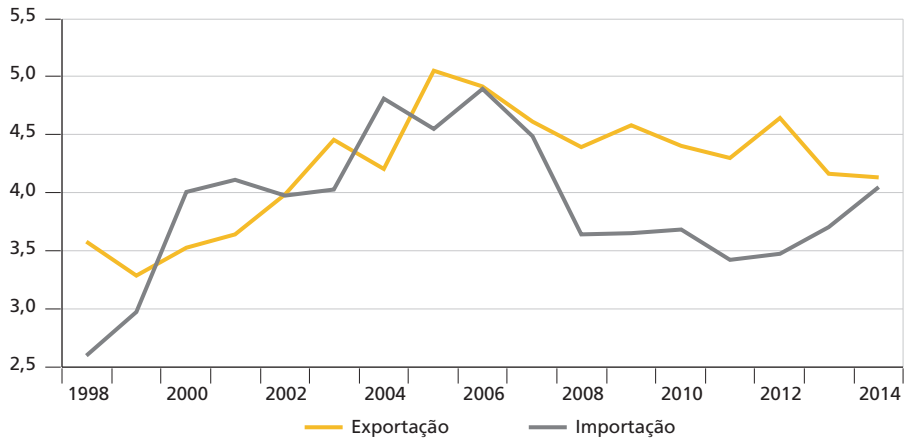
atingindo 13,1% em 2014 (gráfico 4A). Em termos de origem das importações, o Nafta e o Mercosul constituem-se os principais parceiros comerciais da Bahia, com 18,5% e 15,1% do total das importações, respectivamente (gráfico 4B). Esse padrão de comércio, no entanto, é diferente do restante do país (gráficos 4C e 4D).

GRÁFICO 2
Comércio internacional do estado da Bahia: volume exportado e importado
(Em US$ bilhões a preços de 2014 FOB)



Fonte: MDIC.
Elaboração dos autores.

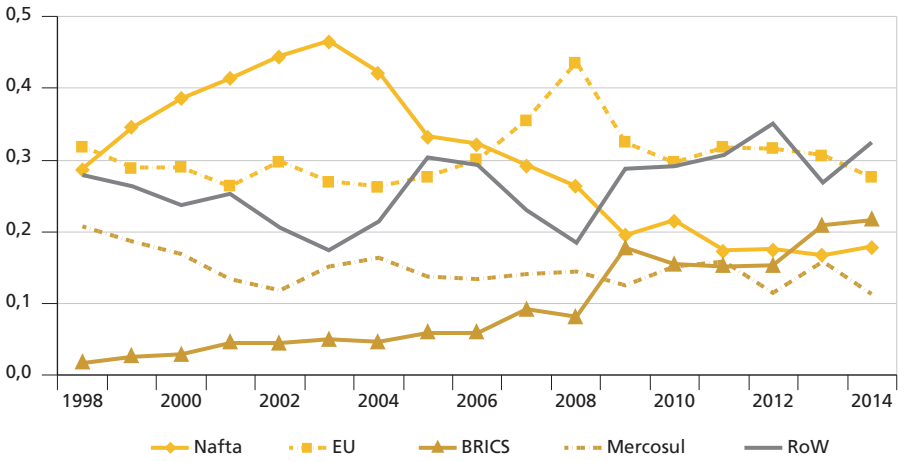
GRÁFICO 3
Participação do estado da Bahia no comércio exterior do Brasil: proporção do volume exportado e importado
(Em %)



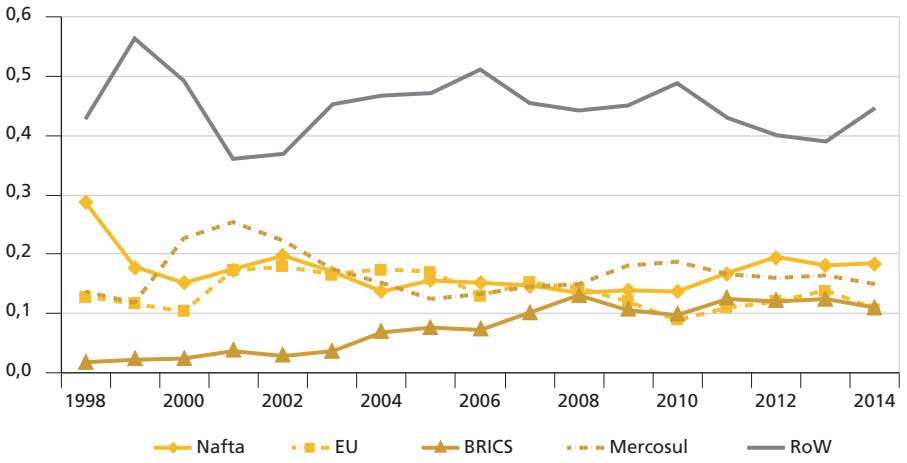
Fonte: MDIC.
Elaboração dos autores.

GRÁFICO 4
Comércio internacional do estado da Bahia: exportações e importações, por destino e origem (1998-2014)
(Em %)

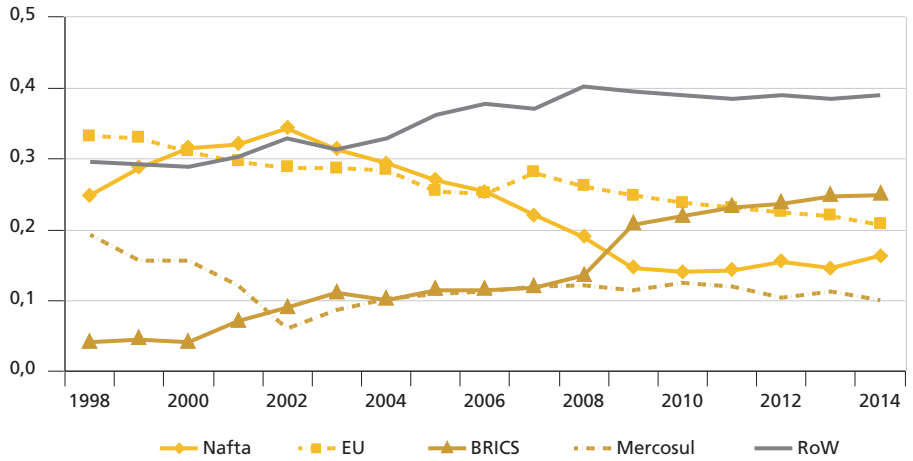
4A – Comércio internacional do estado da Bahia: destino das exportações



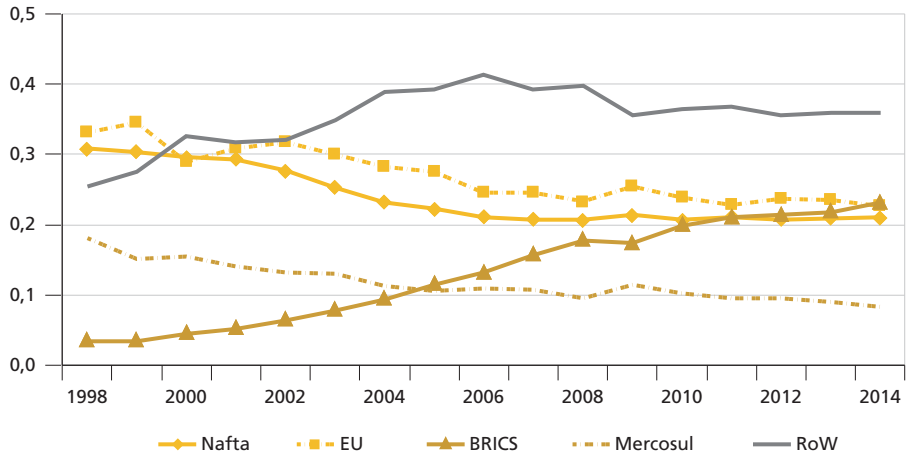
4B – Comércio internacional do estado da Bahia: origem das importações



4C – Comércio internacional do Brasil: destino das exportações



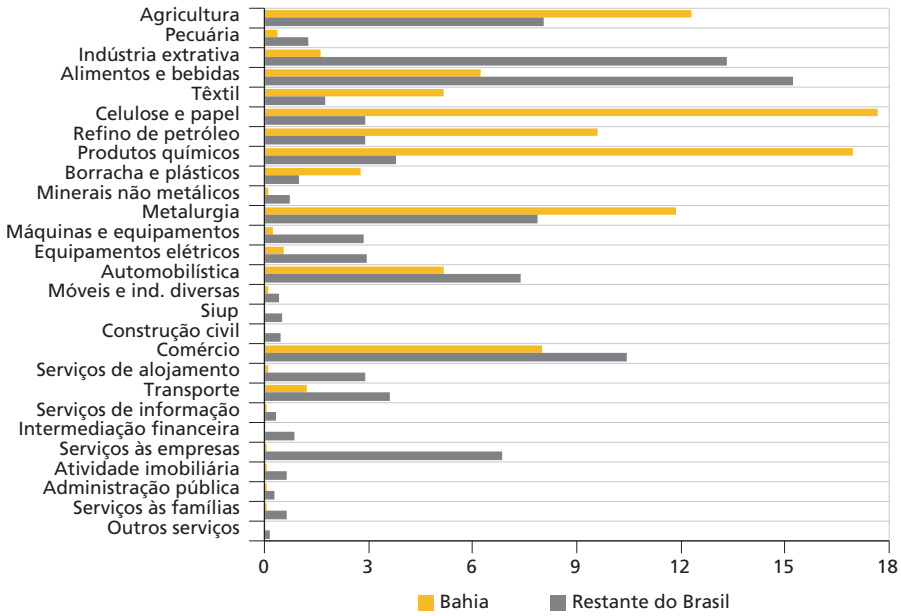
4D – Comércio internacional do Brasil: origem das importações



Fonte: MDIC.
Elaboração dos autores.
Obs.: RoW refere-se ao restante do mundo.

O gráfico 5 apresenta a distribuição setorial das exportações do estado da Bahia e do restante do Brasil para o ano de 2009. As exportações baianas estão concentradas nas atividades de celulose e papel (17,7%), produtos químicos (17%), agricultura (12,3%), metalurgia (11,9%) e refino do petróleo (9,6%). Esses setores representam 68,5% das exportações baianas. O restante do Brasil também apresenta pauta de exportação pouco diversificada e concentrada em bens primários e de pouco valor agregado ao longo na cadeia produtiva. As exportações do restante do Brasil são constituídas principalmente pelos setores de alimentos e bebidas (14,9%), indústria extrativa (12,9%), comércio (10,4%), agricultura (8,2%) e metalurgia (8%).

GRÁFICO 5
Participação das exportações setoriais no total exportado: Bahia e restante do Brasil (2009)
(Em %)



Fonte: MIP-IR BA-RB (2009).
Elaboração dos autores.

Conforme abordado anteriormente, o valor das exportações não é totalmente gerado na economia doméstica, visto que incorpora uma parcela de insumos importados à produção. Desse modo, torna-se relevante identificar o efeito real do comércio exterior no estado da Bahia e no restante do Brasil. Com esse objetivo, além das estatísticas tradicionais de comércio, é importante identificar o conteúdo importado presente nas exportações, o que é feito na próxima seção, por meio das medidas de fragmentação produtiva dessas regiões.

5 RESULTADOS

5.1 Fragmentação internacional: uma análise para Bahia e restante do Brasil

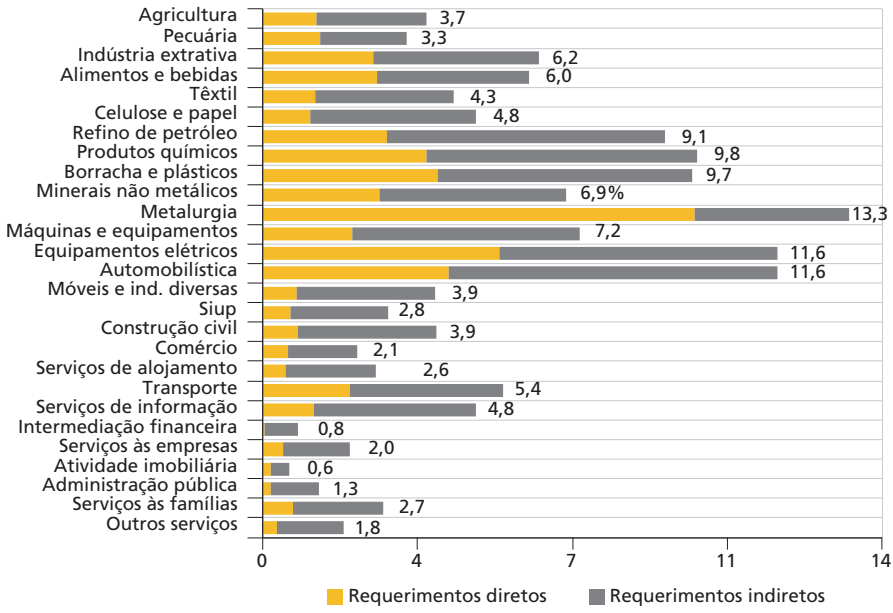
O conteúdo importado nas exportações foi calculado usando MIP-IR BA-RB para o ano de 2009, a partir da equação (5). A medida do conteúdo importado nas exportações representa o valor que atravessa as fronteiras, pelo menos mais de uma vez. No estado da Bahia, observa-se que o conteúdo importado representava 7,3% das exportações no ano de 2009, enquanto 9,7% das exportações do restante do Brasil, para o mesmo ano, eram formadas por insumos intermediários importados.

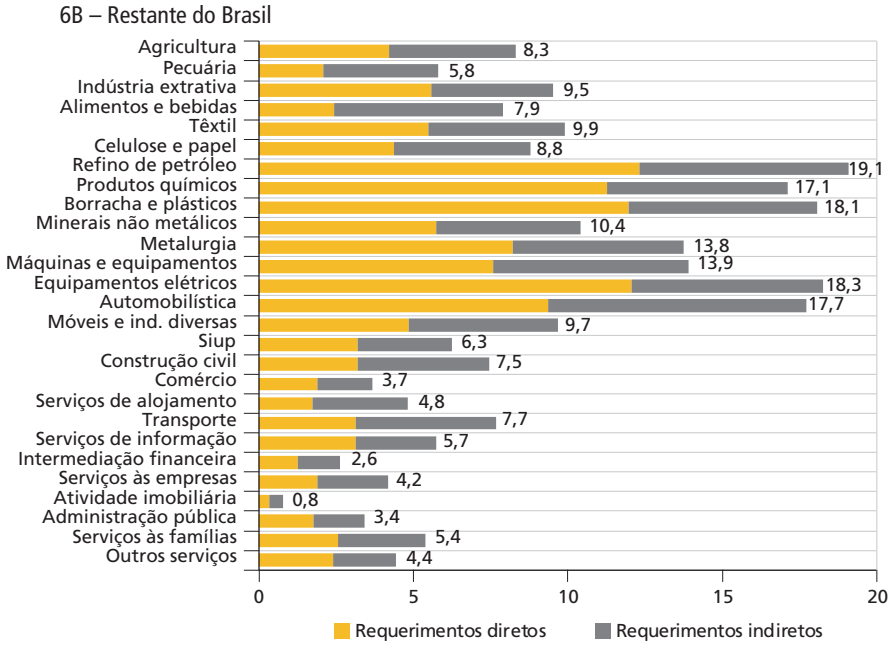
O gráfico 6 mostra a importância relativa do conteúdo importado nas exportações do estado da Bahia e do restante do Brasil em requerimentos diretos e indiretos, por setor de atividade, o que permite mensurar a participação dessas duas economias nas cadeias globais de produção. A diferença setorial na quantidade de conteúdo importado presente nas exportações das duas regiões é determinada pela proporção relativa de insumos intermediários importados na produção e pela estrutura de *linkages* setoriais.

A quantidade de conteúdo importado nas exportações foi maior, tanto para o estado da Bahia quanto para o restante do Brasil, principalmente nas indústrias de transformação: refino de petróleo, produtos químicos, borracha e plástico, minerais não metálicos, metalurgia, máquinas e equipamentos, materiais elétricos e automobilística. Todos os setores produtivos na Bahia apresentam menor participação de importados nas exportações, quando comparados com o restante do Brasil. Esse resultado pode estar relacionado com a estrutura de produção, a capacidade de integração ao mercado mundial e a própria integração dentro do país. No estado da Bahia, à exceção da indústria metalúrgica, a maior parte do conteúdo importado nas exportações foi originada de requerimentos indiretos.

GRÁFICO 6
Requerimentos diretos e indiretos de conteúdo importado nas exportações: Bahia e restante do Brasil (2009)
(Em %)

6A – Bahia





Elaboração dos autores.

A tabela 1 apresenta os resultados para a decomposição da participação do conteúdo importado, em relação ao produto setorial, nas exportações do estado da Bahia e do restante do Brasil para o ano de 2009. Em média, as exportações contribuem pouco para a produção setorial nas duas regiões. Na Bahia, as indústrias de celulose e papel, metalurgia, têxtil e produtos químicos são as mais voltadas para o mercado internacional. Embora haja tendência de uma relação positiva entre o conteúdo importado nas exportações e a importância do mercado externo, medido pela participação das exportações no produto setorial, não existe uma relação causal direta entre as duas medidas. Há setores, por exemplo, materiais elétricos, que possuem elevada participação de conteúdo importado (11,6%, gráfico 6A), mas com produção voltada para o mercado interno, apenas 2,3% da produção é exportada (tabela 1).

TABELA 1
Participação das exportações e do conteúdo importado por setor: Bahia e restante do Brasil (2009)
(Em % do produto setorial)

Setor	Bahia			Restante do Brasil		
	e^X	Λ	$e^X - \lambda$	e^X	Λ	$e^X - \lambda$
1. Agricultura	15,87	0,59	15,28	15,62	1,30	14,32
2. Pecuária	1,39	0,05	1,35	4,11	0,24	3,87
3. Indústria extrativa	6,19	0,39	5,80	33,71	3,21	30,50
4. Alimentos e bebidas	8,31	0,50	7,81	13,58	1,07	12,51
5. Têxtil	16,85	0,73	16,12	5,49	0,54	4,95
6. Celulose e papel	58,89	2,84	56,05	9,37	0,82	8,55
7. Refino de petróleo	7,45	0,68	6,77	5,98	1,14	4,84
8. Produtos químicos	13,71	1,35	12,37	6,77	1,16	5,61
9. Borracha e plásticos	10,04	0,98	9,07	5,66	1,02	4,64
10. Minerais não metálicos	1,16	0,08	1,08	4,50	0,47	4,03
11. Metalurgia	20,96	2,78	18,19	15,61	2,15	13,46
12. Máquinas e equipamentos	2,57	0,18	2,39	11,06	1,54	9,52
13. Equipamentos elétricos	2,33	0,27	2,06	7,86	1,44	6,42
14. Automobilística	10,55	1,23	9,32	11,65	2,06	9,59
15. Móveis e indústrias diversas	1,25	0,05	1,21	3,19	0,31	2,88
16. Siup	0,00	0,00	0,00	0,99	0,06	0,93
17. Construção civil	0,00	0,00	0,00	0,56	0,04	0,52
18. Comércio	4,43	0,10	4,34	6,61	0,24	6,37
19. Serviços de alojamento	0,17	0,00	0,17	8,17	0,39	7,78
20. Transporte	1,36	0,07	1,29	4,47	0,34	4,12
21. Serviços de informação	0,01	0,00	0,01	0,50	0,03	0,47
22. Intermediação financeira	0,00	0,00	0,00	0,89	0,02	0,86
23. Serviços às empresas	0,00	0,00	0,00	9,78	0,41	9,37
24. Atividade imobiliária	0,00	0,00	0,00	0,87	0,01	0,87
25. Administração pública	0,02	0,00	0,02	0,14	0,00	0,14
26. Serviços às famílias	0,01	0,00	0,01	1,69	0,09	1,59
27. Outros serviços	0,00	0,00	0,00	0,24	0,01	0,23
Total (% produto setorial)	**5,65**	**0,41**	**5,23**	**6,13**	**0,60**	**5,54**

Elaboração dos autores.
Obs.: e^X: exportações totais; λ: conteúdo importado nas exportações; $e^X - \lambda$: exportações líquidas.

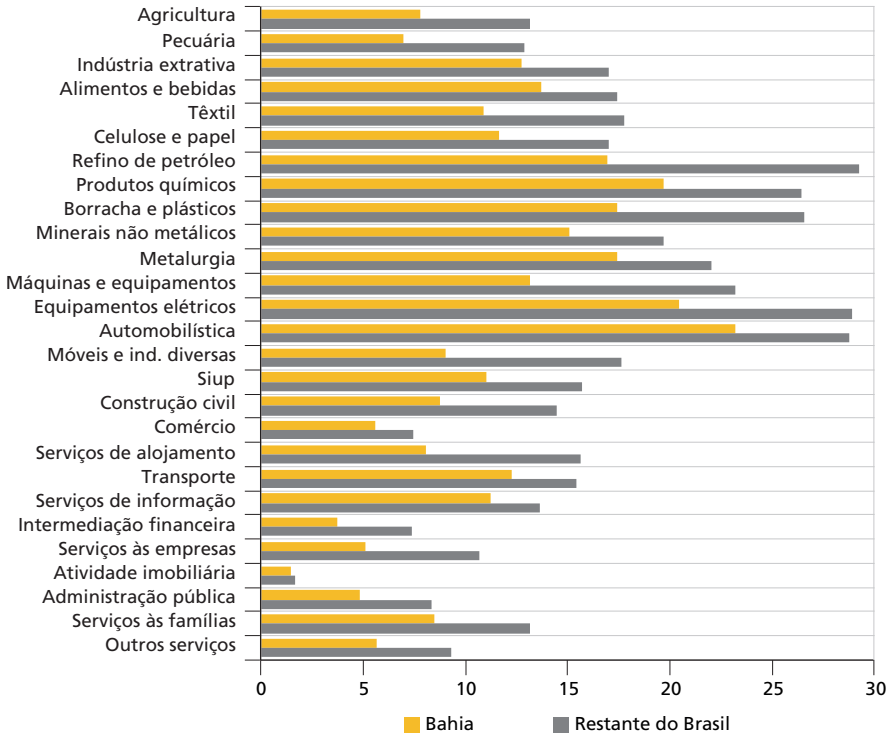
As estimativas da participação do valor adicionado externo nas exportações na Bahia e no restante do Brasil foram calculadas a partir da equação (8). Em média, 14,1% do valor adicionado às exportações baianas possuem origem externa ao país, enquanto para o restante do Brasil, em média, 17,5% das exportações eram compostas por valor adicionado incorporado em etapas anteriores da produção, a partir da importação de insumos intermediários.

O gráfico 7 mostra os resultados do valor adicionado externo nas exportações por setor de atividade. A maior parte da produção ocorre em território nacional; assim, as atividades econômicas possuem menor proporção de valor adicionado estrangeiro nas exportações, comparado ao valor adicionado doméstico. As indústrias de transformação foram as que apresentaram a maior proporção de valor adicionado externo, com destaque para as indústrias de refino de petróleo, produtos químicos, borracha e plástico, materiais elétricos e automobilística. A participação do valor adicionado estrangeiro mais elevado nesses setores sugere, especialmente, maior necessidade de se incorporar tecnologia produzida fora do país nessas atividades. Além disso, tais atividades também podem estar relacionadas à maior presença de multinacionais, que realizam apenas parte de sua produção no Brasil, gerando, assim, a fragmentação internacional da produção por meio do comércio intraindústria.

Em termos nacionais, a inserção da economia brasileira nas cadeias globais de valor ainda é pequena, quando comparada a outros países. O valor adicionado estrangeiro presente nas exportações brasileira para o ano de 2009, calculado neste estudo, foi de 17,4%. Nas estimativas da TiVA esse valor foi de 20,4% para o mesmo ano (OECD, 2016). Essa diferença é justificada pelo uso das matrizes de insumo-produto no contexto inter-regional global e com abertura setorial diferente. As estimativas da TiVA para outros países foram: Alemanha (24,6%), Estados Unidos (23,8%), França (23,2), Rússia (22,9%), Reino Unido (22,6%), China (22,8%), México (21,9%), África do Sul (21,1%), Japão (19,9%) e Índia (19,1%).[13] Percebe-se que, em geral, os países com maior nível de desenvolvimento industrial são mais integrados às cadeias globais de valor.

13. Os valores calculados pela TiVA para o ano de 2009 podem ser considerados atípicos devido aos efeitos da crise internacional de 2008. A escolha de 2009 foi feita para fins de comparação com os resultados deste estudo. No entanto, os valores dessa estatística para o ano de 2011 apresentam classificação semelhante para a integração dos países nas cadeias de valor: Brasil (23,4%), Alemanha (27,3%), Estados Unidos (25,4%), França (26,3), Rússia (26,4%), Reino Unido (25,2%), China (24,1%), México (25,5%), África do Sul (23,3%), Japão (22,0%) e Índia (19,3%).

GRÁFICO 7
Participação do valor adicionado externo nas exportações por setor: Bahia e restante do Brasil (2009)
(Em %)



Elaboração dos autores.

A segmentação da produção em diferentes regiões também pode ocorrer dentro do mesmo território. No entanto, a fragmentação produtiva inter-regional pode apresentar dinâmica diferente da fragmentação internacional. Para identificar essas diferenças, a próxima subseção analisa o conteúdo importado e o valor adicionado externo na economia baiana considerando a importação de insumos do restante do Brasil e do restante do mundo.

5.2 Fragmentação inter-regional

Nesta subseção, as medidas de conteúdo importado e valor adicionado nas exportações foram contabilizadas considerando exógenos o comércio com o restante do Brasil e com o restante do mundo, na modelagem de insumo-produto. Assim, a inversa

de Leontief foi calculada apenas com os fluxos de comércio inter-setorial com origem e destino dentro do estado da Bahia – $Z_{ij}^{BA,BA}$ do quadro A.1, no apêndice. Os conceitos de importação e de valor adicionado estrangeiro levam, portanto, em consideração o fluxo de insumos importados pela Bahia no comércio internacional e inter-regional. As exportações são identificadas a partir do fluxo de comércio inter-regional (para o restante do Brasil) e internacional (para outros países). Esses resultados são diferentes da seção anterior, no qual essas medidas referiam-se exclusivamente ao comércio internacional.

A importância das exportações, para o restante do Brasil e restante do mundo, no PIB da economia baiana é de 39,3% (tabela 2). No entanto, esse valor não representa a contribuição real das exportações, visto que 20,2% das exportações baianas são de insumos intermediários importados, utilizados no processo produtivo e reexportados. Assim, retirando-se o conteúdo importado, as exportações líquidas do estado da Bahia representaram 31,3% do seu PIB no ano de 2009.

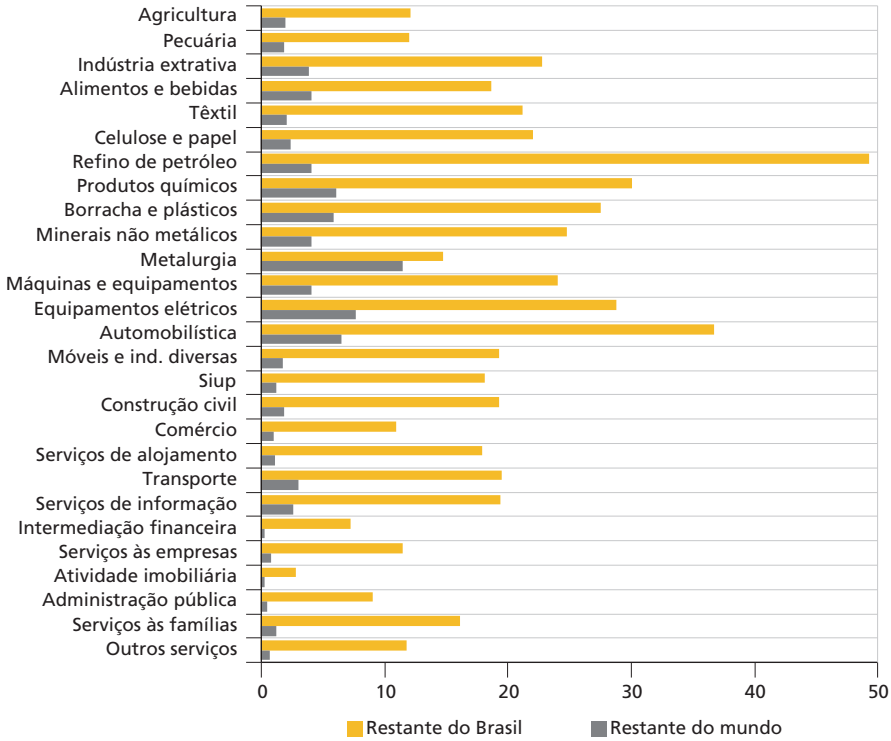
TABELA 2
Exportações e conteúdo importado: Bahia (2009)
(Em %)

Região	Participação no PIB			Participação de conteúdo importado nas exportações totais
	Exportações totais	Conteúdo importado	Exportações líquidas	
Restante do Brasil	28,3	7,4	20,8	26,3
Restante do mundo	11,0	0,5	10,5	4,6
Total	39,3	7,9	31,3	20,2

Elaboração dos autores.

Visto que a cadeia de produção pode ser fragmentada tanto em nível global quanto em nível regional, algumas indústrias localizam suas atividades de montagem perto dos mercados finais, enquanto outras indústrias são menos concentradas regionalmente. Portanto, a fragmentação da produção pode variar consideravelmente entre os diferentes tipos de atividades econômicas. No caso da Bahia, gráfico 8, a produção de bens e de serviços depende, principalmente, de insumos importados do restante do Brasil, revelando, assim, maior interação em termos inter-regionais.

GRÁFICO 8
Participação do conteúdo importado nas exportações por setor e origem do conteúdo importado: Bahia (2009)
(Em %)



Elaboração dos autores.

A tabela 3 apresenta as participações das exportações totais (restante do Brasil e restante do mundo) e do conteúdo importado do estado da Bahia em relação à produção setorial. Os setores nos quais as exportações mais contribuem para a produção total são as indústrias de celulose e papel (70,7%), automóveis (59,5%), metalurgia (59,8%) e de produtos químicos (55,3%). No entanto, descontado os conteúdos importados, que posteriormente são exportados, as exportações líquidas desses setores contabilizam 53,5%, 33,8%, 44,2% e 35,3% da produção setorial, respectivamente.

Na Bahia, a metalurgia é o setor que contém a maior quantidade de conteúdo importado (11,4%) do restante do mundo nas exportações (gráfico 8). O setor destaca-se principalmente pela produção de ferroligas, cobre e produção de

aço, com 35,1% das exportações destinadas ao mercado internacional e 64,9% destinadas para o restante do Brasil (tabela 3). Esse setor também concentra a maior participação no PIB industrial do estado (18,6%).

O valor adicionado externo nas exportações do estado foi estimado em 35,3% (tabela 3). As maiores participações do valor adicionado externo foram identificadas nas indústrias de refino de petróleo (57,4%) e automobilística (51,3%). Essas indústrias possuem exportações voltadas principalmente para o restante do Brasil (80% e 82%, respectivamente), sendo integradas em uma cadeia de produção com ligações mais fortes no contexto nacional do que internacional.

TABELA 3
Exportações, conteúdo importado e valor adicionado estrangeiro por setor: Bahia (2009)
(Em %)

Setor	Destino das exportações		Participação no produto total				Valor adicionado estrangeiro nas exportações
			Exportações totais	Conteúdo importado		Exportações líquidas	
	Restante do Brasil	Restante do mundo		Restante do Brasil	Restante do mundo		
1. Agricultura	28,46	71,54	22,18	2,67	0,44	19,07	17,06
2. Pecuária	70,81	29,19	4,76	0,57	0,09	4,10	16,48
3. Indústria extrativa	76,42	23,58	26,25	5,97	1,01	19,26	31,27
4. Alimentos e bebidas	68,50	31,50	26,37	4,92	1,07	20,38	28,82
5. Têxtil	40,34	59,66	28,25	5,98	0,59	21,68	28,05
6. Celulose e papel	16,71	83,29	70,71	15,58	1,68	53,45	29,37
7. Refino de petróleo	80,10	19,90	37,42	18,46	1,51	17,45	57,35
8. Produtos químicos	75,19	24,81	55,29	16,64	3,38	35,26	43,50
9. Borracha e plásticos	78,60	21,40	46,93	12,90	2,74	31,29	38,64
10. Minerais não metálicos	89,78	10,22	11,39	2,82	0,46	8,10	34,92
11. Metalurgia	64,94	35,06	59,79	8,79	6,83	44,17	29,09
12. Máquinas e equipamentos	88,21	11,79	21,82	5,24	0,89	15,69	32,02
13. Equipamentos elétricos	92,26	7,74	30,15	8,67	2,32	19,17	42,70

(Continua)

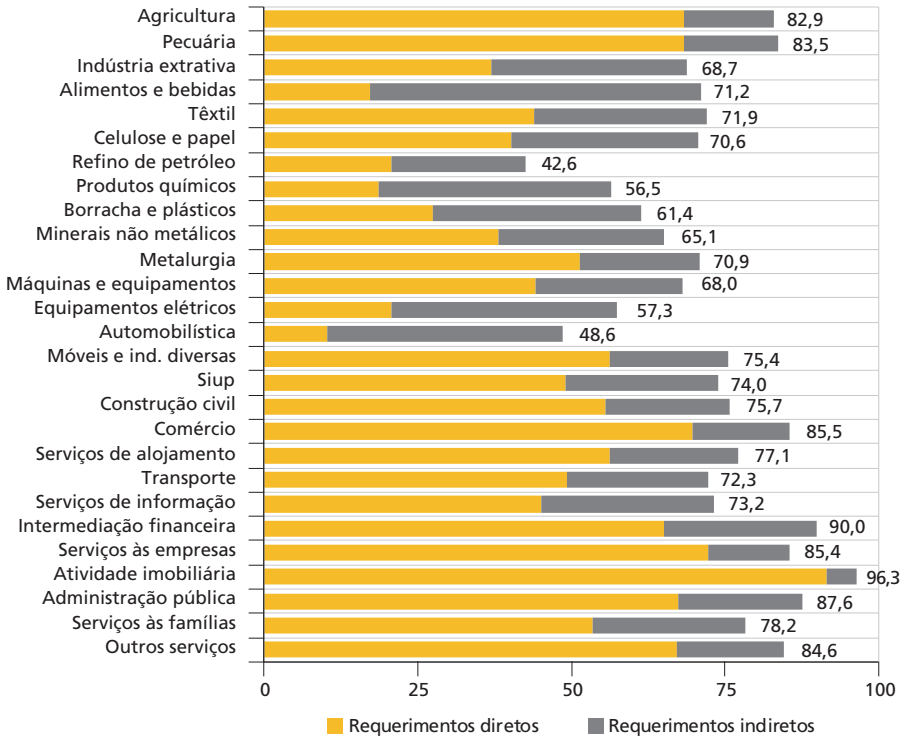
(Continuação)

Setor	Destino das exportações		Participação no produto total				Valor adicionado estrangeiro nas exportações
			Exportações totais	Conteúdo importado		Exportações líquidas	
	Restante do Brasil	Restante do mundo		Restante do Brasil	Restante do mundo		
14. Automobilística	82,29	17,71	59,54	21,84	3,88	33,81	51,38
15. Móveis e indústrias diversas	93,77	6,23	20,13	3,89	0,35	15,90	24,56
16. Siup	100,00	0,00	1,80	0,33	0,02	1,45	26,04
17. Construção civil	100,00	0,00	0,74	0,14	0,01	0,58	24,25
18. Comércio	84,13	15,87	27,91	3,04	0,29	24,58	14,48
19. Serviços de alojamento	80,17	19,83	0,87	0,16	0,01	0,70	22,93
20. Transporte	94,54	5,46	24,98	4,87	0,75	19,36	27,67
21. Serviços de informação	99,65	0,35	3,33	0,65	0,09	2,60	26,79
22. Intermediação financeira	100,00	0,00	0,01	0,00	0,00	0,01	10,03
23. Serviços às empresas	99,93	0,07	0,92	0,11	0,01	0,81	14,56
24. Atividade imobiliária	99,38	0,62	0,15	0,00	0,00	0,14	3,73
25. Administração pública	89,34	10,66	0,17	0,02	0,00	0,16	12,41
26. Serviços às famílias	96,18	3,82	0,34	0,06	0,00	0,28	21,83
27. Outros serviços	100,00	0,00	0,00	0,00	0,00	0,00	15,43
Total	**71,97**	**28,03**	**20,15**	**5,16**	**0,95**	**14,05**	**35,28**

Elaboração dos autores.

O valor adicionado doméstico presente na produção setorial na Bahia, gráfico 9, denota a importância da economia local na agregação de valor dos bens produzidos pelo estado. No total, 64,7% do valor adicionado nos bens produzidos na Bahia foram gerados dentro do estado, sendo os requerimentos indiretos (28,3%) menores em relação aos requerimentos diretos (36,4%). Essa medida apresentou elevada variação setorial. Enquanto o valor adicionado originado na economia doméstica na indústria automobilista foi de 48,6% (com a menor proporção de requerimentos diretos, apenas 10,8%), na atividade imobiliária, 96,3% do valor adicionado foi formado por atividade doméstica, com 91,5% de requerimentos diretos.

GRÁFICO 9
Participação do valor adicionado doméstico por setor: Bahia (2009)
(Em %)



Elaboração dos autores.

Os resultados acerca dos fluxos de comércio inter-regional permitem compreender melhor sobre a importância relativa das exportações no contexto de uma região dentro do território nacional. A análise possibilita identificar setores que podem perder vantagem relativa após considerar apenas as exportações líquidas ou o valor adicionado doméstico, dado o erro em contabilizar parte da produção de outras regiões (restante do mundo e restante do Brasil) como sendo da estrutura produtiva do estado.

No caso baiano, é possível citar perdas de vantagens relativas para as indústrias de refino de petróleo e automobilística (que representam 9,6% e 5,2% das exportações estaduais, respectivamente – gráfico 5), nas quais a menor parcela das suas exportações incorporam atividades produzidas dentro do estado da Bahia (42,6% e 48,6%, respectivamente – gráfico 9). No entanto, as indústrias de celulose e papel e de metalurgia, que também se destacam nas exportações (17,7% e 11,9%,

respectivamente – gráfico 5), possuem a maior parte de suas exportações composta por bens e serviços produzidos na Bahia (70,6% e 70,9%, respectivamente – gráfico 9). Portanto, uma vez que seja levada em consideração a produção doméstica, as exportações relacionadas às atividades de refino de petróleo e automobilística poderiam tornar-se menos importantes em termos de geração de renda para o estado. Entretanto, nessa análise de vantagens relativas decorrentes da integração às cadeias globais de valor, devemos considerar os *linkages* de insumo-produto entre os setores de atividade. Nessa perspectiva, os resultados deste trabalho são complementares ao estudo realizado por Perobelli *et al.* (2015), que apresentam uma análise completa dos encadeamentos produtivos na estrutura industrial da Bahia.

6 CONSIDERAÇÕES FINAIS

O processo de globalização tem levado a mudanças na estrutura de comércio internacional, gerando alterações no fluxo de mercadorias e de serviços entre os países. O desenvolvimento de economias periféricas, como a brasileira, está ocorrendo em um ambiente de crescente integração econômica internacional. Para sustentar o seu crescimento, faz-se necessário inserir-se, de forma competitiva, nos fluxos internacionais de comércio e de investimento.

O presente estudo teve como objetivo analisar a inserção produtiva no contexto da fragmentação internacional e inter-regional da produção a partir da estrutura de exportações do estado da Bahia, em particular, e do restante do Brasil. Para tal, utilizou-se a metodologia de insumo-produto, o que tornou possível medir o conteúdo doméstico e a participação do valor adicionado doméstico nas exportações dessas duas regiões.

Os resultados deste estudo ampliam a compreensão sobre a importância do comércio global e regional para as economias brasileira e baiana. Pode-se verificar que a integração do estado da Bahia com o restante do Brasil é maior do que com o restante do mundo. Essas duas regiões apresentam pouca inserção no comércio exterior com elevada heterogeneidade setorial. As indústrias de transformação, em especial refino de petróleo, produtos químicos, borracha e plástico, materiais elétricos e automobilística, são as mais integradas às cadeias globais de valor.

Tomando por base a importância do mercado externo no contexto do crescimento regional, mapear a fragmentação produtiva e as interdependências da Bahia revela alternativas de crescimento para o estado. Isso permite delinear políticas setoriais específicas e, assim, maximizar os resultados para essa região em termos de aumento de renda, de produção e maior participação nas cadeias produtivas.

Importante salientar que a abordagem deste trabalho é inovadora, pois lança luz sobre a estrutura de inserção de uma Unidade da Federação, dentro do novo contexto de análise proposto pela literatura. Não coube aqui fazer avaliações sobre

vantagens e/ou desvantagens de estar mais integrado à economia do restante do Brasil ou ao restante do mundo, mas mapear o conteúdo de integração em seus diversos aspectos, a fim de subsidiar possíveis tomadas de decisão por parte dos formuladores de política. Ademais, nota-se a necessidade de a produção nacional estar conectada aos mercados internacionais. Nesse sentido, é importante identificar quais são as indústrias que proporcionam maior entrada de divisas, e, daí, planejar esforços e investimentos que resultem na maior geração de renda para o estado, a exemplo daquelas apontadas no estudo.

Pesquisas futuras podem acompanhar a evolução da integração do estado da Bahia e do restante do Brasil nas cadeias de produção nacional e global. Desse modo, poder-se-ia avaliar se a agregação de valor pela economia local tem aumentado ou diminuído ao longo do tempo. Pesquisas futuras também podem levar em consideração os países de origem e de destino dos fluxos de comércio, o que possibilitaria analisar se a fragmentação da produção é essencialmente regional, ocorrendo principalmente entre um grupo de países vizinhos e blocos comerciais regionais, ou essencialmente global, envolvendo também países de fora da América do Sul. Esta é uma questão com implicações importantes para os formuladores de políticas comerciais, pois, sendo as cadeias de produção fragmentadas dentro de países vizinhos, acordos regionais de comércio poderiam ser suficientes para aumentar a inserção na cadeia de valor, enquanto as cadeias de valor essencialmente globais exigiriam acordos multirregionais.

REFERÊNCIAS

ARAÚJO, I. F.; PEROBELLI, F. S.; FARIA, W. R. Fragmentação internacional da produção e a inserção do Brasil em cadeias globais de valor. *In:* ENCONTRO NACIONAL DE ECONOMIA, 44., 2016, Foz do Iguaçu. **Anais** [...]. Foz do Iguaçu: Associação Nacional dos Centros de Pós-Graduação em Economia, 2016.

BACEN – BANCO CENTRAL DO BRASIL. Grau de abertura ao comércio externo: uma análise regional. **Boletim Regional do Banco Central do Brasil**, Brasília, p. 91-95, 2013.

BAHIA. Superintendência de Estudos Econômicos e Sociais da Bahia. **Tabela de recursos e usos do estado da Bahia 2009**. Salvador: SEI, 2013. v. 1.

_____. _____. **Bahia em números 2013**. Salvador: SEI, 2014. v. 13.

BALDWIN, R. E. **Globalisation**: the great unbundling(s). Globalisation challenges for Europe. Helsinki: Office of the Prime Minister of Finland, 2006.

_____. **Trade and industrialisation after globalisation's 2nd unbundling**: how building and joining a supply chain are different and why it matters. Cambridge: National Bureau of Economic Research, 2011. (Working Paper, n. 17716).

BALDWIN, R. E.; EVENETT, S. J. Value creation and trade in 21st-century manufacturing. **Journal of Regional Science**, v. 55, n. 1, p. 31-50, 2015.

BEMS, R.; JOHNSON, R. C.; YI, K. M. Vertical linkages and the collapse of global trade. **American Economic Review**, v. 101, p. 308-312, 2011.

BRENE, P. R. A. **Estimação de matrizes insumo-produto municipais por meio do quociente locacional**: limites e possibilidades da proposta metodológica. 2013. Tese (Doutorado em Economia) – Universidade Federal do Paraná, Curitiba, 2013.

BROWN, W. M.; ANDERSON, W. P. The influence of industrial and spatial structure on Canada-U.S. Regional trade. **Growth and Change**, v. 30, n. 1, p. 23-47, 1999.

CHEN, X. *et al.* **The estimation of domestic value-added and employment induced by exports**: an application to Chinese exports to the United States. Stanford: Stanford University, 2004.

CHEN, X. *et al.* Domestic value added and employment generated by Chinese exports: a quantitative estimation. **China Economic Review**, v. 23, n. 4, p. 850-864, 2012.

COSTINOT, A.; VOGEL, J.; WANG, S. An elementary theory of global supply chains. **Review of Economic Studies**, v. 80, n. 1, p. 109-144, 2013.

DIETZENBACHER, E. *et al.* The construction of world input-output tables in the WIOD Project. **Economic Systems Research**, v. 25, n. 1, p. 71-98, 2013.

DIETZENBACHER, E.; GUILHOTO, J. J. M.; IMORI, D. **The role of Brazilian regions in the global value chain**. São Paulo: Núcleo de Economia Regional e Urbana da Universidade de São Paulo, 2013. (Texto para Discussão Nereus, n. 13).

DOMINGUES, E. P.; HADDAD, E. A. Política tributária e re-localização. **Revista Brasileira de Economia**, v. 57, n. 4, p. 849-871, 2003.

FLAM, H.; FLANDERS, M. J. **Heckscher-Ohlin trade theory**. Cambridge: The MIT Press, 1991.

FERRAZ, L.; GUTIERRE, L.; CABRAL, R. A. indústria brasileira na era das cadeias globais de valor. *In:* BARBOSA, N. *et al.* (Eds.). **Indústria e desenvolvimento produtivo no Brasil**. Rio de Janeiro: Elsevier, 2016.

FUJII-GAMBERO, G.; CERVANTES-MARTÍNEZ, R. The weak linkages between processing exports and the internal economy: the Mexican case. **Economic Systems Research**, v. 29, n. 4, p. 1-13, 2017.

FUJITA, M.; HAMAGUCHI, N. Supply chain internationalization in East Asia: inclusiveness and risks. **Papers in Regional Science**, v. 95, n. 1, p. 81-100, 2016.

GEREFFI, G.; HUMPHREY, J.; STURGEON, T. The governance of global value chains. **Review of International Political Economy**, v. 12, n. 1, p. 78-104, 2005.

GROSSMAN, G. M.; ROSSI-HANSBERG, E. Trading tasks: a simple theory of offshoring. **American Economic Review**, v. 98, n. 5, p. 1978-1997, 2008.

GUILHOTO, J. J. M.; IMORI, D. **Brazilian role in the global value chains**. São Paulo: USP, 2014. (Working Paper Series, n. 2014-24).

GUILHOTO, J. J. M.; SESSO FILHO, U. A. Estimação da matriz insumo-produto a partir de dados preliminares das contas nacionais. **Economia Aplicada**, v. 9, n. 2, p. 277-299, 2005.

_____; _____. Estimação da matriz insumo-produto utilizando dados preliminares das contas nacionais: aplicação e análise de indicadores econômicos para o Brasil em 2005. **Economia e Tecnologia**, v. 23, p. 53-62, 2010.

HADDAD, E. A. *et al.* Macroeconomia dos estados e matriz interestadual de insumo-produto. **Revista Economia Aplicada**, v. 6, n. 4, p. 875-895, 2002.

HADDAD, E. A. *et al.* Assessing the ex ante economic impacts of transportation infrastructure policies in Brazil. **Journal of Development Effectiveness**, v. 3, n. 1, p. 44-61, 2011.

HADDAD, E. A.; DOMINGUES, E. P. **Matriz inter-regional de insumo-produto**: São Paulo/resto do Brasil. São Paulo: USP, 2003. (Texto para Discussão Nereus, n. 10-2003).

HADDAD, E. A.; PEROBELLI, F. S. Integração regional e padrão de comércio dos estados brasileiros. *In:* ANITA, K. (Org.). **Unidade e fragmentação**: a questão regional no Brasil. São Paulo: Perspectiva, 2002.

HULU, E.; HEWINGS, G. J. D. The development and use of interregional input-output models for Indonesia under conditions of limited information. **Review of Urban and Regional Development Studies**, v. 5, n. 2, p. 135-153, 1993.

HUMMELS, D.; ISHII, J.; YI, K.-M. The nature and growth of vertical specialization in world trade. **Journal of International Economics**, v. 54, n. 1, p. 75-96, 2001.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Sistema de Contas Nacionais do Brasil – referência 2000**. Rio de Janeiro: IBGE, 2011.

_____. **Sistema de Contas Regionais do Brasil 2012**. Rio de Janeiro: IBGE, 2014.

_____. **Sistema de Contas Regionais do Brasil 2015**. Rio de Janeiro: IBGE, 2017.

IMORI, D. **Brazilian regions in the global value chain**: trade and the environment. 2016. Thesis (Economics Doctorate Degree) – Universidade de São Paulo, São Paulo, 2016.

ITO, T.; ROTUNNO, L.; VÉZINA, P. L. Heckscher–Ohlin: evidence from virtual trade in value added. **Review of International Economics**, v. 25, n. 3, p. 427-446, 2017.

JOHNSON, R. C.; NOGUERA, G. Accounting for intermediates: production sharing and trade in value added. **Journal of International Economics**, v. 86, n. 2, p. 224-236, 2012.

JONES, R. **Globalization and the theory of input trade**. Cambridge: MIT Press, 2000.

KOOPMAN, R.; WANG, Z.; WEI, S. J. Estimating domestic content in exports when processing trade is pervasive. **Journal of Development Economics**, v. 99, n. 1, p. 178-189, 2012.

_____; _____; _____. Tracing value-added and double counting in gross exports. **American Economic Review**, v. 104, n. 2, p. 459-494, 2014.

LOS, B.; TIMMER, M. P.; DE VRIES, G. J. How global are global value chains? A new approach to measure international fragmentation. **Journal of Regional Science**, v. 55, n. 1, p. 66–92, 2015.

_____; _____; _____. Tracing value-added and double counting in gross exports: comment. **The American Economic Review**, v. 106, n. 7, p. 1958-1966, 2016.

MILLER, R. E.; BLAIR, P. D. **Input-output analysis**: foundations and extensions. Cambridge: Cambridge University Press, 2009.

MILLER, R. E.; TEMURSHOEV, U. Output upstreamness and input downstreamness of industries/countries in world production. **International Regional Science Review**, v. 40, n. 5, p. 443-475, 2017.

OECD – ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). **TiVA – Trade in Value Added**. Paris: OECD, 2016.

OLIVEIRA, I. T. M.; REIS, C. F. B.; BLOCH, C. D. A inserção do Brasil no comércio internacional de serviços e suas relações com cadeias globais de valor. *In:* OLIVEIRA, I. T. M.; CARNEIRO, F. L.; SILVA FILHO, E. S. (Orgs.). **Cadeias globais de valor, políticas públicas e desenvolvimento**. Brasília: Ipea, 2017.

PEROBELLI, F. S. *et al*. Estrutura de interdependência inter-regional no Brasil: uma análise espacial de insumo-produto para os anos de 1996 e 2002. **Pesquisa e Planejamento Econômico**, v. 40, n. 2, p. 281-326, 2010.

PEROBELLI, F. S. *et al.* Estimativa da matriz de insumo-produto da Bahia (2009): características sistêmicas da estrutura produtiva do estado. **Revista Econômica do Nordeste**, v. 46, n. 4, p. 97-116, 2015.

PEROBELLI, F. S.; HADDAD, E. A. Padrões de comércio interestadual no Brasil, 1985 e 1997. **Revista de Economia Contemporânea**, v. 10, n. 1, p. 61-88, 2006a.

_____; _____. Exportações internacionais e interações regionais: uma análise de equilíbrio geral. **Revista de Estudos Econômicos**, v. 36, p. 833-866, 2006b.

PEROBELLI, F. S.; HADDAD, E. A.; DOMIGUES, E. P. Interdependence among the Brazilian states: an input-output approach. *In:* FARSHCHI, M. A.; JANNE, O. E. M.; MCCANN, P. (Eds.). **Technological change and mature industrial regions**. Northampton: Edward Elgar Publishing, 2009.

PORSSE, A. A.; HADDAD, E. A.; RIBEIRO, E. P. **Estimando uma matriz de insumo-produto inter-regional Rio Grande do Sul-restante do Brasil**. São Paulo: USP, 2003. (Texto para Discussão Nereus, n. 20-2003).

PRETE, D. D.; RUNGI, A. Organizing the global value chain: a firm-level test. **Journal of International Economics**, v. 109, p. 16-30, 2017.

RÉBORA, S.; VIVANCO, D. Domestic value-added and import content of exports: evidence from the Chilean input-output tables 2008-2012. *In:* INTERNATIONAL INPUT-OUTPUT CONFERENCE, 23., 2015, México. **Annals** [...]. México: International Input-Output Association, 2015.

RICHARDSON, H. W. **Regional economics**: location theory, urban structure and regional change. New York: Pareger Publisher, 1969.

THORSTENSEN, V.; FERRAZ, L. O isolamento do Brasil em relação aos acordos e mega-acordos comerciais. **Boletim de Economia e Política Internacional**, n. 16, p. 5-17, 2014.

TIMMER, M. P. *et al.* Slicing up global value chains. **Journal of Economic Perspectives**, v. 28, n. 2, p. 99-18, 2014.

TUKKER, A.; DIETZENBACHER, E. Global multiregional input–output frameworks: an introduction and outlook. **Economic Systems Research**, v. 25, n. 1, p. 1-19, 2013.

VEIGA, P. M.; RIOS, S. P. Inserção em cadeias globais de valor e políticas públicas: o caso do Brasil. *In:* OLIVEIRA, I. T. M.; CARNEIRO, F. L.; SILVA FILHO, E. S. (Orgs.). **Cadeias globais de valor, políticas públicas e desenvolvimento**. Brasília: Ipea, 2017.

YI, K.-M. Can vertical specialization explain the growth of world trade? **Journal of Political Economy**, v. 111, n. 1, p. 52-102, 2003.

APÊNDICE

QUADRO A.1

Estrutura da tabela de usos e recurso (setor x setor) com as transações inter-regionais da Bahia e do restante do Brasil

Insumo		Produto	Insumos intermediários domésticos						Demanda final (E+C+G+I)		Demanda total
			Bahia (BA)			Restante do Brasil (RB)					
			1	...	N	1	...	N	Bahia	RB	
Insumos intermediários domésticos	Bahia (BA)	1	$Z_{ij}^{BA,BA}$			$Z_{ij}^{BA,RB}$			$Y_j^{BA,BA}$	$Y_j^{BA,BR}$	X_j^{BA}
		⋮									
		N									
	Restante do Brasil (RB)	1	$Z_{ij}^{RB,BA}$			$Z_{ij}^{RB,RB}$			$Y_j^{RB,BA}$	$Y_j^{RB,RB}$	X_j^{RB}
		⋮									
		N									
Insumos intermediários produzidos fora do Brasil		1	$Z_{ij}^{M,BA}$			$Z_{ij}^{M,RB}$			$Y_j^{M,BA}$	$Y_j^{M,RB}$	M_j
		⋮									
		N									
Valor adicionado			VA_i^{BA}			VA_i^{RB}					
Valor bruto da produção			X_i^{BA}			X_i^{RB}					

Fonte: MIP-IR BA-RB (2009), disponível em: http://www.ufjf.br/lates/.

Elaboração dos autores.

Obs.: Os insumos importados podem ter origem no comércio inter-regional (*R*), $Z_{ij}^{RB,BA}$ e $Z_{ij}^{BA,RB}$, ou no comércio internacional (*M*), $Z_{ij}^{M,BA}$ e $Z_{ij}^{M,RB}$.

ANEXO

QUADRO A.1
Atividades na matriz de insumo-produto inter-regional da Bahia e restante do Brasil

Setor	Descrição
1. Agricultura	Agricultura e silvicultura
2. Pecuária	Pecuária e pesca
3. Indústria extrativa	Indústria extrativa
4. Alimentos e bebidas	Alimentos, bebidas e fumo
5. Têxtil	Têxtil, vestuário e couros
6. Celulose e papel	Celulose, produção de papel e de madeira – exclusive móveis
7. Refino de petróleo	Produtos do refino de petróleo e coque
8. Produtos químicos	Produtos químicos
9. Borracha e plásticos	Borracha e plásticos
10. Minerais não metálicos	Cimento e outros produtos de minerais não metálicos
11. Metalurgia	Metalurgia
12. Máquinas e equipamentos	Máquinas e equipamentos, inclusive manutenção e reparos
13. Equipamentos elétricos	Máquinas, aparelhos, materiais elétricos e equipamentos de informática
14. Automobilística	Automóveis, camionetas e utilitários, caminhões e ônibus e peças e acessórios
15. Móveis e indústrias diversas	Móveis e produtos das indústrias diversas
16. Siup	Produção e distribuição de eletricidade, gás, água, esgoto e limpeza urbana
17. Construção civil	Construção civil
18. Comércio	Comércio e serviços de manutenção e reparação
19. Serviços de alojamento	Serviços de alojamento e alimentação
20. Transporte	Transporte, armazenagem e correio
21. Serviços de informação	Serviços de informação
22. Intermediação financeira	Intermediação financeira, seguros e previdência complementar e serviços relacionados
23. Serviços às empresas	Serviços prestados às empresas
24. Atividade imobiliária	Atividades imobiliárias e aluguéis
25. Administração pública	Administração, saúde e educação públicas e seguridade social
26. Serviços às famílias	Serviços prestados às famílias e associativas
27. Outros serviços	Outros serviços

Fonte: MIP-IR BA-RB (2009).

NOTA AOS COLABORADORES DE PESQUISA E PLANEJAMENTO ECONÔMICO

1. A revista só analisa, com vistas a eventual publicação, artigos com conteúdo inédito, tanto no país quanto no exterior. Além disso, o seu tema deve se inserir em uma das áreas da ciência econômica, contribuindo de modo significativo ao avanço do conhecimento científico nessa área.

2. Resenhas de livros recentemente publicados poderão ser consideradas para publicação, mas resenhas temáticas e os textos essencialmente descritivos não serão, de um modo geral, aceitos.

3. As contribuições não serão remuneradas, e a submissão de um artigo à revista implica a transferência dos direitos autorais ao Ipea, caso ele venha a ser publicado.

4. Em geral, os artigos submetidos à revista devem ser escritos em português. Em casos excepcionais, poderão ser recebidos textos em língua inglesa para análise, mas se ele vier a ser aceito para publicação, o autor deverá se responsabilizar por sua tradução.

5. Só serão publicados artigos em português, mas sua versão em inglês poderá ser disponibilizada no sítio da revista na internet. Os anexos muito longos ou complexos para serem publicados, bem como as bases de dados necessárias para reproduzir os resultados empíricos do trabalho, serão também oferecidos aos leitores em versão virtual.

6. Caso o trabalho seja aceito para publicação, cada autor receberá 3 (três) exemplares do número da revista correspondente.

7. Para submeter um trabalho à revista, o autor deve acessar a página de Pesquisa e Planejamento Econômico na internet, em http://ppe.ipea.gov.br/index.php/ppe/index, e seguir os procedimentos ali descritos.

8. Os artigos recebidos que estejam de acordo com as instruções acima serão avaliados pelo Corpo Editorial com o auxílio de dois pareceristas escolhidos pelo Editor. O trabalho dos pareceristas é feito observando-se o método duplamente cego: o autor não saberá quem são os pareceristas, nem estes quem é o autor. Dessa análise poderá resultar a aceitação do artigo, condicionada, ou não, à realização de alterações; sua rejeição, com ou sem a recomendação de nova submissão após modificações; ou a rejeição definitiva. No caso de uma segunda submissão, o artigo será novamente avaliado por pareceristas, podendo vir a ser enquadrado em qualquer das situações acima. A rotina de análise se repete até que uma decisão final de rejeição ou aceitação seja alcançada. O processamento do artigo é conduzido pelo Editor, a quem cabe também a comunicação com os autores.

9. A decisão final quanto à publicação dos artigos cabe ao Corpo Editorial, que se reúne ordinariamente para decidir a composição de cada um dos números da revista, por recomendação do Editor. A aprovação do artigo para publicação só então é comunicada aos autores dos artigos respectivos, por escrito.

Ipea – Instituto de Pesquisa Econômica Aplicada

Assessoria de Imprensa e Comunicação

EDITORIAL

Coordenação
Ipea

Revisão e editoração
Editorar Multimídia

Capa
Luís Cláudio Cardoso da Silva

Projeto Gráfico
Renato Rodrigues Bueno

The manuscripts in languages other than
Portuguese published herein have not been proofread.

Composto em Adobe Garamond Pro 11 (texto)
Frutiger 67 Bold Condensed (títulos, gráficos e tabelas)
Brasília-DF

Missão do Ipea
Aprimorar as políticas públicas essenciais ao desenvolvimento brasileiro por meio da produção e disseminação de conhecimentos e da assessoria ao Estado nas suas decisões estratégicas.

ISSN 0100-0551

Apoio editorial
CNPq

ipea Instituto de Pesquisa Econômica Aplicada

MINISTÉRIO DA
ECONOMIA

PÁTRIA AMADA BRASIL
GOVERNO FEDERAL