

**IMPACTOS DO PROGRAMA BOLSA FAMÍLIA SOBRE A
DIVERSIFICAÇÃO DO CONSUMO DE ALIMENTOS NO BRASIL**

Aléssio Tony Cavalcanti de Almeida
Shirley Pereira de Mesquita
Magno Vamberto Batista da Silva

**EXPANSÃO DA ESCOLARIDADE E REDUÇÃO DA DESIGUALDADE
REGIONAL DE RENDA NO BRASIL ENTRE 1995 E 2011:
PROGRESSOS RECENTES E DESAFIOS PRESENTES**

Rodrigo Carvalho Oliveira
Raul da Mota Silveira Neto

**DETERMINANTES DAS MIGRAÇÕES INTERESTADUAIS NO BRASIL:
EVIDÊNCIAS A PARTIR DE UM MODELO GRAVITACIONAL**

Hilton Martins de Brito Ramalho
Erik Figueiredo
José Luis da Silva Netto Júnior

**CONCORRÊNCIA NO SETOR BANCÁRIO BRASILEIRO:
BANCOS INDIVIDUAIS VERSUS CONGLOMERADOS BANCÁRIOS**

Marcelo Cardoso
Paulo F. Azevedo
Klenio Barbosa

**INFLUÊNCIA DO AMBIENTE ECONÔMICO SOBRE A CONFIANÇA DO CONSUMIDOR
E AS DECISÕES DE CONSUMO NO BRASIL: UMA ANÁLISE EMPÍRICA
CONSIDERANDO OS EFEITOS DAS CREDIBILIDADES MONETÁRIA E FISCAL**

Gabriel Caldas Montes
Bruno Campos Visconti

**ELEIÇÃO NAS ESCOLAS: UMA ANÁLISE DO IMPACTO DO DIRETOR ELEITO
SOBRE O DESEMPENHO EDUCACIONAL NO ESTADO DA BAHIA**

Fernanda Julyanna Silva dos Santos
Raquel Menezes Bezerra Sampaio
Luciano Menezes Bezerra Sampaio

pesquisa e planejamento econômico - ppe

ipea

Governo Federal

Ministério do Planejamento, Desenvolvimento e Gestão

Ministro – Romero Jucá Filho



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

Presidente

Manoel Carlos de Castro Pires

Diretor de Desenvolvimento Institucional

Alexandre dos Santos Cunha

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

Roberto Dutra Torres Junior

Diretor de Estudos e Políticas Macroeconômicas

Mathias Jourdain de Alencastro

Diretor de Estudos e Políticas Regionais, Urbanas e Ambientais

Marco Aurélio Costa

Diretora de Estudos e Políticas Setoriais de Inovação, Regulação e Infraestrutura

Fernanda De Negri

Diretor de Estudos e Políticas Sociais, Substituto

José Aparecido Carlos Ribeiro

Diretor de Estudos e Relações Econômicas e Políticas Internacionais, Substituto

Cláudio Hamilton Matos dos Santos

Chefe de Gabinete

Fabio de Sá e Silva

Assessor-chefe de Imprensa e Comunicação

João Cláudio Garcia Rodrigues Lima

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

Daniilo 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** 2016

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 do Planejamento, Desenvolvimento e Gestão.

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

IMPACTOS DO PROGRAMA BOLSA FAMÍLIA SOBRE A DIVERSIFICAÇÃO DO CONSUMO DE ALIMENTOS NO BRASIL 7

Aléssio Tony Cavalcanti de Almeida

Shirley Pereira de Mesquita

Magno Vamberto Batista da Silva

EXPANSÃO DA ESCOLARIDADE E REDUÇÃO DA DESIGUALDADE REGIONAL DE RENDA NO BRASIL ENTRE 1995 E 2011: PROGRESSOS RECENTES E DESAFIOS PRESENTES 41

Rodrigo Carvalho Oliveira

Raul da Mota Silveira Neto

DETERMINANTES DAS MIGRAÇÕES INTERESTADUAIS NO BRASIL: EVIDÊNCIAS A PARTIR DE UM MODELO GRAVITACIONAL 67

Hilton Martins de Brito Ramalho

Erik Figueiredo

José Luis da Silva Netto Júnior

CONCORRÊNCIA NO SETOR BANCÁRIO BRASILEIRO: BANCOS INDIVIDUAIS *VERSUS* CONGLOMERADOS BANCÁRIOS 113

Marcelo Cardoso

Paulo F. Azevedo

Klenio Barbosa

INFLUÊNCIA DO AMBIENTE ECONÔMICO SOBRE A CONFIANÇA DO CONSUMIDOR E AS DECISÕES DE CONSUMO NO BRASIL: UMA ANÁLISE EMPÍRICA CONSIDERANDO OS EFEITOS DAS CREDIBILIDADES MONETÁRIA E FISCAL 147

Gabriel Caldas Montes

Bruno Campos Visconti

ELEIÇÃO NAS ESCOLAS: UMA ANÁLISE DO IMPACTO DO DIRETOR ELEITO SOBRE O DESEMPENHO EDUCACIONAL NO ESTADO DA BAHIA 177

Fernanda Julyanna Silva dos Santos

Raquel Menezes Bezerra Sampaio

Luciano Menezes Bezerra Sampaio

CONTENTS

- IMPACTS OF BOLSA FAMILIA PROGRAM ON THE DIVERSIFICATION OF FOOD CONSUMPTION IN BRAZIL** 7
Aléssio Tony Cavalcanti de Almeida
Shirley Pereira de Mesquita
Magno Vamberto Batista da Silva
- EDUCATION EXPANSION AND REDUCTION OF BRAZILIAN REGIONAL INCOME INEQUALITY BETWEEN 1995 AND 2011: RECENT PROGRESS AND CHALLENGES** 41
Rodrigo Carvalho Oliveira
Raul da Mota Silveira Neto
- A GRAVITY MODEL FOR INTER STATE MIGRATION IN BRAZIL** 67
Hilton Martins de Brito Ramalho
Erik Figueiredo
José Luis da Silva Netto Júnior
- COMPETITION IN THE BRAZILIAN BANKING SECTOR: INDIVIDUALIZED BANKS VERSUS BANKING CONGLOMERATES** 113
Marcelo Cardoso
Paulo F. Azevedo
Klenio Barbosa
- INFLUENCE OF THE ECONOMIC ENVIRONMENT ON CONSUMER CONFIDENCE AND CONSUMPTION DECISIONS IN BRAZIL: AN EMPIRICAL ANALYSIS CONSIDERING THE EFFECTS OF MONETARY CREDIBILITY AND FISCAL CREDIBILITY** 147
Gabriel Caldas Montes
Bruno Campos Visconti
- ELECTION IN SCHOOLS: AN ANALYSIS OF PRINCIPAL'S ELECTION ON THE STUDENTS' PROFICIENCY LEVELS IN THE STATE OF BAHIA** 177
Fernanda Julyanna Silva dos Santos
Raquel Menezes Bezerra Sampaio
Luciano Menezes Bezerra Sampaio

IMPACTOS DO PROGRAMA BOLSA FAMÍLIA SOBRE A DIVERSIFICAÇÃO DO CONSUMO DE ALIMENTOS NO BRASIL¹

Aléssio Tony Cavalcanti de Almeida²

Shirley Pereira de Mesquita³

Magno Vamberto Batista da Silva⁴

O objetivo do presente estudo é analisar o efeito do Programa Bolsa Família (PBF) sobre o nível de diversificação do consumo de bens alimentares entre as famílias beneficiadas residentes em áreas urbanas do Brasil, considerando inclusive propriedades nutricionais da cesta de consumo. A estratégia empírica baseia-se no método de *propensity score matching* e nos microdados da Pesquisa de Orçamentos Familiares (POF) 2008-2009. Por meio dessas informações são mensurados os efeitos médios de tratamento sobre duas medidas de diversificação do consumo de alimentos: o índice de Berry e o índice de diversificação saudável. Os resultados indicam que o PBF influencia positivamente o nível de diversificação do consumo de alimentos, mas esse efeito não se mostra significativo quando se considera o indicador que capta informações sobre a qualidade dos itens consumidos. Portanto, os incrementos de renda do programa para as famílias tratadas não geram melhorias em termos de uma dieta com mais itens saudáveis.

Palavras-chave: programa bolsa família; consumo alimentar; índice de Berry; diversificação saudável.

IMPACTS OF BOLSA FAMILIA PROGRAM ON THE DIVERSIFICATION OF FOOD CONSUMPTION IN BRAZIL

This paper aims to analyze the effect of the Bolsa Familia Program (PBF) on the level of diversification of food consumption among beneficiary families residing in urban areas of Brazil, also considering nutritional properties of the consumption basket. The empirical strategy is based on propensity score matching approach and on microdata from the Household Budget Survey (POF) 2008-2009. Through these information are measured average treatment effects on two measures of diversification of food consumption: the Berry index and the healthy diversification index. The results indicate that the PBF influences positively the level of diversification of food intake, but this effect does not appear significant when considering the indicator that captures information about the quality of items consumed. Therefore, increments the program's income for treated families do not generate improvements in a diet with healthier items.

Keywords: bolsa familia program; food consumption; Berry index; healthy diversification.

JEL: D04; D12; H31.

1. Os autores agradecem aos comentários e às sugestões dos pareceristas anônimos. Qualquer erro remanescente é de exclusiva responsabilidade dos autores.

2. Professor do Programa de Pós-Graduação em Economia da Universidade Federal da Paraíba (PPGE/UFPB). E-mail: <alessio@ccsa.ufpb.br>.

3. Professora do Programa de Pós-Graduação em Economia da Universidade Federal da Paraíba (PPGE/UFPB). E-mail: <shirley_mesquita@yahoo.com.br>.

4. Professor do Programa de Pós-Graduação em Economia da Universidade Federal da Paraíba (PPGE/UFPB). E-mail: <magnobs@yahoo.com>.

1 INTRODUÇÃO

A redução da pobreza e da desigualdade de renda é o principal objetivo de muitos programas governamentais em países da América Latina. No Brasil, a iniciativa que melhor retrata esse perfil de política pública é o Programa Bolsa Família (PBF). O projeto foi criado em 2003 a partir da junção de diversos outros programas, como o Bolsa Escola, Bolsa Alimentação e Auxílio-Gás, com o intuito de combater a pobreza e a extrema pobreza, bem como promover segurança alimentar e nutricional⁵ em todo o país.

Na literatura sobre o PBF, estuda-se com frequência o impacto sobre a pobreza e a distribuição de renda (Ferraz, 2008; Sátyro e Soares, 2009; Barros *et al.*, 2007), frequência escolar (Kassouf e Glewwe, 2008; Hermeto e Romero, 2009; Ribeiro e Cacciomali, 2012), trabalho infantil (Araújo, Ribeiro e Neder, 2010; Melo e Duarte, 2010), saúde (Camelo, Tavares e Saiani, 2009), segurança alimentar (Camelo, Tavares e Saiani, 2009; Traldi, 2011), consumo (Rodrigues, Gomes e Dias, 2008; Resende e Oliveira, 2008; Baptistella, 2012), entre outros. Dentro dos estudos que avaliam os efeitos sobre o consumo das famílias, o maior destaque é para os gastos com alimentos, sendo que a maior parte dos recursos advindos do programa é direcionada para esse fim (Rodrigues, Gomes e Dias, 2008).

Para o caso brasileiro, é possível encontrar pesquisas, como Menezes, Silveira e Azzoni (2008), Hoffmann (2010), Coelho, Aguiar e Eales (2010) e Pereda e Alves (2012), que avaliam a demanda por alimentos considerando, em especial, a importância dos fatores socioeconômicos das famílias para um consumo mais qualificado em termos de atributos nutricionais e de tipos de bens alimentares, com destaque para as diferentes elasticidades renda e preço entre famílias pobres e ricas. Contudo, a literatura nacional tem pouco explorado a relação específica entre o impacto de programas de transferência de renda (como o PBF) e possíveis alterações nos indicadores de diversificação saudável do consumo de alimentos provocados por esse programa, pois, como salientado por Pereda e Alves (2012), existe a possibilidade de incrementos na renda de famílias pobres resultar em um aumento mais do que proporcional na procura por bens alimentares com características nutricionais indesejáveis.

Segundo Royo-Bardonaba (2007), o organismo humano precisa de uma dieta com aproximadamente quarenta nutrientes para manter o bom funcionamento, sendo necessário, portanto, o consumo de diferentes tipos de alimentos diariamente. De acordo com a pirâmide alimentar do Brasil, exposta em Philippi *et al.* (1999) e

5. Segundo a Lei Orgânica de Segurança Alimentar e Nutricional (LOSAN), Lei nº 11.346, de 15 de setembro de 2006, entende-se por Segurança Alimentar e Nutricional (SAN) a realização do direito de todos ao acesso regular e permanente a alimentos de qualidade, em quantidade suficiente, sem comprometer o acesso a outras necessidades essenciais, tendo como base práticas alimentares promotoras de saúde que respeitem a diversidade cultural e que sejam ambiental, cultural, econômica e socialmente sustentáveis.

Brasil (2013), é recomendado que os brasileiros alimentem-se com diferentes produtos, mas que obedeçam determinadas porções diárias para obter equilíbrio nutricional.

Além dos aspectos nutricionais, a diversificação do consumo alimentar é importante também na dimensão econômica. Kohlmeier *et al.* (1993) destacam que uma alimentação diversificada, por exemplo, melhora o nível de concentração dos indivíduos e, assim, a *performance* escolar, podendo influenciar de forma relevante a formação do capital humano, que é um dos principais determinantes da condição de ocupação no mercado de trabalho e do nível de renda dos indivíduos. Ademais, um fato que merece destaque na literatura é que a demanda por uma cesta de alimentos diversificada está diretamente relacionada com fatores socioeconômicos (Lee, 1987; Lee e Brown, 1989; Thieles e Weiss, 2003; Drescher *et al.*, 2009).

No tocante aos objetivos do PBF, destaca-se a promoção da segurança alimentar e nutricional das famílias beneficiadas, que correspondiam a mais de 14 milhões de famílias no ano de 2013 e que receberam aproximadamente R\$ 25 bilhões do programa, conforme informações do Ministério do Desenvolvimento Social e Combate à Fome (MDS). Dado o universo de famílias beneficiadas e o montante de recursos alocados com esse programa anualmente, o objetivo do presente artigo é analisar o efeito dessa intervenção sobre o nível de diversificação saudável do consumo alimentar das famílias beneficiadas residentes em áreas urbanas do Brasil. Para tanto, foi utilizado o método de *propensity score matching* (PSM) e os microdados da Pesquisa de Orçamentos Familiares (POF) de 2008-2009 do Instituto Brasileiro de Geografia e Estatística (IBGE).

O trabalho encontra-se estruturado em cinco seções, incluindo esta introdução. A segunda seção apresenta uma revisão da literatura sobre o tema. A terceira aborda a estratégia empírica e a base de dados. Na quarta, é feita a análise do impacto do PBF sobre o nível de diversificação dos gastos com alimentos das famílias selecionadas. E, por fim, a quinta seção é reservada às principais conclusões.

2 BOLSA FAMÍLIA E CONSUMO ALIMENTAR

O Ministério do Desenvolvimento Social e Combate à Fome (MDS) define o PBF como um programa de transferência direta de renda que beneficia as famílias brasileiras em situação de pobreza e extrema pobreza, isto é, com renda familiar *per capita* entre R\$ 70,01 e R\$ 140,00 e inferior a R\$ 70,00 mensais, respectivamente. O Artigo 4º do Decreto nº 5.209/2004, que regulamenta a Lei nº 10.836/2004 sobre o programa, aponta seus principais objetivos, a saber: promoção do acesso à rede de serviços públicos, em especial, de saúde, de educação e de assistência social; combate à fome; promoção da segurança alimentar e nutricional; estímulo à emancipação sustentada das famílias que vivem em situação de pobreza e extrema pobreza; combate à pobreza; e promoção da intersetorialidade, da complementaridade e da sinergia das ações sociais do Poder Público.

Em relação aos critérios de elegibilidade, o Decreto nº 6.917/2009 definiu as seguintes condições: renda mensal *per capita* de até R\$ 140 no domicílio; presença de crianças ou adolescentes de 0 a 17 anos, apenas para as famílias em situação de pobreza⁶, e participação da família no Cadastro Único para Programas Sociais (CadÚnico). No entanto, nem todas as famílias que atendem aos critérios de seleção recebem os rendimentos do programa, devido à limitação de recursos governamentais. Após a seleção, a permanência no programa está condicionada à manutenção das crianças e adolescentes em idade escolar, frequentando a escola, e ao cumprimento dos cuidados básicos de saúde, quais sejam, seguir o calendário de vacinação para as crianças entre 0 e 6 anos e a agenda pré e pós-natal para as gestantes e mães em amamentação. Quanto ao benefício, seu valor varia entre R\$ 22,00 e R\$ 200,00, por família, dependendo da renda mensal *per capita* e do número de crianças e adolescentes de até 15 anos e de jovens de 16 e 17 anos no domicílio (MDS, 2013).

Segundo o MDS (2013), os principais resultados do programa estão relacionados à redução da extrema pobreza e da desigualdade, bem como às melhorias na situação alimentar e nutricional das famílias beneficiadas. O PBF, conforme o Instituto Brasileiro de Análises Sociais e Econômicas (Ibase), 2008, atua como uma fonte de renda direcionada principalmente para o consumo de alimentos, dessa forma reduzindo o problema da insegurança alimentar entre as famílias beneficiadas⁷.

Aspectos destacados na literatura sobre a demanda por bens alimentares no Brasil revelam que incrementos de renda para famílias pobres, tal como preconizado pelo PBF, afetam positivamente a estrutura de consumo desses bens, mas com diferentes implicações no teor nutricional da cesta consumida. Nessa direção, Menezes, Silveira e Azzoni (2008) estimam as elasticidades preço, preço-cruzada e renda da demanda para uma cesta de bens alimentares básicos, que são importantes para melhorar as necessidades calóricas das famílias. Na estratégia empírica, é usado o modelo *almost ideal demand system* (AIDS), com base nos dados da POF de 1987-1988 e de 1995-1996. Especificamente à elasticidade renda, os autores encontraram resultados positivos e estatisticamente significantes em todos os grupos de bens alimentares. Registraram também que esta elasticidade é maior para as famílias mais pobres em comparação com as mais ricas para todas as cestas de bens alimentares, sendo consistente com os resultados esperados, considerando-se a grande desigualdade de renda no Brasil.

Também nessa perspectiva, Coelho, Aguiar e Eales (2010) estudam a demanda por bens alimentares no Brasil, estimando a demanda para dezoito

6. Para famílias em situação de extrema pobreza, a participação no PBF não está condicionada à idade dos membros do domicílio.

7. Estudos realizados pelo Ministério da Saúde (MS) sobre o Bolsa Alimentação, que a partir de 2003 passou a integrar o PBF em conjunto com outros programas federais de transferência de renda, demonstram a existência de impactos positivos deste programa tanto em termos qualitativos quanto quantitativos sobre a composição da cesta de bens alimentares, isto é, as famílias beneficiadas tendem a gastar mais em alimentos e possuem uma dieta mais diversificada do que as famílias não beneficiadas (Brasil, 2004, 2005).

produtos com dados da POF de 2002-2003, por meio do modelo *quadratic almost ideal demand system* (QUAIDS). Em geral, os resultados indicaram que quanto maior a renda familiar, menor é a probabilidade de gastos em consumo básicos, enquanto para carnes, leite e outros produtos a correlação é positiva. Os cálculos das elasticidades renda mostraram que dos dezoito bens, um terço deles são ditos superiores e nenhum é inferior.

Por seu turno, Hoffmann (2010) mensura as elasticidades renda no Brasil para várias categorias de despesas, incluindo produtos alimentares, por meio da POF 2008-2009. Para tanto, o autor estima a despesa média para dez classes de renda familiar *per capita* como função da despesa média *per capita*, ajustada a uma função poligonal com três segmentos. Além disso, são estimadas as elasticidades renda do consumo de bens físicos. Em geral, os resultados apontam que as elasticidades renda média das despesas com alimentos das famílias são quase sempre maiores que 1, bem como evidenciam que estas elasticidades são maiores que as obtidas relativamente ao consumo físico de alimentos, uma vez que o produto adquirido pelas famílias mais ricas é de melhor qualidade e mais caro, quando comparado ao das famílias mais pobres.

Em um estudo mais recente, Pereda e Alves (2012) analisam a demanda por nutrientes a fim de investigar a mudança de composição na qualidade da dieta alimentar da população domiciliar brasileira. O trabalho faz uma conexão entre um sistema de demanda por alimentos e por nutrientes, o que se configura como uma inovação em relação aos trabalhos anteriores. É estimada, então, a demanda por nutrientes com base no modelo *quadratic almost ideal demand system* (QUAIDS), usando a POF 2002-2003 para os domicílios urbanos brasileiros. Os cálculos das elasticidades renda evidenciaram que proteínas, lipídios, colesterol e fibras são considerados itens de luxo para os domicílios mais pobres e são normais para domicílios de renda maior. Em relação aos incrementos de renda para as famílias mais pobres, os achados da pesquisa revelam dois pontos conflitantes: por um lado, mais renda aumentaria mais que proporcionalmente o consumo de proteínas e fibras; por outro, mais renda conduziria a elevação mais do que proporcional no consumo de lipídios e de colesterol, considerados nocivos à saúde das pessoas.

Dentro desse contexto, é possível encontrar na literatura nacional estudos que destaquem o papel do incremento de renda proveniente do PBF sobre o combate à insegurança alimentar e ao seu efeito sobre o consumo de alimentos⁸. O estudo de Rodrigues, Gomes e Dias (2008) sobre os impactos do PBF no consumo e nos preços dos alimentos revela uma relação positiva entre as transferências governamentais e o consumo de alimentos entre as famílias beneficiadas pelo programa, bem como

8. Resende e Oliveira (2008) fazem um estudo sobre os efeitos do Bolsa Escola, um dos componentes do atual PBF, utilizando dados da POF 2002-2003 e PSM. Os principais resultados mostram um efeito positivo desse programa de transferência de renda sobre o consumo das famílias pobres, com destaque para o aumento dos gastos em alimentação.

mostra que esse efeito positivo é potencializado pela alta elasticidade renda da demanda por alimentos nos domicílios beneficiados. Dessa forma, confirma-se a hipótese de que uma expansão na renda das famílias em situação de pobreza ou extrema pobreza afeta significativamente o comportamento do consumo de alimentos.

O trabalho de Duarte, Sampaio e Sampaio (2009) avalia o impacto das transferências de renda do PBF sobre os gastos com alimentos de famílias rurais, analisando os dados fundamentados no uso do modelo PSM. Os resultados apontam que a maior parcela da renda das famílias é gasta com alimentação e que, portanto, o programa exerce um impacto positivo sobre a estrutura de consumo alimentar. Os achados empíricos de Camelo, Tavares e Saiani (2009) mostram que o PBF contribui para a redução da condição de insegurança alimentar leve dos beneficiados⁹ e para o aumento da probabilidade de a criança obter melhores indicadores antropométricos.

Em contraste, Pereira *et al.* (2011), em estudo para os municípios brasileiros com baixo índice de IDH, encontram evidências de que a maior parte das famílias inseridas no PBF não conseguiu sair da situação de insegurança alimentar e que a maior parcela dos seus gastos advindos das transferências é destinada à compra de alimentos.

Em estudo para o estado do Ceará, o PBF assina um papel importante no bem-estar das famílias como política de curto prazo, tendo impacto positivo sobre indicadores de saúde, educação e estado nutricional das famílias beneficiadas (Rocha, Khan e Lima, 2009). Os autores destacam que o consumo médio de bens da cesta alimentar básica é superior nas famílias beneficiadas em relação às não beneficiadas, e que para cada R\$ 1,00 transferido para as famílias R\$ 0,72 são gastos com alimentação. Outro estudo realizado por Aires, Gomes e Esmeraldo (2011) para o Ceará indica que as famílias beneficiadas pelo PBF apresentam aumentos na quantidade e no acesso aos alimentos, contudo sem garantias de melhoria significativa nos índices de segurança alimentar.

Os resultados de Traldi (2011) indicam que, apesar do maior gasto com alimentos, 95% das famílias do município de Araraquara, em São Paulo, que recebem os rendimentos do PBF, permanecem na situação de insegurança alimentar. Por sua vez, Assunção, Leitão e Inácio (2012) indicam que o recebimento do benefício direcionado às famílias de pescadores artesanais do estado de Pernambuco aumenta o consumo de alimentos e a introdução de novos itens na dieta das famílias. Os achados de Paula *et al.* (2012) também mostraram uma alteração na estrutura de consumo, principalmente evidenciado no aumento do consumo de leite das famílias beneficiadas pelo PBF em Belo Horizonte, Minas Gerais.

9. A situação de insegurança alimentar é, comumente, classificada em três categorias: leve, quando há receio de passar fome em um futuro próximo; moderada, quando há restrição na quantidade de comida para a família; e alta, nos casos de falta de alimento na mesa (Fiocruz, 2011).

De forma geral, a literatura específica que avalia o papel do PBF no consumo de alimentos mostra-se inconclusiva quanto aos resultados na segurança alimentar das famílias beneficiadas. A grande parte dessas pesquisas preocupa-se em verificar a mudança provocada nos gastos totais com alimentos ou em despesas por grupos alimentares, não considerando, portanto, os efeitos do programa em um indicador sintético que contemple a diversificação e a questão nutricional da cesta de alimentos consumida. Assim, este trabalho pretende preencher essa lacuna de acordo com a avaliação do impacto do PBF sobre medidas que captam a diversificação alimentar em termos quantitativos e qualitativos.

3 ESTRATÉGIA EMPÍRICA

Para a execução desta pesquisa, foi necessário construir a variável de interesse, qual seja o nível de diversificação saudável de consumo de alimentos, e classificar as famílias beneficiadas e não beneficiadas pelo PBF. Isso foi possível por meio da utilização dos dados da POF 2008-2009 e da adoção de uma metodologia capaz de medir os efeitos dessa intervenção governamental sobre o nível de diversificação das famílias com bens alimentares, aspecto relacionado a um dos alvos do programa, que é melhorar a segurança alimentar e nutricional dos beneficiados. Sendo assim, fez-se o uso da abordagem PSM, que será descrita ao longo da subseção 3.1, a seguir. Por sua vez, a subseção 3.2 apresenta o índice de diversificação saudável de alimentos.

3.1 *Propensity score matching* (PSM)

A avaliação de impacto tem como escopo examinar se um determinado programa está atingindo os resultados desejados, bem como se permite quantificar a magnitude e o sinal do efeito médio da intervenção sobre um determinado grupo de indivíduos (Cameron e Trivedi, 2005). Para Caliendo e Kopeinig (2008), inferências acerca do impacto do tratamento sobre um indivíduo envolvem “especulação” sobre como teria sido o desempenho do agente caso ele não tivesse recebido o tratamento. Para a construção de modelos que objetivam mensurar o impacto de um programa, é necessário dispor de informações sobre os indivíduos, o tratamento e os resultados potenciais. Contudo, a grande dificuldade de se estimar o efeito do tratamento (τ) para o indivíduo i , como mostrado na Equação 1, é que na prática somente se observa um dos resultados potenciais. Isto é, dado que o indivíduo i participa de um programa, só é possível identificar o resultado potencial fruto da participação do programa, $Y_i(1)$, ao passo que o resultado contrafactual para esse mesmo agente não seja observável, $Y_i(0)$.

$$\tau_i = Y_i(1) - Y_i(0) \quad (1)$$

O problema técnico na literatura envolvendo avaliação de impacto é que a variável referente à participação ou não no programa é potencialmente correlacionada ao termo de erro estocástico, o que traz para o parâmetro que mede o sinal e magnitude do efeito médio do tratamento um viés atrelado ao problema de autosseleção¹⁰. Um dos métodos que se propõe a minimizar o viés da estimação desse efeito com o conjunto de dados observáveis, que é o utilizado no presente artigo, foi proposto por Rosenbaum e Rubin (1983), conhecido como *propensity score matching* (PSM). Conforme Khandker, Koolwal e Samad (2010), o PSM propõe resumir as características de cada unidade em um único *propensity score*, que viabiliza o pareamento, levando em consideração a criação de um grupo de controle próximo ao grupo de tratamento em termos de características observáveis (X). De acordo com isso, os resultados dos tratados e não tratados com *propensity scores* semelhantes podem ser comparados para, assim, obter-se o efeito do programa.

Becker e Ichino (2002) apontam que o PSM é uma abordagem que possibilita a correção da estimação do efeito de tratamento por meio do controle desses problemas de aleatoriedade fundamentado na ideia de que o viés é reduzido quando a comparação dos resultados ocorre a partir de grupos tratados e de controles mais semelhantes possíveis. A maior parte da literatura sobre avaliação de impacto admite a hipótese que a variável atrelada ao tratamento satisfaz condições de exogeneidade, implicando que diferenças de resultados entre indivíduos tratados e não-tratados são atribuídas ao tratamento (Caliendo e Kopeinig, 2008). Assim, o *propensity score*, $p(X)$ é definido como a probabilidade condicional de receber um tratamento dado por características de pré-tratamento observáveis:

$$p(X) \equiv Pr(P|X) = \Phi\{h(X_i)\} \quad (2)$$

Em que: T é o indicador binário de exposição ao tratamento, isto é, $T = (0,1)$; X é o vetor de características observáveis; $h(X_i)$ é a forma funcional que inclui as covariadas; Φ representa a função de distribuição acumulada. Supondo que Φ segue uma distribuição de probabilidade logística, pode-se estimar $p(X)$ por meio de um modelo de resposta qualitativa logit.

Conforme Rosenbaum e Rubin (1983), se o *propensity score* $p(X_i)$ é conhecido, o efeito médio do tratamento sobre o grupo tratado (τ_{ATT}) pode ser estimado por intermédio da Equação 3 abaixo.

$$\tau_{ATT} = E\{E[Y_{1i}|T_i = 1, p(X_i)] - E[Y_{0i}|T_i = 0, p(X_i)|T_i = 1]\} \quad (3)$$

10. Ver, por exemplo, Heckman (1979) para mais detalhes sobre o problema de viés de seleção e suas consequências para a tendenciosidade dos estimadores.

Em que: Y_{1i} e Y_{0i} são os resultados potenciais nas duas situações contrafactuais (tratamento e não tratamento).

Para o cálculo de τ_{ATT} , é necessário considerar algumas suposições, entre elas destaca-se que as unidades com igual *propensity score* devem ter a mesma distribuição de características observáveis e não observáveis independentemente da situação do tratamento. É como se, segundo Menezes-Filho (2012), na comparação entre dois indivíduos, o único fator que os diferencia é a participação no programa. Portanto, a validade do PSM depende de duas condições: (1) independência condicional¹¹, ou seja, os fatores não observados não afetem a participação no programa; (2) suporte comum ou sobreposição considerável nos *propensity scores* entre os participantes e não participantes do programa (Khandker, Koolwal e Samad, 2010, p. 53).

É válido enfatizar que a estimação do *propensity score* não é suficiente para o cálculo da Equação (3), sendo que a probabilidade de observar duas unidades com igual valor do *propensity score* é muito baixa, dado que $p(X)$ é uma variável contínua. Dessa forma, existem alguns algoritmos de pareamento para superar este problema, entre os quais foi usada neste artigo a técnica de pareamento por vizinho mais próximo (com reposição), considerando uma distância máxima tolerável (*calliper*) de 0,1% e um corte (*trimming*) nas observações do grupo de tratamento e controle abaixo dos valores do percentil 5. Ademais, as estimativas do ATT foram realizadas por meio do método *bootstrap* com mil replicações, com a finalidade de reduzir o viés no cálculo do erro-padrão do parâmetro de interesse. Ademais, foi incluído no Apêndice C deste artigo uma tabela com as estimativas do ATT para outras formas de pareamento: dez vizinhos mais próximos, vizinho mais próximo sem reposição; raio; Kernel.

3.2 Índice de diversificação saudável

Considerando os objetivos do PBF, esta pesquisa pretende avaliar os impactos deste programa sobre o nível de diversificação de alimentos consumidos pelas famílias beneficiadas. Para atender tal propósito, foi necessário inicialmente computar as informações do consumo de diferentes itens alimentares que pudessem assim subsidiar a construção da variável de interesse.

O indicador de diversidade saudável de itens alimentares consumidos por uma família, com base em Drescher, Thiele e Mensink (2007), foi mensurado considerando três dimensões: *quantidade*, captado pelo número de itens consumidos; *distribuição*, obtido por meio da participação dos itens alimentares na cesta total de bens alimentares; *peso nutricional*, capturado pelas ponderações dos diferentes grupos (alimentos vegetais, produtos de origem animal e gorduras e óleos) e

11. Sobre a hipótese de independência condicional, Khandker, Koolwal e Samad (2010) enfatizam que ela não pode ser testada diretamente, visto que tal critério vai depender das características específicas e do desenho de cada programa.

subgrupos alimentares, que foram hierarquizados dos itens mais saudáveis (na base da pirâmide nutricional) aos menos saudáveis (no topo da pirâmide).

Tradicionalmente, os índices de diversificação eram calculados considerando como mais diversificada a cesta com maior participação de itens alimentares diferentes consumidos pelas famílias, independentemente de sua classificação nutricional. Trabalhos como Lee (1987), Lee e Brown (1989), Moon *et al.* (2002), Thiele e Weiss (2003) e Stewart e Harris (2005) adotaram esse tipo de medida.

O índice de Berry (Berry, 1971), também conhecido como índice Simpson ou inverso do índice de Herfindahl-Hirschman, tem sido comumente aplicado em estudos econômicos como medida de diversificação alimentar (Gollop e Monahan, 1991; Thiele e Weiss, 2003; Stewart e Harris, 2005). O índice representa o somatório das parcelas de cada item individual, sendo definido por: $IB_i = 1 - H_i = 1 - \sum_{j=1}^n s_{ij}^2$, em que H_i é o índice de Herfindahl-Hirschman por família i ; s_{ij} é a fração do produto j em relação ao total de itens alimentares consumidos pela família i e n representa o número total de itens alimentares.

O índice de Berry¹² (IB) trata-se de uma medida de diversidade proporcional, isto é, quantifica a parcela do alimento i na quantidade total. O IB varia entre 0 e $1 - 1/n$, assumindo o valor zero quando não existe diversidade, isto é, quando o indivíduo consome somente um item alimentar, e o valor máximo $1 - 1/n$, se os itens forem igualmente distribuídos na cesta de consumo, isto é, valores próximos de um indicam um número de alimentos consumidos tendendo ao infinito ($n \rightarrow \infty$).

Não obstante, o consumo de mais itens alimentares e, assim, um maior índice de Berry não implica, por si só, a família possuir uma dieta mais saudável. Nessa direção, o índice de diversidade de alimentação saudável, doravante denotado de índice *healthy food diversity* (HFD), proposto por Drescher, Thiele e Mensink (2007), tem como escopo dispor de um indicador de diversificação que reflita de forma mais apropriada as questões nutricionais. O índice HFD pondera o índice de Berry por um fator que leva em conta o “valor de saúde” da cesta de consumo, de modo que os seus resultados consigam melhor diferenciar a diversidade de itens alimentares saudáveis de não saudáveis.

Conforme Drescher, Thiele e Mensink (2007), o cálculo do valor de saúde (VS) da cesta de consumo da família requer os chamados fatores de saúde (FS) para diferentes tipos de alimentos com base na pirâmide alimentar, tendo em vista que uma alimentação com boas propriedades nutricionais deve priorizar, em especial, a variedade e o número de porções recomendadas. Assim, o VS é definido por: $VS = \sum_{i=1}^n (FS_i) \cdot s_i$, em que $FS = G_w \cdot G_b$ e s_i é a fração consumida do item i na

12. O valor do índice de Berry dificilmente é interpretado em termos absolutos. Define-se que valores altos indiquem alto grau de diversificação, e que ele só pode ser comparado, se considerado à mesma cesta de bens para todas as famílias da amostra.

cesta de consumo¹³. Dessa forma, para o cálculo do FS, é preciso dispor do peso nutricional do grupo (G_w) e do peso nutricional do subgrupo (G_b) do produto. Seguindo a lógica dos citados autores, mas atribuindo os pesos do grupo do produto conforme as peculiaridades da pirâmide alimentar brasileira¹⁴, G_w , portanto, é formado por três grupos de alimentos com os pesos fundamentado no número de porções recomendadas, conforme os pesos a seguir: alimentos vegetais (74%), produtos de origem animal (21%) e gorduras e óleos (5%).

É válido destacar que, na metodologia do índice HFD, cada grupo de produtos é composto por cinco subgrupos (G_b), hierarquizados pelo nível de qualidade nutricional em relação à posição na pirâmide alimentar, recebendo os seguintes valores percentuais: 36%, 28%, 20%, 12% e 4%. Por exemplo, o fator de saúde correspondente ao item maçã possui um FS de 0,2664 (= 0,74 x 0,36), enquanto o FS do item manteiga é de 0,006 (= 0,05 x 0,12). O detalhamento do FS para cada um dos itens considerados nesta pesquisa está na tabela A1 no apêndice A.

Sendo assim, os resultados do PBF serão avaliados em termos do indicador de diversificação de alimentos tradicional, o índice de Berry, e, de forma especial, no indicador que reflete a direção da qualidade nutricional da cesta de alimentos consumida pela família, o índice HFD.

4 BASE DE DADOS E VARIÁVEIS

A fonte de dados usada no presente trabalho é proveniente da Pesquisa de Orçamentos Familiares (POF) 2008-2009, do Instituto Brasileiro de Geografia e Estatística (IBGE), que possibilita identificar os domicílios que tiveram rendimentos advindos de programas sociais – como o Bolsa Família –, bem como permite a análise dos gastos em diversos itens de consumo. A POF 2008-2009 é formada por sete questionários, que trazem informações detalhadas sobre o consumo das famílias brasileiras presentes na amostra, apresentando também dados acerca das características gerais do domicílio (rendimento e condições estruturais) e das pessoas que o compõem.

Vale destacar que todas as informações trabalhadas neste estudo levam em conta domicílios que apresentam apenas uma unidade de consumo e uma família, que representam 94,6% da amostra total da POF e que estão situadas da zona urbana. Esses primeiros filtros permitem: uniformizar a nomenclatura de unidade de consumo, família e domicílio, pois todo domicílio terá uma única unidade de consumo e uma única família; reduzir as chances de sub ou sobrestimar os impactos

13. Os pesos atribuídos em Drescher, Thiele e Mensink (2007) e Drescher *et al.* (2009) seguem a pirâmide alimentar da Sociedade de Nutrição da Alemanha, em que os pesos dos grupos foram 73%, 25% e 2%, respectivamente, para alimentos vegetais, produtos de origem animal e gorduras e óleos.

14. As informações sobre a estrutura da pirâmide alimentar do Brasil estão detalhadas em Philippi *et al.* (1999) e Brasil (2013).

do PBF em domicílios que, por exemplo, possuem mais de uma família beneficiada; evitar problemas na mensuração da variável dependente para moradores de áreas rurais, no tocante a questões relacionadas ao autoconsumo de bens que poderiam provocar viés nos resultados do efeito médio de tratamento. Para atender ao objetivo deste trabalho, foram computadas informações de consumo de cada família em relação a 123 itens de bens alimentares¹⁵. Destaca-se que as bebidas alcoólicas foram desconsideradas para o efeito do cálculo do índice de diversificação.

Neste estudo, o grupo de tratamento foi obtido por meio da informação do recebimento do valor do PBF por algum membro da família. No entanto, aquelas que recebiam também outros benefícios de programas sociais federais, como Programa de Erradicação do Trabalho Infantil (PETI), Benefício de Prestação Continuada da Assistência Social (BPC) e auxílio a portadores de deficiência física, foram desconsideradas para não ocorrer sobreposição de programas de transferência de renda. Do total de famílias residentes em áreas urbanas que receberam o benefício do PBF, 142 delas (2,8%) participavam também de outros programas de transferência de renda

Além da medida que representa o nível de diversificação, considerada como variável de resultado, as covariadas utilizadas para a determinação da participação no programa Bolsa Família foram divididas em quatro blocos:

- *atributos da família*: renda familiar *per capita* líquida dos programas federais de transferência de renda, número de membros por faixa de idade (de 0 a 11 anos, de 12 a 17 anos, de 18 a 29 anos e com 30 ou mais anos), proporção de analfabetos na família com 21 anos ou mais de idade, estrutura familiar (biparental, família com chefe e cônjuge, e monoparental, família apenas com chefe e sem cônjuge), *status* de ocupação remunerada da mãe e do pai e a interação entre mãe com ocupação remunerada e casada. Para a quantificação do número de membros na família não foi considerada a quantidade de empregada doméstica e seus agregados;
- *atributos do chefe da família*: faixa de escolaridade (analfabeto, ensino fundamental incompleto, ensino fundamental completo e ensino médio completo ou mais anos de estudo), idade¹⁶, gênero (um, se mulher), estado e raça (um, se o chefe declara-se branco);
- *acesso à infraestrutura básica*: uma variável *dummy* que capta se na residência da família o escoadouro sanitário do banheiro ou sanitário

15. O número total de bens alimentares classificados na POF 2008-2009 é de 1.583, agrupados em 53 itens dentro de 16 categorias. Para esse estudo, a classificação foi expandida de 53 para 123 itens dentro das mesmas categorias iniciais. A descrição desses itens encontra-se na tabela A1 no apêndice A deste artigo.

16. Neste artigo, as famílias chefiadas por indivíduos com menos de 18 anos de idade foram desconsideradas.

do domicílio é feito utilizando-se a fossa rudimentar, direto para o rio, lago ou mar, vala ou não existe;

- *localização geográfica do domicílio*: região de residência (Norte, Nordeste, Sudeste, Sul e Centro-Oeste); variável *dummy* para família residente em área metropolitana. É válido destacar que a análise é restrita para famílias de áreas urbanas.

Esses quatro blocos de variáveis captam de forma geral os requisitos à candidatura de uma família ao programa e aos escopos definidos pelo PBF, em que as características familiares representam os principais determinantes para acesso ao programa, com destaque para o nível de renda *per capita* e para o número de crianças e adolescentes entre 0 e 17 anos de idade. A escolha dessas características observáveis está condizente com as covariadas usadas nos trabalhos que estimaram as probabilidades de participação no PBF (usando técnicas de pareamento), como Tavares *et al.* (2009); Araújo, Ribeiro e Neder (2010); Mattos, Maia e Marques (2010); Melo e Duarte (2010); e Ribeiro e Cacciamali (2012). A tabela 1, a seguir, apresenta a descrição e a estatística descritiva das variáveis.

TABELA 1
Descrição das variáveis e estatística descritiva da base de dados

Variáveis	N	Média	Desvio-padrão	Mínimo	Máximo
Variável de resultado					
Índice de diversificação (Berry) – em %	33.255	80,797	14,509	12,444	97,442
Índice de diversificação saudável (HFD) – em %	33.255	7,722	3,610	0,360	22,909
Indicação de participação no programa					
Grupo de tratamento	33.255	0,102	0,303	0,000	1,000
Bloco de covariadas para o propensity score					
I – Atributos da família					
Renda familiar <i>per capita</i> líquida dos programas de transferência	33.255	1176,940	2058,298	-946,870	87430,750
Número de pessoas entre 0 e 11 anos de idade	33.255	0,580	0,883	0,000	10,000
Número de pessoas entre 12 e 17 anos de idade	33.255	0,342	0,639	0,000	5,000
Número de pessoas entre 18 e 29 anos de idade	33.255	0,642	0,833	0,000	10,000
Número de pessoas com 30 ou mais anos de idade*	33.255	1,590	0,820	0,000	9,000
Proporção de analfabetos na família	33.255	0,073	0,217	0,000	1,000
Status de ocupação do pai	33.255	0,602	0,490	0,000	1,000
Status de ocupação da mãe	33.255	0,471	0,499	0,000	1,000
Núcleo familiar					
Família tipo 1 (biparental)	33.255	0,671	0,470	0,000	1,000
Família tipo 2 (monoparental)*	33.255	0,329	0,470	1,000	1,000
Mãe ocupada e casada	33.255	0,337	0,473	0,000	1,000

(Continua)

(Continuação)

Variáveis	N	Média	Desvio-padrão	Mínimo	Máximo
II – Atributos do chefe da família					
Idade	33.255	47,543	15,381	18,000	103,000
Faixa de escolaridade do chefe					
Analfabeto	33.255	0,083	0,276	0,000	1,000
Ensino fundamental incompleto	33.255	0,398	0,489	0,000	1,000
Ensino fundamental completo	33.255	0,144	0,351	0,000	1,000
Ensino médio completo ou mais anos de estudo*	33.255	0,375	0,484	0,000	1,000
Raça (cor branca)	33.255	0,522	0,500	0,000	1,000
Gênero (mulher)	33.255	0,323	0,468	0,000	1,000
III – Acesso à infraestrutura básica					
Esgotamento sanitário precário ou inexistente	33.255	0,221	0,415	0,000	1,000
IV – Localização geográfica					
Família reside em área metropolitana	33.255	0,434	0,496	0,000	1,000
<i>Região de residência da família</i>					
Norte	33.255	0,060	0,238	0,000	1,000
Nordeste	33.255	0,234	0,423	0,000	1,000
Sudeste*	33.255	0,473	0,499	0,000	1,000
Sul	33.255	0,159	0,366	0,000	1,000
Centro-Oeste	33.255	0,074	0,262	0,000	1,000

Fonte: Elaboração própria, a partir dos dados da POF 2008-2009.

Nota: O número de observações (ampliada pelo fator de expansão da POF) corresponde a 38.559.912.

*Categoria de referência na regressão

O número de observações na POF 2008-2009, após a exclusão daquelas unidades que dispunham de alguma informação perdida para ao menos uma das variáveis apresentadas na tabela 1, é de 33.255 famílias – que representam mais 38 milhões de famílias na amostra ampliada pelo fator de expansão ajustado para estimativas da POF. Sendo que destas, apenas 10% recebem benefícios do PBF¹⁷. Assim, a estimação do efeito do PBF sobre o índice de diversificação (IB e, de forma particular, HFD) exposta na próxima seção requer *a priori* a identificação do grupo de controle que seja o mais próximo possível do grupo de tratamento em relação às características observáveis dos quatro blocos de variáveis expostas na tabela 1.

5 RESULTADOS

Esta seção de resultados apresenta os elementos que subsidiam as discussões dos impactos do PBF sobre a diversificação do consumo de alimentos. Dessa forma, as duas próximas subseções exibem informações sobre o grau de ajustamento do

17. Essa proporção está ajustada pelo fator de expansão, de modo que ele indica que, do total de observações (pela amostra ampliada), mais de 6 milhões de famílias recebem o benefício do PBF.

modelo *logit* usado na estimação do *propensity score* e o efeito médio de tratamento do programa sobre as medidas de diversificação (IB e HFD).

5.1 Distribuição do índice de diversificação e análise do grau de ajuste do *propensity score*

A diversificação dos itens alimentares é importante para a segurança alimentar e nutricional das famílias. Contudo, a diversificação medida por indicadores tradicionais na literatura, como IB, por si só não garante que a família tenha uma cesta de consumo mais saudável, por isso que os resultados desta pesquisa são realizados de forma particular em função do índice HFD que pondera cada item da cesta de consumo das famílias conforme predicados nutricionais. Desse modo, a tabela 2 apresenta os valores médios dos índices de Berry e HFD por decil de renda para as famílias brasileiras.

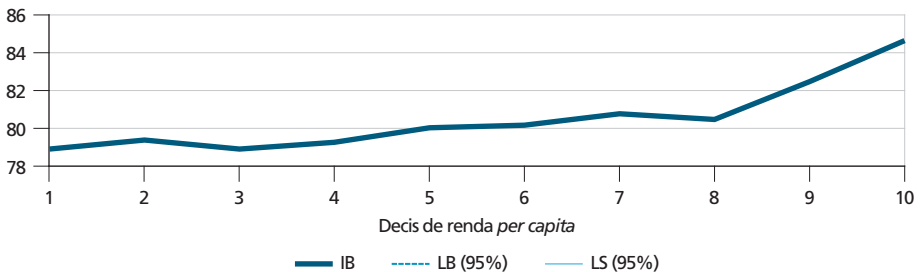
TABELA 2
Distribuição dos índices de diversificação por classes de renda familiar *per capita*

Classes de renda <i>per capita</i>	Máx.	Média	Média	Erro-padrão	IC (95%) LI	IC (95%) LS
	Renda <i>per capita</i>		Índice de Berry (IB)			
Abaixo do 1º decil	161,52	110,52	78,9014	0,0088	78,8841	78,9187
Entre o 1º e o 2º decil	246,17	204,84	79,3026	0,0083	79,2863	79,3189
Entre o 2º e o 3º decil	329,22	287,48	78,9662	0,0084	78,9498	78,9826
Entre o 3º e o 4º decil	423,11	376,88	79,2653	0,0082	79,2492	79,2814
Entre o 4º e o 5º decil	532,62	476,56	80,0104	0,0076	79,9956	80,0252
Entre o 5º e o 6º decil	667,76	600,01	80,0948	0,0073	80,0805	80,1092
Entre o 6º e o 7º decil	861,61	756,75	80,7489	0,0070	80,7352	80,7626
Entre o 7º e o 8º decil	1.175,96	1.006,76	80,4734	0,0071	80,4594	80,4874
Entre o 8º e o 9º decil	1.892,79	1.472,84	82,2965	0,0067	82,2835	82,3096
Acima do 9º decil	87.430,75	4.350,80	84,5892	0,0055	84,5784	84,6000
Índice de diversificação saudável (HFD)						
Abaixo do 1º decil	161,52	110,52	6,7060	0,0019	6,7022	6,7099
Entre o 1º e o 2º decil	246,17	204,84	7,0231	0,0018	7,0195	7,0266
Entre o 2º e o 3º decil	329,22	287,48	7,1765	0,0018	7,1730	7,1801
Entre o 3º e o 4º decil	423,11	376,88	7,0909	0,0018	7,0874	7,0945
Entre o 4º e o 5º decil	532,62	476,56	7,4167	0,0018	7,4132	7,4202
Entre o 5º e o 6º decil	667,76	600,01	7,5997	0,0018	7,5962	7,6033
Entre o 6º e o 7º decil	861,61	756,75	7,7372	0,0018	7,7336	7,7407
Entre o 7º e o 8º decil	1.175,96	1.006,76	7,8115	0,0017	7,8082	7,8148
Entre o 8º e o 9º decil	1.892,79	1.472,84	8,2961	0,0018	8,2926	8,2995
Acima do 9º decil	87.430,75	4.350,80	9,0198	0,0017	9,0166	9,0231

Fonte: Elaboração própria a partir dos dados da POF 2008-2009.

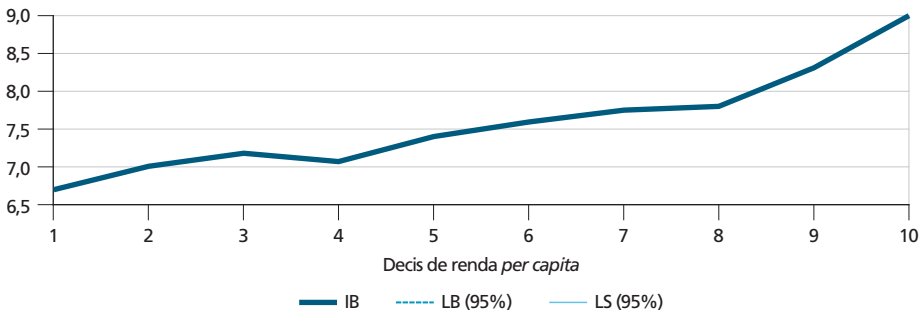
As informações da tabela 2 indicam, de maneira geral, uma tendência de crescimento da diversificação do consumo de alimentos à medida que o nível de renda cresce. Assim, sem controlar para outras características da família (como escolaridade, número de membros, localização do domicílio etc.) que podem afetar a decisão de consumo por itens que elevem o teor nutricional da alimentação, pode-se verificar que as famílias com menor rendimento apresentam uma cesta menos diversificada em termos do consumo de diferentes itens e dos predicados nutricionais recomendados. Nessa direção, as figuras 1 e 2 exibem as médias das medidas de diversificação do consumo alimentar (IB e HFD) por decil de renda familiar *per capita*, com o objetivo de facilitar a visualização da relação direta entre esses indicadores e o nível de renda. Conforme essas figuras, essa tendência crescente é mais consistente para o índice de diversificação saudável do que para o índice de Berry, muito embora em ambos os casos as famílias mais ricas possuem uma cesta de consumo mais diversificada do que famílias com menor nível de renda. Tais fatos já evidenciam a necessidade de políticas nutricionais direcionadas para melhorar o hábito de consumo saudável de alimentos no país e programas específicos para as famílias mais pobres.

FIGURA 1
Média do índice de diversificação de Berry (IB) por decil de renda familiar *per capita*



Fonte: Elaboração própria a partir dos dados da POF 2008-2009.

FIGURA 2
Média do índice de diversificação Saudável (HFD) por decil de renda familiar *per capita*



Fonte: Elaboração própria a partir dos dados da POF 2008-2009.

No entanto, como observado em Pereda e Alves (2012), incrementos na renda de famílias mais pobres implicam dois efeitos conflitantes em relação à demanda por nutrientes, pois o aumento de renda gera uma elevação mais que proporcional em alimentos com nutrientes prejudiciais (lipídios e colesterol) e benéficos (proteínas e fibras) para a saúde dos indivíduos, indicando a possibilidade de acréscimos na renda de famílias pobres não atuar na direção de uma demanda por alimentos mais saudáveis. Nesse sentido, as evidências apontadas pelos autores reforçam os questionamentos sobre a efetividade dos programas de transferência de renda, como o PBF, no que diz respeito aos seus efeitos sobre a segurança alimentar e nutricional dos beneficiados. Por isso, o interesse do presente estudo em identificar os impactos do PBF sobre o nível de diversificação saudável do consumo de alimentos das famílias tratadas.

Tendo em vista as mais de 30 mil famílias presentes na amostra deste artigo, é extremamente relevante para os fins da medição dos impactos do PBF a identificação do grupo de controle. Como salienta Wooldridge (2010), a não participação no programa não garante que todos os indivíduos possam ser enquadrados como grupo de controle, pois poder-se-ia comparar dois grupos em que o efeito médio de tratamento sobre a variável de interesse não seria uma medida relevante para a política pública, dado que o grupo de controle seria formado por unidades que, *ceteris paribus*, nunca seriam elegíveis para tratamento.

Pensando nas características de elegibilidade ao PBF e outras variáveis que podem afetar a participação no programa (como características de localização geográfica da família) foi feito o uso de ferramentas estatísticas específicas para aferir o desempenho da intervenção governamental em estudo, tendo em vista as particularidades da base de dados em seção transversal. Sendo assim, a tabela 3 evidencia os resultados do estágio inicial do método PSM, que consiste na estimação das probabilidades de participação em um programa condicionado a um vetor de características observáveis, para assim viabilizar um melhor pareamento entre os grupos de tratamento e controle.

TABELA 3

Estimação do *propensity score* fundamentado nas características observáveis para famílias residentes em áreas urbanas do Brasil. Variável dependente: *dummy* de participação no programa

Variáveis	Coefficiente	Erro-padrão	P-valor
Atributos da família			
Renda familiar <i>per capita</i> líquida	-0,0044	0,0002	0,0000
Pessoas entre 0 e 11 anos de idade	0,3845	0,0313	0,0000
Pessoas entre 12 e 17 anos de idade	0,5562	0,0393	0,0000
Pessoas entre 18 e 29 anos de idade	-0,0902	0,0343	0,0090

(Continua)

(Continuação)

Variáveis	Coefficiente	Erro-padrão	P-valor
Proporção de analfabetos na família	-0,2763	0,1857	0,1370
Status de ocupação do pai	0,2878	0,0955	0,0030
Status de ocupação da mãe	0,7131	0,1242	0,0000
Núcleo familiar			
Família tipo I (biparental)	0,3183	0,1110	0,0040
Mãe ocupada e casada	-0,2395	0,1385	0,0840
Atributos do chefe da família			
Idade	-0,0120	0,0027	0,0000
Faixa de escolaridade do chefe			
Analfabeto	1,1140	0,1657	0,0000
Ensino fundamental incompleto	0,8303	0,0870	0,0000
Ensino fundamental completo	0,3567	0,1068	0,0010
Raça (cor branca)	-0,1863	0,0683	0,0060
Gênero (mulher)	0,4511	0,0916	0,0000
Acesso à infraestrutura básica			
Esgotamento sanitário precário ou inexistente	0,1911	0,0620	0,0020
Localização geográfica			
Família reside em região metropolitana	-0,5213	0,0675	0,0000
<i>Região de residência da família</i>			
Norte	0,5223	0,1019	0,0000
Nordeste	0,8820	0,0839	0,0000
Sul	-0,2748	0,1196	0,0220
Centro-Oeste	-0,3495	0,1142	0,0020
Intercepto	-1,8627	0,2332	0,0000
N		33255	
N (Fator de expansão da POF)		38559912	
Pseudo-R² (McFadden)		0,3670	
Pseudo-R² (McKelvey e Zavoina)		0,9410	
Wald X²		2122,570	
P-valor (Wald)		0,0000	

Fonte: Elaboração própria a partir dos dados da POF 2008-2009.

Nota: Regressão logit. Variável dependente binária: um, se a família recebe algum benefício do PBF, e zero, caso contrário. Os parâmetros estimados representam a contribuição de cada uma das características observáveis sobre o log das chances de uma família participar do programa governamental em análise.

Antes de avaliar os coeficientes da tabela 3 e o teste de balanceamento das diferenças das médias entre as covariadas do grupo de tratamento e controle após o pareamento, é importante analisar o poder preditivo do modelo de resposta binária usado na estimação do *propensity score*. Uma primeira medida da qualidade

de ajustamento disponível na tabela 3 é o pseudo- R^2 de McFadden, com base no valor do log da verossimilhança ajustado e no valor do log da verossimilhança da regressão apenas com intercepto, em que o grau de ajuste dessa medida foi de 36,7%. Pelo pseudo- R^2 de McKelvey e Zavoina, que relaciona a variabilidade da previsão da variável latente com a variabilidade total, o modelo teve um nível de ajuste de 94%. Por essas duas medidas de ajustamento, conforme Hosmer e Lemeshow (2000) e Long e Freese (2001), o modelo em questão possui um bom grau preditivo¹⁸.

Ademais, outra medida de grau de ajustamento calculada encontra-se detalhada na tabela 4. Essa tabela mensura a qualidade do grau de ajuste, a partir do cruzamento entre os valores da variável de resposta prevista (\hat{T}) e observada (T) referentes à participação no PBF. Para a probabilidade prevista maior ou igual a 50%, considerou-se $\hat{T} = 1$, e para a probabilidade menor que 50%, tomou-se $\hat{T} = 0$.

TABELA 4
Resultados da tabela de classificação para a participação no programa conforme os resultados do modelo *logit*

Classificação Prevista	Resposta observada		Total
	$T = 1$	$T = 0$	
$\hat{T} = 1$	1.755	970	2.725
$\hat{T} = 0$	2.774	27.756	30.530
Total	4.529	28.726	33.255
Sensibilidade	-	-	38,75%
Especificidade	-	-	96,62%
Taxa de falsos negativos	-	-	61,25%
Taxa de falsos positivos	-	-	3,38%
Classificações corretas	-	-	88,74%

Fonte: Elaboração própria.

A proporção de casos corretamente previstos foi de 88,74%, indicando que o número de casos classificados corretamente supera as ocorrências classificadas erroneamente, bem como ratificam a qualidade do ajustamento do modelo. As informações da tabela 4 sinalizam também a existência de um possível suporte comum entre unidades que recebem o bolsa família e unidades que não recebem o benefício, mas que na classificação prevista poderiam participar do programa. Nesse caso, 970 famílias foram classificadas como participantes, quando na verdade não recebem o benefício do programa. Dessa maneira, essa interseção entre famílias não tratadas que poderiam participar do programa e as famílias que são corretamente

18. É válido realçar que os resultados do teste de Wald, que possui como hipótese nula que todos os parâmetros são simultaneamente iguais a zero, indicam que a hipótese nula pode ser rejeitada, a um nível de 1%.

classificadas como participantes é um sinal de que uma das condições necessárias para o PSM está sendo respeitada, que é a hipótese de suporte comum.

Em relação às outras informações da tabela 4, constata-se que a medida de sensibilidade foi de 38,75%, indicando que aproximadamente 40% das famílias que recebem o bolsa família são corretamente identificadas. Por sua vez, a medida de especificidade mostra que 96,62% de famílias não tratadas são corretamente classificadas como famílias que não deveriam receber o benefício do bolsa família. Assim, por meio da área calculada para a curva de Característica de Operação do Receptor¹⁹ (ROC), que relaciona os indicadores de sensibilidade e de especificidade, nota-se que o modelo estimado auferiu uma área sob a curva ROC de 0,90. A informação extraída pela área sob a curva ROC indica que o modelo consegue discriminar de forma satisfatória²⁰ famílias que participam e não participam do PBF.

Retornando a análise para os coeficientes associados, as covariadas do modelo logit expostos na tabela 3, verifica-se em geral que os quatro blocos de variáveis apresentam parâmetros com significância estatística e sinais condizentes com as expectativas de famílias candidatas a participar do programa. Inicialmente, analisando os atributos da família, destaca-se que quanto maior o rendimento *per capita* do domicílio menores as chances de participação no PBF, ao passo que a maior quantidade de indivíduos com idade entre 0 e 17 anos na família aumenta as chances de participação. Essas informações estão consistentes com os critérios de elegibilidade do programa que prioriza aspectos como renda familiar e a idade dos indivíduos. Outra variável importante é a estrutura do núcleo familiar, pois as famílias biparentais (formadas por chefe e cônjuge) têm mais chances de serem beneficiárias do programa em relação às monoparentais, bem como pais e mães com ocupação remunerada aumentam as chances de inserção no programa, sobretudo para o caso das mulheres. Observando o resultado da interação entre o *status* de ocupação da mãe e o estado civil (significativo estatisticamente a 10,0 %), o coeficiente das mulheres casadas que atuam no mercado de trabalho relaciona-se negativamente com a participação no programa.

Nas características da pessoa de referência, em geral, destaca-se que a educação reduz progressivamente a probabilidade de participação no Programa, por exemplo, uma família com chefe analfabeto tem, aproximadamente, três vezes mais chances de participação no PBF em comparação à categoria omitida, ensino médio completo ou mais anos de estudo. Por sua vez, quando o chefe tem o ensino fundamental completo, as chances de participação da família no programa aumentam, mas numa proporção menor que indivíduos menos escolarizados. Os chefes de famílias que

19. A Figura B1, contendo a curva ROC, encontra-se no apêndice B deste artigo.

20. Conforme a regra exposta em Hosmer e Lemeshow (2000, p. 162), um modelo com área sob a curva ROC maior ou igual a 0,70 tem desempenho satisfatório em termos de discriminação de indivíduos de diferentes grupos de respostas.

se declaram de cor branca apresentaram, em média, uma associação positiva com as chances de participação no programa, enquanto a associação mostra-se negativa para chefes com mais idade. Quando o chefe é do sexo feminino, esse resultado é positivo, fato que pode estar relacionado às regras do PBF, que priorizam o recebimento e a administração dos recursos pelas mulheres.

No tocante às características de infraestrutura do domicílio, a precariedade ou inexistência dos serviços de esgotamento sanitário na casa das famílias, um indicador de pobreza, mantém associação positiva com as chances de participar do PBF. Além disso, essa probabilidade é influenciada pela localização geográfica de residência, onde as famílias residentes nas regiões Sul e Centro-Oeste têm um log de chances reduzido em, 0,27 e 0,35, respectivamente, em relação às famílias da região Sudeste. Nas regiões Norte e Nordeste, as chances aumentam, principalmente para as famílias nordestinas, que possuem 2,4 vezes mais chances de receberem o benefício do PBF do que aquelas residentes na região Sudeste. Tal fato é condizente com as desigualdades regionais do país, em que o eixo Norte-Nordeste apresenta os piores indicadores socioeconômicos e, portanto, é o foco principal dos programas de transferências de renda do governo. Nessa mesma direção, os resultados indicam que famílias residentes fora das regiões metropolitanas têm maiores chances de receber o benefício.

Assim, os resultados do modelo logit mostram, sobretudo, a importância dos fatores socioeconômicos para as probabilidades de participar do PBF, pois as famílias com o contexto econômico mais desfavorável e com outras privações no domicílio (como problemas com o esgotamento sanitário e chefe com baixa escolaridade) apresentam um perfil que se enquadra dentro do escopo do programa.

Após a análise dos resultados do modelo logit e do seu grau de ajustamento, é relevante testar o balanceamento dos quatro blocos de características observáveis entre as famílias tratadas e não tratadas após o pareamento, pois a qualidade desse balanceamento é de suma relevância para a obtenção do efeito médio de tratamento. Nesse sentido, a tabela 5 exibe a média para cada uma das características observáveis consideradas nos dois grupos e o teste de hipótese (teste t) de que o valor médio de cada variável é o mesmo no grupo de tratamento e no grupo de controle. É válido destacar que a técnica de pareamento adotada foi a do vizinho mais próximo, respeitando um *calliper* de 0,1% e o *trimming* de 5%.

TABELA 5
Teste de balanceamento das covariadas entre as famílias tratadas e não tratadas após o pareamento

Covariadas	Média		Diferença (A/B-1) Em %	Teste-t	
	Grupo tratado (A)	Grupo de controle (B)		t	P-valor
Atributos da família					
Renda familiar <i>per capita</i> líquida	226,240	222,720	1,580	1,050	0,292
Pessoas entre 0 e 11 anos de idade	1,326	1,364	-0,028	-1,470	0,142
Pessoas entre 12 e 17 anos de idade	0,770	0,756	0,019	1,330	0,184
Pessoas entre 18 e 29 anos de idade	0,734	0,753	-0,025	-1,520	0,129
Proporção de analfabetos na família	0,131	0,129	0,016	0,360	0,720
Status de ocupação do pai	0,649	0,635	0,022	1,230	0,219
Status de ocupação da mãe	0,512	0,521	-0,017	-0,830	0,407
Núcleo familiar					
Família tipo I (biparental)	0,743	0,730	0,018	1,580	0,114
Mãe ocupada e casada	0,352	0,355	-0,009	-0,310	0,756
Atributos do chefe da família					
Idade	41,005	41,267	-0,006	-0,980	0,329
<i>Faixa de escolaridade do chefe</i>					
Analfabeto	0,157	0,158	-0,006	-0,120	0,907
Ensino fundamental incompleto	0,592	0,590	0,003	0,190	0,846
Ensino médio incompleto	0,129	0,128	0,012	0,220	0,824
Raça (cor branca)	0,234	0,232	0,009	0,230	0,821
Gênero (mulher)	0,356	0,379	-0,061	-1,750	0,080
Acesso à infraestrutura					
Esgotamento sanitário precário ou inexistente	0,486	0,496	-0,021	-1,000	0,318
Localização geográfica					
Família reside em região metropolitana	0,261	0,262	-0,002	-0,050	0,961
<i>Região de residência da família</i>					
Norte	0,163	0,162	0,007	0,140	0,886
Nordeste	0,598	0,596	0,003	0,150	0,880
Sul	0,042	0,043	-0,032	-0,320	0,752
Centro-Oeste	0,065	0,076	-0,139	-1,850	0,064

Fonte: Elaboração própria a partir dos dados da POF 2008-2009.

Para a efetividade do PSM, como destacado por Rosenbaum e Rubin (1983) e Khandker, Koolwal e Samad (2010), é necessário que a hipótese de balanceamento seja atendida em sua plenitude, na qual os grupos de tratamento e de controle devem ter em média as mesmas características observáveis. Sendo assim, conforme apresentado na tabela 5, para um nível de significância de ao menos 5%, todas as variáveis explicativas apresentam a mesma média, independentemente da situação do tratamento, indicando que a hipótese nula não pode ser rejeitada para nenhuma das covariadas. Com base nesses resultados, a próxima subseção realiza a avaliação do PBF.

5.2 Avaliação do efeito do PBF sobre a diversificação alimentar

A suposição que baliza esta pesquisa está relacionada com os incrementos de renda promovidos pelo PBF para famílias pobres e com o aumento na quantidade consumida de diferentes itens alimentares, sendo que trabalhos empíricos para o Brasil, como Menezes, Silveira e Azzoni (2008) e Rodrigues, Gomes e Dias (2008), mostram que as famílias pobres possuem alta elasticidade renda para alimentos. Mas o interesse maior dessa investigação é identificar se as famílias beneficiadas apresentam uma cesta de consumo que incorpore diversidade de itens e rigor nutricional, dado que um dos objetivos do PBF gira em torno da segurança alimentar e nutricional dos beneficiados.

Após a estimação do *propensity score* e identificação do grupo de controle, a tabela 6 a seguir apresenta o impacto do PBF sobre as duas medidas de diversificação para as famílias beneficiadas em áreas urbanas do país.

TABELA 6

Efeito médio de tratamento do PBF sobre as medidas de diversificação alimentar
(Vizinho mais próximo com reposição, *calliper* de 0,1% e *trimming* de 5%)

Variável de resultado	Grupo tratado	Grupo de controle	Diferença (ATT)	Erro-padrão (<i>Bootstrap</i>)	Estat.-t	P-valor
Índice de diversificação de Berry (IB)	80,2998	79,4051	0,8947	0,40608	2,20	0,028**
Índice de diversificação saudável (HFD)	7,0482	7,0479	0,0003	0,08974	0,00	0,998

Fonte: Elaboração própria a partir dos dados da POF 2008-2009.

* $p < 0,10$; ** $p < 0,05$; *** $p < 0,01$

Nota: Erro-padrão calculado via método *bootstrap* com 1.000 replicações.

Os resultados da tabela 6 mostram que o efeito médio de tratamento sobre o grupo tratado (ATT) no período estudado apresentou resultados positivos e estatisticamente significativos a 5%, no que concerne ao aumento da diversificação dos itens alimentares medidos pelo IB, que não leva em conta as propriedades nutricionais recomendadas. Dessa forma, as famílias residentes na zona urbana que recebem os benefícios do PBF apresentaram um índice de Berry ligeiramente superior

ao das famílias não tratadas, no qual o efeito do programa foi de aproximadamente 0,9%. Esse efeito mostra-se condizente com as suposições acerca de incrementos de renda para famílias pobres incorporarem novos itens na cesta de alimentos, tendo em vista a alta elasticidade renda para bens alimentares.

Não obstante, ao verificar o ATT para o índice de diversificação saudável, que pondera cada item consumido por fatores nutricionais, constata-se que a diferença entre os indicadores dos grupos tratados e de controle não é estatisticamente diferente de zero, evidenciando que o PBF não gera impactos em termos de uma diversificação no consumo de alimentos saudáveis.

Considerando as estimativas do ATT obtidos por outras técnicas de pareamento (ver Tabela C1 no Apêndice), verifica-se que, de forma global, os resultados exibidos na tabela 6 são mantidos, pois as famílias beneficiadas possuem uma diferença positiva no IB, quando comparadas às famílias não tratadas, e o efeito médio do programa sobre o HFD permanece sem significância estatística.

Portanto, os achados sinalizam que o pequeno aumento na variedade de itens consumidos pelas famílias tratadas não foi acompanhado por melhorias na qualidade da alimentação, visto que as famílias presentes no grupo de controle apresentaram estatisticamente o mesmo nível do índice de diversificação saudável. Dessa maneira, as preocupações evidenciadas em Pereda e Alves (2012) sobre a relação positiva entre renda e consumo de itens com propriedades nutricionais nocivas para as famílias pobres, parecem balizar os impactos não significativos do PBF sobre o consumo diversificado com propriedades mais saudáveis.

Tendo em vista a importância de uma cesta de consumo diversificada que respeite uma dieta balanceada em termos de nutrientes, o PBF poderia atuar de forma mais efetiva na mudança da qualidade alimentar das famílias beneficiadas. O PBF exige uma série de condicionalidades por parte dos beneficiários, que abarcam as áreas de educação, tal como a exigência de matrícula da criança na rede regular de ensino e frequência escolar de no mínimo 85% e saúde. No campo da saúde, as condicionalidades, conforme o MDS (2013), são as seguintes:

- inscrição no pré-natal e comparecimento às consultas na unidade de saúde por parte de gestantes e nutrízes;
- participação das atividades educativas ofertadas pelas equipes de saúde sobre aleitamento materno e promoção da alimentação saudável;
- condução da criança às unidades de saúde ou aos locais de vacinação e realização do acompanhamento do estado nutricional e do desenvolvimento.

Dessa forma, tais condicionalidades atreladas à área de saúde deveriam mostrar-se mais efetivas na parte referente à promoção do consumo de alimentos

saudáveis. Assim, o Programa de Saúde da Família (PSF), que atua na atenção primária dos cuidados com a saúde e possui equipes multiprofissionais, poderia exercer esse importante canal para a melhoria no nível de diversificação saudável de alimentos, visto que a necessidade do cumprimento das condicionalidades por parte das famílias participantes do programa.

6 CONSIDERAÇÕES FINAIS

A diversificação alimentar, do ponto de vista nutricional, é essencial para o desenvolvimento saudável do indivíduo, enquanto, sobre aspectos econômicos, for importante, pois incide indiretamente na formação do capital humano. Diante desse contexto, o presente artigo investigou o impacto do PBF em medidas de diversificação do consumo de alimentos das famílias tratadas.

De forma geral, os resultados auferidos neste trabalho sinalizam que o programa de transferência de renda direta – o Bolsa Família –, apesar de apresentar um resultado positivo sobre o nível de consumo por uma cesta com mais variedade de itens alimentares, não consegue contribuir para mudanças em termos de uma cesta diversificada por mais itens saudáveis. Conforme Wahlqvist (2003), uma cesta diversificada de itens de alimentação é importante para o estoque de saúde corrente e do futuro dos indivíduos, atuando, por exemplo, em melhores perspectivas de uma vida mais prolongada e com mais qualidade.

Sendo assim, uma melhor gestão das condicionalidades do PBF em termos, por exemplo, do cumprimento ativo dos itens relacionados ao acompanhamento nutricional e da promoção da alimentação saudável, poderia ampliar ainda mais os efeitos dessa intervenção governamental em relação ao consumo com mais variedade de itens e, sobretudo, que respeitem as boas propriedades nutricionais.

REFERÊNCIAS

AIRES, K. DA S.; GOMES, D. S.; ESMERALDO, G. G. S. L. Bolsa Família como Política Pública no Assentamento Rural: Contribuições para a Segurança Alimentar. *In*: V JOINPP, 2011, São Luís. **Anais...** São Luís: Jornada Internacional de Políticas Públicas, 2011. Disponível em: <<http://www.joinpp.ufma.br/jornadas/joinpp2011/CdVjornada/vjornada.html>>. Acesso em: 5 abr. 2013.

ARAÚJO, G. S.; RIBEIRO, R.; NEDER, H. D. Impactos do Programa Bolsa Família sobre o Trabalho de Crianças e Adolescentes Residentes na Área Urbana em 2006. **Revista Economia**, v. 11, n. 4, p. 57-102, 2010.

ASSUNÇÃO, V. K.; LEITAO, M. R. F. A.; INÁCIO, P. H. D. Comer mais e melhor: Os impactos do programa Bolsa Família na alimentação de famílias de pescadores artesanais de Pernambuco. **Amazônica**, v. 4, n. 2, p. 336-353, 2012.

BAPTISTELLA, J. C. F. **Avaliação de Programas Sociais**: uma análise do impacto do Bolsa Família sobre o consumo de alimentos. 2012. 101 f. Dissertação de Mestrado em Economia, Universidade Federal de São Carlos, Sorocaba, 2012.

BARROS, R. P. B.; CARVALHO, M.; FRANCO, S.; MENDONÇA, R. **A queda recente da desigualdade de renda no Brasil**. Rio de Janeiro: Ipea, 2007 (Texto para Discussão n. 1258).

BECKER, S. O.; Inchino, A. Estimation of average treatment effects based on propensity scores. **The Stata Journal**, v. 2, n. 4, p. 358-377, 2002.

BERRY, C. H. Corporate growth and diversification. **Journal of law and economics**, v. 14, n. 2, p. 371-383, 1971.

BRASIL. **Avaliação do Programa Bolsa Alimentação**: Primeira Fase. Ministério da Saúde. Brasília-DF. 2004.

_____. **Avaliação do Programa Bolsa Alimentação**: Segunda Fase. Ministério da Saúde. Brasília, DF. 2005.

_____. **Dez passos para uma alimentação saudável**. Ministério da Saúde. Brasília, DF. 2013. Disponível em: <http://189.28.128.100/nutricao/docs/geral/guia_alimentar_bolso.pdf>. Acesso em: 22 jan. 2014.

_____. Instituto Brasileiro de Análises Sociais e Econômicas. **Repercussões do Programa Bolsa Família na segurança alimentar e nutricional das famílias beneficiadas**. Rio de Janeiro: Ibase, 2008 (Relatório Técnico). Disponível em: <http://www.ibase.br/userimages/ibase_bf_sintese_site.pdf>. Acesso em: 20 jun. 2013.

_____. Ministério do Desenvolvimento Social e Combate à Fome. **Bolsa Família**, 2013. Disponível em: <<http://www.mds.gov.br/bolsafamilia>>. Acesso em: 12 abr. 2013.

CAMELO, R. S.; TAVARES, P. A.; SAIANI, C. S. Alimentação, nutrição e saúde em programas de transferência de renda: evidências para o Programa Bolsa Família. **Economia**, v. 10, n.4, p. 685-713, 2009.

CAMERON, A. C.; TRIVEDI, P. K. **Microeconometrics**: methods and applications. Cambridge University Press, 2005.

CALIENDO, M.; KOPEINIG, S. Some practical guidance for the implementation of Propensity Score Matching. **Journal of Economic Surveys**, v. 22, n. 1, p. 31-72, 2008.

COELHO, A. B.; AGUIAR, D. R. D.; EALES, J. S. Food Demand in Brazil: an Application of Shonkwiler & Yen Two-Step Estimation Method. **Estudos Econômicos**, v. 40, n. 1, p. 185-211, 2010.

DUARTE, G. B.; SAMPAIO, B.; SAMPAIO, Y. Programa Bolsa Família: impacto das transferências sobre os gastos com alimentos em famílias rurais. **Revista de Economia e Sociologia Rural**, v. 47, n. 4, p. 903-918, 2009.

DRESCHER, L. S., THIELE, S., MENSINK, G. B. M. A New Index to Measure Healthy Food Diversity Better Reflects a Healthy Diet Than Traditional Measures. **The Journal of Nutrition**, v. 137, p. 647-651, 2007.

DRESCHER, L. S. *et al.* Consumer demand for healthy eating considering diversity: an economic approach for German individuals. **International Journal of Consumer Studies**, v. 33, n.6, p. 684-696, 2009.

FERRAZ, L. F. **Programa Bolsa Família: Impactos na distribuição de renda.** Monografia (Especialização em Orçamento Público), Tribunal de Contas da União. Brasília, 2008.

FUNDAÇÃO OSWALDO CRUZ. **O que é insegurança alimentar.** 2011. Disponível em: <<http://www.fiocruz.br/jovem/cgi/cgilua.exe/sys/start.htm?infoid=703&sid=25>>. Acesso em: 10 abr. 2013.

GOLLOP, F. M.; MONAHAN, J. L. A Generalized Index of Diversification: Trends in U.S. Manufacturing. **The Review of Economics and Statistics**, MIT Press, v. 73, n. 2, p. 318-30, 1991.

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

HERMETO, A. M.; ROMERO, J. R. Avaliação de Impacto do Programa Bolsa Família sobre Indicadores Educacionais: uma abordagem de regressão descontínua. In: 37º Encontro Nacional de Economia, 2009, Foz do Iguaçu. **Anais do 37º Encontro Nacional de Economia**, 2009.

HOSMER, David W.; LEMESHOW, Stanley. **Applied Logistic Regression.** 2. ed. New York: John Wiley & Sons, 2000.

KASSOUF, A. L.; GLEWWE, P. The Impact of the Bolsa Escola/Familia Conditional Cash Transfer Program on Grade Advancement and Dropout Rates in Brazil. In: **5th Midwest International Economic Development Conference**, Madison, Wiconsin, 2008.

KHANDKER, R. S.; KOOLWAL, G. B.; SAMAD, H. A. **Handbook on Impact Evaluation: Quantitative Methods and Practices.** Washington, DC: The World Bank, 2010.

KOHLMEIER, L. *et al.* Lycopene and myocardial infarction risk. **The Euramic Study. Am. J. Epidemiol**, v. 146, p. 618-626, 1993.

LEE, J. Y. The Demand for Varied Diet with Econometric Models for Count Data. **American Journal of Agricultural Economics**, v. 69, n. 3: 687-692, 1987.

LEE, J. Y.; BROWN; M. G. Consumer Demand for Food Diversity. **Southern Journal of Agricultural Economics**, v. 163, p. 47-53, 1989.

LONG, J. SCOTT; FREESE, JEREMY. **Regression models for categorical dependent variables using stata**. Texas: Stata Press, 2001.

MATTOS, E.; MAIA, S.; MARQUES, F. Evidências da Relação entre Oferta de Trabalho e Programas de Transferência de Renda no Brasil: Bolsa Escola versus Renda Mínima. **Pesquisa e Planejamento Econômico**, v. 40, n. 2, p. 237-279, 2010.

MELO, R. M. S.; DUARTE, G. B. Impacto do Programa Bolsa Família sobre a frequência escolar: o caso da agricultura familiar no Nordeste do Brasil. **Revista de Economia e Sociologia Rural**, v. 48, n. 3, p. 635-657, 2010.

MENEZES, T. A.; SILVEIRA, F. G.; AZZONI, C. R. Demand elasticities for food products: a two-stage budgeting system. **Applied economics**, v. 40, n. 19, p. 2557-2572, 2008.

MENEZES FILHO, N. (Org.). **Avaliação econômica de projetos sociais**. São Paulo: Dinâmica Gráfica e Editora, 2012.

MOON, W. *et al.* Demand for food variety in an emerging market economy. **Applied Economics**, n. 34, p. 573-581, 2002.

PEREDA, P. C.; ALVES, D. C. O. Qualidade alimentar dos brasileiros: teoria e evidência usando demanda por nutrientes. **Pesquisa e Planejamento Econômico**, v. 42, n. 2, p. 239-260, 2012

PEREIRA, L. L. *et al.* Avaliação do Programa Bolsa Família em municípios de baixo Índice de Desenvolvimento Humano e cumprimento das condicionalidades de saúde. **Tempus: Atas de Saúde Coletiva**, v. 5, p. 139-165, 2011.

PHILIPPI, S. T. *et al.* Pirâmide Alimentar Adaptada: guia para escolha dos alimentos. **Revista de Nutrição**, v. 12, n. 1, p. 65-80, 1999.

RESENDE, A. C. C.; OLIVEIRA, A. M. H. C. Avaliando resultados de um programa de transferência de renda: o impacto do Bolsa-Escola sobre os gastos das famílias brasileiras. **Estudos Econômicos**, v. 38, n. 2, p. 235-265, 2008.

RIBEIRO, R.; CACCIAMALI, M. C. Impactos do Programa Bolsa-Família sobre os Indicadores Educacionais. **Economia**, v. 13, n. 2, p. 415-446, 2012.

ROCHA, L.A.; KHAN, A. S.; LIMA, P.V.P.S. Impacto do Programa Bolsa família sobre o bem-estar das famílias beneficiadas no estado do Ceará. *In*: Carvalho, E. B. S.; Holanda, M. C.; Barbosa, M. P. (Org.). **Economia do Ceará em Debate 2008**, Fortaleza: Ipece, 2009.

RODRIGUES, C. T.; GOMES, A. P.; DIAS, R. S. A expansão do programa Bolsa Família e as implicações sobre consumo e preços de alimentos: O caso de Viçosa/MG. In: 46º Congresso da SOBER, 2008, Rio Branco. **Anais...** Rio Branco: Sociedade Brasileira de Economia, Administração e Sociologia Rural, 2008. Disponível em: <<http://purl.umn.edu/106090>>. Acesso em: 5 abr. 2013.

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

ROYO-BARDONABA, M. A. (Org.). **Nutrición en Salud Pública**. Madrid: Ministerio de Sanidad y Consumo, Instituto de Salud Carlos III, 2007.

SATYRO, N.; SOARES, S. S. **O Programa Bolsa Família: desenho institucional, impactos e possibilidades futuras**. Rio de Janeiro: Ipea, 2009 (Texto para Discussão n. 1424).

STEWART, H.; HARRIS, J. M. Obstacles to Overcome in Promoting Dietary Variety: The Case of Vegetables. **Review of Agricultural Economics**, v. 27, n. 1, p. 21-36, 2005.

TAVARES, P. A. *et al.* Uma avaliação do programa Bolsa Família: focalização e impacto na distribuição de renda e pobreza. **Pesquisa e Planejamento Econômico**, v. 39, n. 1, p. 25-58, 2009.

TRALDI, D. R. C. **Efetividades e Entraves do Programa Bolsa Família no Município de Araraquara/SP: um olhar sobre a segurança alimentar dos beneficiários**. Araraquara, 2011. Dissertação (Mestrado em Desenvolvimento Regional e Meio Ambiente), Centro Universitário de Araraquara, Araraquara, 2011.

THIELE, S.; WEISS, C. Consumer Demand for Food Diversity: Evidence for Germany. **Food Policy**, v. 28, p. 99-115, 2003.

PAULA, D. V. *et al.* Avaliação nutricional e padrão de consumo alimentar entre crianças beneficiárias e não beneficiárias de programas de transferência de renda, em escola municipal do Município de Belo Horizonte, Estado de Minas Gerais, Brasil, em 2009. **Epidemiologia e Serviços de Saúde**, v. 21, n. 3, p. 385-394, 2012.

WAHLQVIST, M. L. Regional food diversity and human health. **Asia Pacific Journal of Clinical Nutrition**, v. 12, n. 3, p. 304-308, 2003.

WOOLDRIDGE, J. M. **Econometric analysis of cross section and panel data**. 2. ed. MIT Press, 2010.

(Originais submetidos em outubro de 2013. Última versão recebida em dezembro de 2014. Aprovada em março de 2015.)

APÊNDICE A

DESCRIÇÃO DOS ITENS ALIMENTARES CONSIDERADOS E DAS COVARIADAS DO MODELO EMPÍRICO

TABELA A1

Descrição dos itens alimentares* com os seus respectivos valores de fator saúde (FS), peso do grupo (GW) e peso do subgrupo (Gb)

Item	GW	Gb	FS	Item	GW	Gb	FS	Item	GW	Gb	FS
Banana	0,74	0,36	0,2664	Outros tubérculos	0,74	0,2	0,148	Açúcar	0,74	0,04	0,0296
Laranja	0,74	0,36	0,2664	Mandioca	0,74	0,2	0,148	Refrigerantes	0,74	0,04	0,0296
Sucos naturais	0,74	0,36	0,2664	Batata inglesa	0,74	0,2	0,148	Chocolate	0,74	0,04	0,0296
Goiaba	0,74	0,36	0,2664	Soja	0,74	0,12	0,0888	Outros enlatados	0,74	0,04	0,0296
Alface	0,74	0,36	0,2664	Café moído	0,74	0,12	0,0888	Achocolatados	0,74	0,04	0,0296
Cebola	0,74	0,36	0,2664	Pão francês	0,74	0,12	0,0888	Outros doces	0,74	0,04	0,0296
Erva mate	0,74	0,36	0,2664	Barra de cereais diet	0,74	0,12	0,0888	Refrigerantes	0,74	0,04	0,0296
Água de coco	0,74	0,36	0,2664	Massas de mandioca	0,74	0,12	0,0888	Sal	0,74	0,04	0,0296
Castanha	0,74	0,36	0,2664	Farinhas diversas	0,74	0,12	0,0888	Pizzas	0,74	0,04	0,0296
Coco	0,74	0,36	0,2664	Massa de milho	0,74	0,12	0,0888	Balás e chicles light	0,74	0,04	0,0296
Cenoura	0,74	0,36	0,2664	Massa de trigo	0,74	0,12	0,0888	Alimentos preparados	0,74	0,04	0,0296
Outras verduras	0,74	0,36	0,2664	Farinha de trigo	0,74	0,12	0,0888	Outras bebidas	0,74	0,04	0,0296
Outras frutas	0,74	0,36	0,2664	Macaráo	0,74	0,12	0,0888	Salgadinhos	0,74	0,04	0,0296
Manga	0,74	0,36	0,2664	Outros grãos	0,74	0,12	0,0888	Sorvete/picolé	0,74	0,04	0,0296
Maracujá	0,74	0,36	0,2664	Amido de cereal	0,74	0,12	0,0888	Chocolates light/diet	0,74	0,04	0,0296
Uva	0,74	0,36	0,2664	Barra de cereais	0,74	0,12	0,0888	Massas light/diet	0,74	0,04	0,0296
Milho	0,74	0,36	0,2664	Arroz	0,74	0,12	0,0888	Açúcar light/diet	0,74	0,04	0,0296
Outras hortaliças	0,74	0,36	0,2664	Feijão	0,74	0,12	0,0888	Complementos alimentares	0,74	0,04	0,0296

(Continua)

(Continuação)											
Item	GW	Gb	FS	Item	GW	Gb	FS	Item	GW	Gb	FS
Açai	0,74	0,36	0,2664	Farina de cereais	0,74	0,12	0,0888	Bebidas light/diet	0,74	0,04	0,0296
Melancia	0,74	0,36	0,2664	Doces light/diet	0,74	0,04	0,0296	Outros pescados	0,21	0,36	0,0756
Ervilha	0,74	0,36	0,2664	Pizzas light/diet	0,74	0,04	0,0296	Frutos do mar	0,21	0,36	0,0756
Mamão	0,74	0,36	0,2664	Biscoito light/diet	0,74	0,04	0,0296	Camarão fresco	0,21	0,36	0,0756
Chá	0,74	0,36	0,2664	Massas diversas	0,74	0,04	0,0296	Peixes frescos	0,21	0,36	0,0756
Tomate	0,74	0,36	0,2664	Condimentos	0,74	0,04	0,0296	Derivados do leite light	0,21	0,28	0,0588
Maçã	0,74	0,36	0,2664	Café preparado	0,74	0,04	0,0296	Leite de vaca	0,21	0,28	0,0588
Melão	0,74	0,36	0,2664	Massas preparadas	0,74	0,04	0,0296	logurtes light/diet	0,21	0,28	0,0588
Frutas cítricas	0,74	0,36	0,2664	Balas e chicletes	0,74	0,04	0,0296	Queijos light/diet	0,21	0,28	0,0588
Outros panificados	0,74	0,28	0,2072	Sanduíches	0,74	0,04	0,0296	Leite em pó	0,21	0,2	0,042
Panificados light/diet	0,74	0,28	0,2072	Maionese	0,74	0,04	0,0296	logurtes	0,21	0,2	0,042
Pão integral	0,74	0,28	0,2072	Legumes em conserva	0,74	0,04	0,0296	Derivados do leite	0,21	0,2	0,042
Arroz integral	0,74	0,28	0,2072	Biscoito	0,74	0,04	0,0296	Outros leites light/diet	0,21	0,2	0,042
Feijão orgânico	0,74	0,28	0,2072	Massa de tomate	0,74	0,04	0,0296	Queijos	0,21	0,2	0,042
Aveia	0,74	0,28	0,2072	Sorvete/picolé light	0,74	0,04	0,0296	Outros leites	0,21	0,2	0,042
Soja orgânica	0,74	0,28	0,2072	Grãos em conserva	0,74	0,04	0,0296	Outras aves	0,21	0,12	0,0252
Peixes em conserva	0,21	0,12	0,0252	Ovo de galinha	0,21	0,12	0,0252	Peru	0,21	0,12	0,0252
Carne bovina segunda	0,21	0,12	0,0252	Outras carnes	0,21	0,12	0,0252	Carne bovina terceira	0,21	0,12	0,0252
Carne de hambúrguer	0,21	0,12	0,0252	Presunto	0,21	0,12	0,0252	Carne bovina primeira	0,21	0,12	0,0252
Frango	0,21	0,12	0,0252	Carnes em conserva	0,21	0,12	0,0252	Carne suína	0,21	0,12	0,0252
Carne caprina	0,21	0,12	0,0252	Azeite de oliva	0,05	0,28	0,014	Peixes industrializados	0,21	0,12	0,0252
Carnes industrializadas	0,21	0,04	0,0084	Margarina	0,05	0,12	0,006	Bacon, mortadela, linguiça	0,21	0,04	0,0084
Óleo de soja	0,05	0,28	0,014	Manteiga	0,05	0,12	0,006	Visceras suína de segunda	0,21	0,04	0,0084

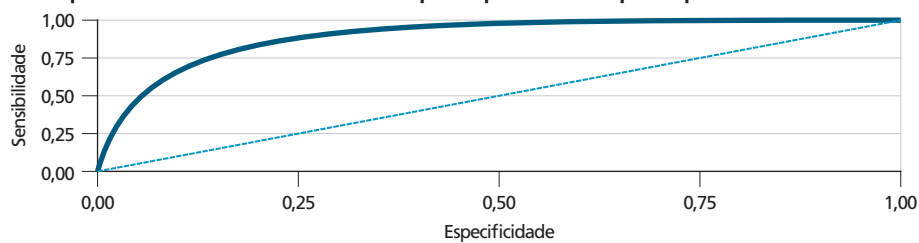
Fonte: Elaboração própria a partir dos dados da POF 2008-2009/IBGE. Classificação baseada em Drescher, Thiele e Mensink (2007) e Brasil (2013).

* Os produtos alimentares totalizam 123 itens.

APÊNDICE B

CURVA DE CARACTERÍSTICA DE OPERAÇÃO DO RECEPTOR (ROC) PARA O MODELO LOGIT

FIGURA B1

Curva ROC: relação entre as medidas de especificidade e sensibilidade para a previsibilidade correta de famílias participantes e não participantes do PBF

Área sob a curva ROC = 0,9005

APÊNDICE C

EFEITO MÉDIO COM DIFERENTES TÉCNICAS DE PAREAMENTO

TABELA C1

Efeito médio de tratamento do PBF sobre as medidas de diversificação alimentar com diferentes técnicas de pareamento

Método de pareamento	Variável de resultado	Grupo tratado	Grupo de controle	Diferença (ATT)	Erro-padrão	Estat.-t	P-valor
Dez vizinhos mais próximos	IB	80,31	79,76	0,55	0,32	1,72	0,0856*
	HFD	7,04	6,99	0,05	0,07	0,65	0,5156
Vizinho mais próximo sem reposição	IB	80,31	79,67	0,63	0,30	2,13	0,0332**
	HFD	7,04	7,09	-0,05	0,07	-0,78	0,4354
Raio	IB	80,30	79,62	0,68	0,32	2,12	0,0340**
	HFD	7,05	7,00	0,04	0,08	0,57	0,5686
Kernel	IB	80,31	79,43	0,88	0,40	2,17	0,0300**
	HFD	7,04	7,06	-0,03	0,09	-0,29	0,7718

Fonte: Elaboração própria a partir dos dados da POF 2008-2009.

* $p < 0,10$; ** $p < 0,05$; *** $p < 0,01$

EXPANSÃO DA ESCOLARIDADE E REDUÇÃO DA DESIGUALDADE REGIONAL DE RENDA NO BRASIL ENTRE 1995 E 2011: PROGRESSOS RECENTES E DESAFIOS PRESENTES

Rodrigo Carvalho Oliveira¹

Raul da Mota Silveira Neto²

A despeito de seus elevados níveis, tem sido verificada uma queda das disparidades de renda no Brasil desde meados da década de 1990, tanto quando se analisa a desigualdade entre pessoas, quanto entre as regiões (Silveira-Neto e Azzoni, 2012; Soares, 2006; Hoffmann, 2006; Barros *et al*, 2010). Tal movimento de desconcentração regional de renda, além disso, tem sido acompanhado de redução das disparidades regionais de escolaridade. Este trabalho tem por objetivo, pois, analisar a importância das fontes de rendas do trabalho associadas à escolaridade sobre a queda na desigualdade de renda regional observada entre 1995 e 2011. Os resultados indicam que a dinâmica da renda do trabalho, com destaque para os indivíduos com no mínimo ensino médio completo, e da renda dos programas sociais (Bolsa Família e Benefício de Prestação Continuada) foram as principais responsáveis pela queda da desigualdade no período entre 1995 e 2011. Por outro lado, a dinâmica de distribuição da renda dos indivíduos com ensino superior entre os estados atuou no sentido de elevar a concentração regional de renda no país.

Palavras-chave: desigualdade regional; educação; programas sociais.

EDUCATION EXPANSION AND REDUCTION OF BRAZILIAN REGIONAL INCOME INEQUALITY BETWEEN 1995 AND 2011: RECENT PROGRESS AND CHALLENGES

Among the socioeconomic problems in Brazil, the personal and regional income inequality and the low educational level are notorious. Despite of the high inequality level, it has been observed a permanent fall both in the personal inequality and in the regional inequality since the final of 1990's (Silveira Neto e Azzoni, 2012; Soares, 2006; Hoffmann, 2006; Barros *et al*, 2010). Besides, there was a simultaneous movement of reduction of regional inequalities of schooling among Brazilian states. The aim of this work is, thus, to decompose the regional Gini index to compute the importance of labors income related to the education, and the importance of non-labors income for the fall of inequality observed in the recent years. The results suggest that the labor income, particularly for individuals with at least secondary education, and the incomes from social programs were the responsible for the fall of inequality in the period between 1995 and 2011. On the other hand, the dynamic of distribution of income for those with a college degree among the states have been acting to avoid stronger reduction in regional inequality.

Keywords: regional inequality; education; social programs.

JEL: R11; R58.

1. Professor do curso de Economia da Universidade Federal da Bahia (UFBA), doutorando em Economia no Pimes/UFPE, bolsista de Doutorado do CNPq.

2. Professor do curso de Economia e de pós-graduação em Economia da Universidade Federal de Pernambuco (Pimes/UFPE), bolsista de Produtividade do CNPq.

1 INTRODUÇÃO

Não parece haver dúvidas entre os pesquisadores sobre a realidade social brasileira e o progresso do país no que diz respeito à redução da pobreza e da desigualdade, uma trajetória de melhoria que se iniciou a partir do fim do período de inflação elevada em meados dos anos 1990 e intensificada nos anos 2000 (Barros, *et al.* 2006; Soares, 2006, Hoffman e Ney, 2008; Osório e Souza, 2012; Souza, 2103). Diretamente ligadas a esses resultados mais favoráveis do ponto de vista da melhoria das condições de vida dos indivíduos, verifica-se também, neste período, como mostraram recentemente Silveira Neto e Azzoni (2011, 2012), uma tendência de diminuição das disparidades regionais de renda entre as unidades federativas do país, fato alentador, dada a conhecida persistência dessas diferenças, sobretudo quando se considera a livre mobilidade espacial de indivíduos (Azzoni, 1997).

Na verdade, como mostraram Silveira Neto e Azzoni (2012), tal redução recente das disparidades regionais de renda parece ser explicada tanto por movimentos associados aos programas de transferência de renda, como por movimentos de renda do trabalho. Note-se que tais movimentos em favor de menores desigualdades regionais observadas para a renda do trabalho, a princípio, são consistentes, por um lado, com a conhecida importância da educação para entender os níveis de desigualdade regional de renda no Brasil (Duarte, Ferreira e Salvato, 2003; Silveira Neto e Menezes, 2008) e, por outro, com o aumento do nível de escolaridade dos indivíduos em todas as unidades da federação (mudanças na oferta de trabalho) e a elevação do número de postos de trabalho para os indivíduos mais escolarizados, movimento associado à desconcentração industrial regional (Silveira Neto e Azzoni, 2011) observada no país na última década (mudanças na demanda de trabalho).

É verdade que, dado o conhecido padrão brasileiro de disparidades regionais, em que os estados mais pobres também apresentam maior concentração relativa de indivíduos abaixo da linha de pobreza (Silveira Neto, 2005), o foco dos atuais programas públicos de transferência de renda nos indivíduos em situação de maior precariedade social termina por impor um viés espacial absolutamente não intencional na alocação de recursos desses programas entre os estados brasileiros.

Contudo, seja porque representam mais de 70% da renda das famílias, seja porque refletem potenciais movimentos associados à produtividade, são as fontes de renda do trabalho associadas aos níveis educacionais os principais protagonistas dos movimentos de distribuição regional de renda no Brasil. Nesse sentido, note-se, por exemplo, que enquanto os estados mais pobres, como Maranhão e Alagoas, viram expandir os anos médios de estudos de sua população em 69,25% e 47,69%, entre 1995 e 2011, os estados de São Paulo e Rio de Janeiro apresentaram expansão da escolaridade de apenas 23,76% e 35,89%, respectivamente, no mesmo período.

Esse ambiente social brasileiro com progresso regional mais homogêneo nos níveis de escolaridade, de políticas sociais e de transferências de renda efetivas certamente tem implicações sobre a distribuição regional de renda entre os estados brasileiros. Considerando-se a ausência de análises que considerem simultaneamente o papel da escolaridade e das políticas sociais no entendimento das disparidades regionais brasileiras no período mais recente, o objetivo central deste trabalho é, pois, fornecer um conjunto de evidências a respeito das contribuições dos diferentes níveis de escolaridade para o entendimento e a redução dos níveis de desigualdade regional de renda observados no Brasil entre 1995 e 2011. Tal tarefa é levada a efeito com fundamento na decomposição do índice de Gini da distribuição regional de renda no país entre diferentes fontes de renda dos domicílios das unidades federativas em uma estratégia semelhante àquela empreendida por Silveira Neto e Azzoni (2012), mas que considera a renda do trabalho tomando como base os diferentes níveis de escolaridade. Adicionalmente, por meio do cálculo de elasticidades desse índice de desigualdade em relação às diferentes fontes de renda dos estados, o trabalho fornece medidas da influência de mudanças nessas diferentes fontes de renda associadas aos diferentes níveis de escolaridade sobre a desigualdade regional de renda no país.

Entre outras evidências, os resultados obtidos na pesquisa indicam que em 2011 cerca de 65% da desigualdade regional de renda entre os estados brasileiros decorria dos níveis (elevados) de concentração da renda dos indivíduos que possuíam escolaridade com, no mínimo, o ensino médio, um percentual superior ao observado em 1995. Além disso, a dinâmica de distribuição regional da fonte de renda associada aos indivíduos mais escolarizados (ensino superior), entre as rendas do trabalho, foi a única a apresentar um movimento em favor de maior concentração regional de renda no Brasil entre 1995 e 2011. Tais resultados indicam que a disparidade regional de renda no país é, hoje, mais dependente da distribuição regional de renda dos indivíduos mais escolarizados.

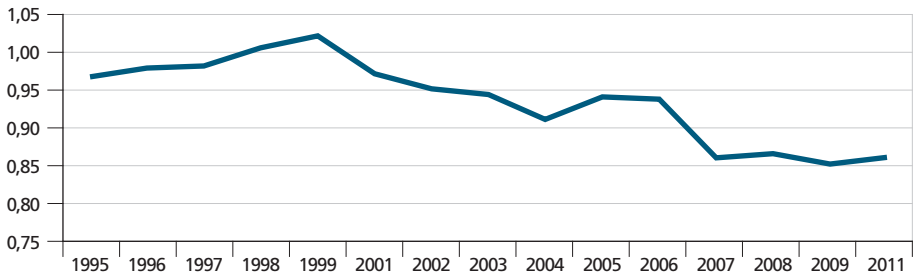
Além dessa introdução, o trabalho estrutura-se em mais três seções. Na próxima seção: *i*) a base de informações utilizadas; *ii*) as diferentes fontes de renda consideradas; e *iii*) as evidências exploratórias iniciais sobre a importância das diferentes fontes de renda associadas aos diferentes níveis de escolaridade para a desigualdade regional de renda no Brasil. Na seção três, são apresentados: *i*) os resultados do trabalho quanto à importância das fontes de renda e seus níveis de distribuição entres os estados brasileiros nos anos de 1995 a 2011; e *ii*) os resultados do trabalho em relação às contribuições das diferentes fontes de renda para a redução da desigualdade regional de renda no país. Na seção quatro são apresentadas as estimativas da elasticidade da desigualdade regional de renda em relação às fontes de renda nos dois referidos anos. As conclusões e as implicações dos resultados são apresentadas na quinta e na última seção.

2 BASE DE DADOS E EVIDÊNCIAS INICIAIS

As informações utilizadas nesta pesquisa provêm dos microdados da Pesquisa Nacional por Amostra de Domicílio (Pnad) do Instituto Brasileiro de Geografia e Estatística (IBGE). A partir de tal base, é possível desagregar as informações para as unidades da federação, como também decompor as informações de renda em relação a diferentes características dos domicílios. Mais especificamente, de acordo com a Pnad, é possível extrair as informações necessárias sobre escolaridade e renda dos domicílios, tanto da renda do trabalho, quanto da renda de outras fontes, como aposentadorias, pensões, renda do aluguel e renda de programas sociais. Tais fontes de renda dos domicílios são, posteriormente, agregadas para as unidades da federação utilizando-se os pesos domiciliares, conforme a estratégia de Silveira Neto e Azzoni (2012). Com base nas informações para diferentes fontes de renda por unidades da federação e das populações, obtêm-se os níveis de renda *per capita* das unidades federativas e as medidas de desigualdade na distribuição da renda *per capita* entre tais unidades. Na presente pesquisa, por permitir diferentes estratégias de decomposição e obtenção de resultados para a elasticidade da desigualdade, utiliza-se o índice de Gini como medida de desigualdade (Silveira Neto e Azzoni, 2012; Soares, 2006; e Hoffmann, 2006).

No gráfico 1, observa-se a evolução do desvio-padrão da distribuição da escolaridade média (número de anos de estudo) entre os estados brasileiros. Neste, pode-se observar uma clara tendência de redução das disparidades educacionais entre os estados brasileiros desde 1999. Esse fato provavelmente contribui para a redução das disparidades econômicas regionais, pois maiores níveis de escolaridade estão associados, de um modo geral, a rendas mais elevadas.

GRÁFICO 1
Evolução do desvio padrão da distribuição do número médio de anos de estudos entre as Unidades da Federação



Fonte: elaboração própria, com base nos dados da Pnad/IBGE.

Evidentemente, o impacto desta tendência de convergência dos níveis de escolaridade média entre os estados do Brasil sobre as disparidades regionais é dependente da natureza da expansão educacional dos estados (ou seja, de quais

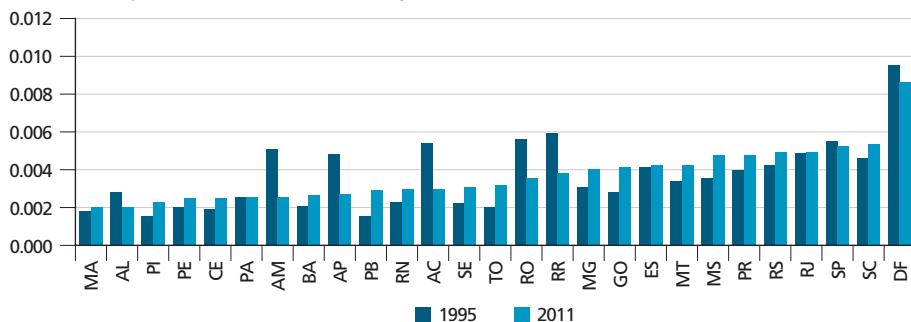
níveis de escolaridade foram expandidos) e das remunerações regionais dos diferentes níveis de escolaridade. Tal questão é tratada a seguir, contudo, nos gráficos 2 e 3, pode-se observar a participação de cada estado no total da renda *per capita* da fonte de renda das pessoas com ensino médio completo e ensino superior completo, respectivamente. Os estados estão ordenados segundo a renda *per capita* total no ano de 1995 e três evidências merecem destaque.

Primeiro, os valores dos gráficos indicam que os estados com maior renda *per capita* total também são aqueles com maior participação nas rendas *per capita* dos indivíduos mais escolarizados. Por exemplo, enquanto para o Distrito Federal a parcela na renda total derivada dos indivíduos com ensino superior estava acima de 1,6%, em 2011, para o Maranhão tal parcela situava-se abaixo de 0,1%, do total. Segundo, a renda *per capita* referente às pessoas com ensino médio (Gráfico 2) é mais dispersa entre os estados do que a renda dos indivíduos com ensino superior (Gráfico 3), a qual apresenta forte concentração nos estados mais ricos. Finalmente, ao longo do tempo, observa-se um maior crescimento da participação dos estados mais pobres na renda dos indivíduos com ensino médio completo que para o caso dos indivíduos com ensino superior.

Mais especificamente, enquanto quatro dos cinco estados mais pobres apresentaram aumento da participação na renda total associada aos indivíduos com ensino médio ou ensino superior incompleto, entre 1995 e 2011, apenas dois dos cinco estados mais ricos apresentaram aumento de suas participações na renda total desta fonte de renda no mesmo período. Para a fonte de renda associada aos indivíduos com ensino superior, todavia, apenas um estado entre os cinco mais pobres apresentou elevação de sua participação no total no mesmo período, número que vai a dois quando se consideradas as cinco unidades federativas mais ricas.

GRÁFICO 2

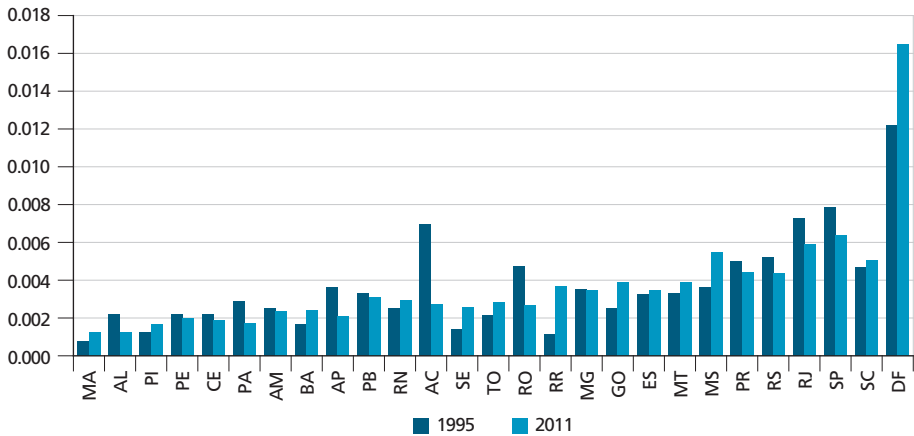
Parcela de renda associada aos indivíduos com ensino médio completo ou superior incompleto – Unidade da Federação



Fonte: elaboração própria, com base nos dados da Pnad/IBGE. Estados ordenados de forma crescente de acordo com a renda *per capita* de 1995.

GRÁFICO 3

Parcela de renda associada aos indivíduos com Superior Completo – Unidade da Federação



Fonte: elaboração própria, com base nos dados da Pnad/IBGE. Estados ordenados de forma crescente de acordo com a renda per capita de 1995.

Com o objetivo de estudar a influência dos movimentos de renda associados a diferentes níveis de escolaridade sobre a desigualdade regional de renda no Brasil entre 1995 e 2011, neste trabalho, são considerados dois grandes grupos de fonte de renda: a renda do trabalho e a renda de fontes não associadas ao trabalho. No sentido de obter uma associação entre a renda gerada e o nível de escolaridade, para o primeiro grupo, considera-se ainda a seguinte subdivisão da renda do trabalho:

- renda do trabalho das pessoas analfabetas ou com fundamental incompleto (até terceira série), denotada por Escol1;
- renda do trabalho das pessoas com fundamental incompleto (até sétima série), denotada por Escol2;
- renda do trabalho das pessoas com ensino fundamental completo e ensino médio incompleto, denotada por Escol3;
- renda do trabalho das pessoas com ensino médio completo e ensino superior incompleto, denotada por Escol4; e
- renda do trabalho das pessoas com ensino superior completo, denotada por Escol5.

Deve-se reconhecer que tais fontes de renda do trabalho associadas à escolaridade refletem movimentos tanto dos condicionantes da oferta, como de condicionantes da demanda de trabalho. Mais especificamente, os níveis e mudanças nessas fontes de renda refletem, ao mesmo tempo, tanto os níveis e elevação da escolaridade da força de trabalho local, como potenciais mudanças na estrutura produtiva que

influenciem os níveis de utilização da força de trabalho das unidades da federação de acordo com a escolaridade. Assim, por exemplo, a elevação da renda do trabalho associada aos níveis de escolaridade do determinado estado brasileiro pode refletir tanto o maior investimento educacional, como uma mudança de demanda associada a um processo de maior industrialização.

Com relação ao segundo (rendas não associadas ao trabalho), há também cinco fontes diferentes de renda:

- renda de pensões e aposentadorias, denotada por APP;
- renda do Programa Bolsa Família, denotada por BF;
- renda do Benefício de Prestação Continuada, denotada por BPC;
- renda do capital e outras fontes, denotada por “Capital”; e
- renda de aluguéis, abonos e doações, denotada por “Aluguéis”.

Embora não seja objetivo principal deste trabalho a análise específica das contribuições das fontes de renda não associadas ao trabalho sobre a desigualdade regional de renda no Brasil, do ponto de vista da política pública, tais contribuições são informativas e úteis, uma vez que permitem dimensionar os movimentos de renda favoráveis ou desfavoráveis à concentração regional associados aos diferentes níveis de escolaridade.

Com respeito a essas fontes de renda, é importante ressaltar que, enquanto para os grupos de renda do trabalho associados à escolaridade, as informações são prontamente disponíveis para processamento com base nos microdados da Pnad, para as fontes de renda dos programas sociais (BF e BPC), do grupo da renda não associada ao trabalho, tais informações não são imediatamente disponíveis nos microdados da Pnad, tendo que ser, portanto, imputadas com base em outras informações. Nesse sentido, nesta pesquisa, tanto para o Programa Bolsa Família, como para o Benefício de Prestação Continuada seguiu-se o expediente utilizado por Souza (2013) e por Soares *et al.* (2010). Especificamente, para o BPC, computou-se como renda deste programa a renda dos indivíduos com valor igual ao salário mínimo na variável V1273 (“capital e outras rendas”) na Pnad. Para o BF, da renda domiciliar obtida a partir da referida variável da Pnad, depois de descontada a renda do BPC, assumiu-se como renda do BF os valores até o limite da transferência máxima do Programa Bolsa Família (R\$ 306,00, em 2011, e R\$ 0,00, em 1995)³. As demais três fontes de renda não associadas ao trabalho são obtidas imediatamente nos microdados da Pnad.

3. Para uma discussão das diferentes maneiras de imputação das rendas dos programas sociais, a partir das informações da Pnad, é feita Soares *et al.* (2010).

Na tabela 1, a seguir, são apresentadas algumas informações obtidas a partir da agregação das variáveis individuais de renda por estado. Os números refletem a distribuição de cada fonte de renda entre as macrorregiões (a partir das linhas na horizontal) e o peso de cada fonte de renda na renda total, na última coluna à direita. Entre as muitas evidências de interesse, é importante destacar diferenças com respeito às duas macrorregiões mais populosas do país, situadas entre os extremos em termos de renda (Nordeste e Sudeste do país).

Nesse sentido, percebe-se que, para região Nordeste, as únicas fontes de renda em que a região detém participações próximas ou acima daquela observada quando se considera a população (27,8%), é aquela do grupo de escolaridade associado ao trabalho dos menos escolarizados (Escol1, com 27,4%) e as fontes de renda dos programas sociais (51,6% e 35,5%, respectivamente, para o BF e para o BPC). Por outro lado, para a região Sudeste, a maior participação da região na renda total quando comparada a sua participação na população ocorre nas fontes de renda do trabalho mais escolarizado, com destaque para a fonte de renda do trabalho associado ao ensino superior (54,5% do total), para a renda do capital e, por fim, a renda de aluguéis e previdência e aposentadoria.

TABELA 1
Participações regionais segundo cada fonte de renda – 2011
Em (%)

	Regiões					% da renda total
	Norte	Nordeste	Sudeste	Sul	Centro-Oeste	
População	8,5	27,8	42,0	14,3	7,5	100,0
Renda do trabalho						
Escol1	11,6	27,4	36,7	14,0	10,2	5,6
Escol2	6,6	15,9	46,5	21,5	9,4	9,6
Escol3	6,5	15,1	48,2	20,2	10,0	9,6
Escol4	6,2	15,9	51,1	17,7	9,1	27,8
Escol5	4,4	13,8	54,5	15,7	11,6	24,6
Outras rendas						
BF	13,0	51,6	22,0	7,5	5,9	0,9
BPC	11,5	35,5	32,8	11,6	8,6	0,6
Capital	6,2	16,1	49,6	22,0	6,0	0,5
Aluguéis	7,0	16,0	46,5	19,6	10,9	1,3
Aposentadorias e pensões	4,1	20,8	50,4	17,7	7,0	19,4
Renda Total	5,8	17,3	49,8	17,5	9,5	100,0

Fonte: elaboração própria com base nos microdados da Pnad/IBGE.

Nota-se, também, que o peso da renda dos menos escolarizados (Escol1) no Nordeste (27,4%) era bastante similar ao percentual desta região na população brasileira. Por outro lado, o peso desta região na fonte de renda dos mais escolarizados (Escol5) era de 13,8%. Tal situação contrasta também com aquela da região Sul: com apenas 7,5% da população total, era gerada na referida região, 11,7% da renda total dos mais escolarizados, com ensino superior (Escol5). Note-se que tais evidências são consistentes com a importância da educação para explicar as desigualdades regionais de renda no Brasil apontadas por Duarte, Ferreira e Salvato (2003) e Silveira Neto e Menezes (2008).

Na tabela 2, pode ser observada a participação das fontes de renda na renda total gerada em cada região, ou seja, como era distribuída a renda dentro de cada região do país no ano de 2011. Com base nessa tabela, pode-se notar, por exemplo, que enquanto 9,2% da renda gerada na região Nordeste derivou da renda das pessoas analfabetas ou que frequentaram até a terceira série do ensino fundamental, no Sudeste essa fonte de renda representava apenas 4,9% do total. Por outro lado, enquanto 24,5% da renda total gerada na região Nordeste era derivada da renda das pessoas com ensino médio completo e superior incompleto, na região Sudeste essa fonte de renda representou 27,7% do total da renda gerada na região.

TABELA 2
Participação de cada fonte de renda na renda total – macrorregiões do Brasil – 2011
Em (%)

	Regiões				
	Norte	Nordeste	Sudeste	Sul	Centro-Oeste
Renda do trabalho					
Escol1	9,8	9,2	4,9	4,5	6,3
Escol2	10,1	9,2	10,1	11,6	10,0
Escol3	9,8	8,3	9,4	11,2	10,4
Escol4	31,9	24,5	27,7	28,3	26,8
Escol5	19,9	19,3	24,5	22,2	29,3
Outras Rendas					
BF	1,9	2,8	0,5	0,4	0,6
BPC	1,0	1,2	0,4	0,4	0,6
Capital	0,7	0,4	0,6	0,6	0,3
Aluguéis	1,5	1,2	1,5	1,5	1,5
Aposentadorias e pensões	13,4	23,9	20,5	19,5	14,1
Total	100,0	100,0	100,0	100,0	100,0

Fonte: Elaboração própria com base nos microdados da Pnad/IBGE.

Três fontes de renda, além disso, merecem destaque. A primeira é a fonte de renda das pessoas mais escolarizadas. Enquanto na região Nordeste o peso desta fonte de renda era de 19,3%, na região Sul e na região Sudeste, o peso desta fonte de renda foi de 22,2% e 24,5%, respectivamente. A segunda é a renda de Aposentadorias e Pensões, a qual apresentou resultados bastante díspares para as diferentes regiões. Enquanto a participação desta fonte de renda na renda total da região Norte era de 13,4%, na região Nordeste e Sudeste, a participação foi de 23,9% e 20,5%, respectivamente. Por fim, é claramente evidenciada a maior importância das fontes de renda associadas aos programas sociais nas duas regiões mais pobres do país (Nordeste e Norte). No Nordeste, em particular, quase 3% da renda total dos domicílios está associada aos recursos do Programa Bolsa Família, um percentual que só chegava a 0,4% na região Sul do país.

3 A IMPORTÂNCIA DAS DIFERENTES FONTES PARA EVOLUÇÃO DA DESIGUALDADE REGIONAL DE RENDA NO BRASIL ENTRE 1995 E 2011

A estratégia a ser utilizada neste estudo para determinar a importância das diferentes fontes de renda na explicação dos níveis de desigualdade regional de renda *per capita* e sua evolução tem como pedra fundamental o cálculo do coeficiente de concentração, que deriva da curva de Lorenz generalizada, proposta por Fei, Ranis e Kuo (1979) e Kakwani (1980). O cálculo deste coeficiente é bastante similar ao cálculo da curva de Lorenz convencional⁴.

Nesse sentido, seja o orçamento domiciliar *per capita* da *i*-ésima pessoa, com $i = 1, \dots, n$. Ordenando-se esses rendimentos em $x_1 < x_2 < \dots < x_n$, é possível construir a curva de Lorenz, a qual mostra como a proporção acumulada da renda varia em função da proporção acumulada da população. Definindo β como a área entre a curva de Lorenz e o eixo das abscissas, sabemos que o índice de Gini (G) é dado por:

$$G = 1 - 2\beta, \text{ em que } 0 < G < 1. \quad (1)$$

Assumindo-se, agora, que x_i é composta de k parcelas:

$$x_i = \sum_{h=1}^k x_{hi}. \quad (2)$$

Em que cada x_{hi} será uma parcela da renda do trabalho associada à escolaridade ou uma parcela da renda não relacionada ao trabalho, as quais foram apresentadas na seção anterior.

4. Ver a respeito, por exemplo, Hoffman (2012).

Mantida a ordenação da população pela renda *per capita* total tal como na curva de Lorenz, podem-se construir as curvas de concentração de cada fonte de renda particular. A curva de concentração da parcela x_{hi} mostra como a proporção acumulada de cada x_{hi} varia em função da proporção acumulada da população.

Definindo β_h como a área entre essa curva e o eixo das abscissas, é possível calcular o coeficiente de concentração da fonte de renda h como:

$$C_h = 1 - 2\beta, \text{ em que } -1 < C_h < 1. \quad (3)$$

A partir da Equação (3) e definindo φ_h como a participação da h -ésima parcela na renda total, Kakwani (1980) demonstra que o índice de Gini pode ser calculado como:

$$G = \sum_{h=1}^k \varphi_h C_h. \quad (4)$$

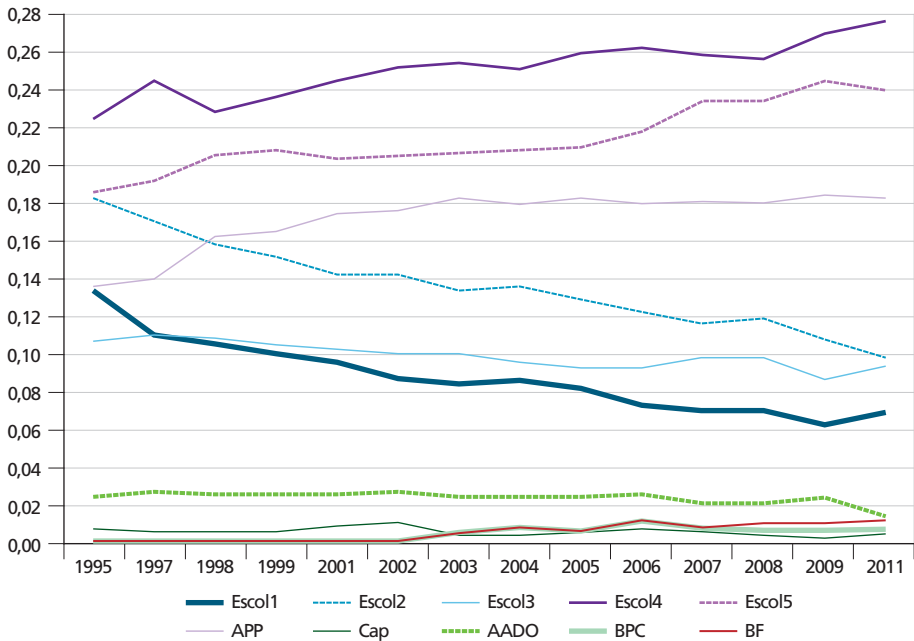
Ou seja, o índice de Gini pode ser decomposto em k parcelas do rendimento, sendo obtido a partir da soma ponderada dos coeficientes de concentração, em que os pesos correspondem às participações das respectivas fontes de renda.

No gráfico 2, pode ser observada a evolução da participação de cada fonte de renda na renda *per capita* total ao longo do tempo⁵. Nota-se que a renda do trabalho das pessoas com menor nível de escolaridade (Escol1, Escol2, Escol3) tem representado uma parcela cada vez menor da renda total. Em 1995, as fontes de renda Escol1 e Escol2 representavam, respectivamente, 13,3% e 18,2% da renda total. Em 2011, o percentual de cada uma destas fontes de renda foi reduzido, respectivamente, para 6,9% e 9,7% da renda total.

Por outro lado, a parcela da renda do trabalho dos indivíduos com maior nível de escolaridade tem tido cada vez mais importância como parcela da renda *per capita* total. Enquanto em 1995 a fonte de renda Escol4 e a fonte de renda Escol5 representavam 22,5% e 18,5% da renda total, em 2011, os pesos destas fontes de renda aumentaram para 27,6% e 23,9%, respectivamente.

5. Nas colunas (1) e (2) da tabela 3, podem ser visualizados os valores exatos da participação das fontes de renda nos anos 1995 e 2011.

GRÁFICO 4
Evolução da participação das fontes de renda na renda total



Fonte: elaboração própria com base nos dados da Pnad/IBGE.

Nas colunas (5) e (6) da tabela 3, apresentada a seguir, podem ser observados os valores das contribuições de cada fonte de renda para o cálculo do índice de Gini em 1995 e em 2011. Os resultados indicam que o conjunto das fontes de renda do trabalho era responsável por 86,5% do índice de Gini calculado em 1995 e por 84,3% em 2011. Por sua vez, as rendas dos indivíduos mais escolarizados (Escol4 e Escol5) eram responsáveis por 52,3% do índice de Gini calculado em 1995 e por 66,4% em 2011.

É possível estender a análise e obter as contribuições dessas diferentes fontes de renda para a redução da desigualdade de renda observada entre estes dois anos a partir de diferentes decomposições do indicador de desigualdade de renda utilizado neste trabalho (índice de Gini). Mais especificamente, a partir do cálculo do coeficiente de concentração é possível obter a variação do índice de Gini entre dois períodos no tempo (Hoffmann, 2006) e, para cada fonte de renda, determinar sua contribuição para a variação observada no índice de Gini.

Para tal, de início, é necessário retomar a equação (4) e expressar a variação da desigualdade entre dois pontos no tempo como:

$$\Delta G = G_2 - G_1 = \sum_{h=1}^k (\varphi_{2h} C_{2h} - \varphi_{1h} C_{1h}). \quad (5)$$

em que o “ Δ ” indica variação entre os dois períodos. A partir da soma e subtração do produto ou , depois de uma fatoração do resultado, é possível obter:

$$\Delta G = \sum_{h=1}^k (C_{2h} \Delta \varphi - \varphi_{1h} \Delta C_h) \quad \text{ou} \quad \Delta G = \sum_{h=1}^k (C_{1h} \Delta \varphi - \varphi_{2h} \Delta C_h). \quad (6)$$

Para se evitar a escolha entre uma dessas formas de decompor G, utiliza-se a média aritmética das duas:

$$\Delta G = \sum_{h=1}^k (C_h^* \Delta \varphi - \varphi_h^* \Delta C_h), \quad (7)$$

em que $C_h^* = (C_{2h} + C_{1h})$ e $\varphi_h^* = (\varphi_{2h} + \varphi_{1h})/2$. Note-se que, na Equação (7), tem-se a variação do índice de Gini entre dois períodos obtida pela variação nas participações ($C_h^* \Delta \varphi$) e pela variação no coeficiente de concentração ($\varphi_h^* \Delta C_h$). Finalmente, como $\sum \varphi_{2h} - \sum \varphi_{1h} = 0$, é possível fazer:

$$\Delta G = \sum_{h=1}^k [(C_h^* - G^*) \Delta \varphi + \varphi_h^* \Delta C_h], \quad (8)$$

em que $G^* = (G_2 + G_1)/2$.

A partir desta última equação, é possível definir o primeiro termo do lado direito como o *efeito participação*, o qual reflete o aumento ou a diminuição da importância (peso) de cada fonte de renda durante o período. Percebe-se que um aumento (diminuição) da participação de uma fonte de renda com concentração acima da desigualdade observada para a renda total causará um aumento (diminuição) da concentração resultante. Por sua vez, o segundo termo do lado direito representa o *efeito concentração*, o qual indica quanto o índice Gini total muda em razão de uma mudança no padrão de concentração de uma fonte de renda. Na tabela 3, a seguir, são apresentados os resultados da decomposição do índice de Gini entre o *efeito concentração* (EC) e o *efeito participação* (EP), bem como o *efeito total* (ET) de cada fonte de renda sobre a variação da desigualdade no período.

Percebe-se, primeiramente, que, da redução de cerca de 6,5% do índice de Gini para desigualdade de renda *per capita* entre os estados brasileiros no período de 1995 a 2011 (o índice vai de 0,201 a 0,197, no período), o *efeito concentração* foi responsável por uma queda bem maior que aquela observada de fato para o índice (178,9% da variação do índice de Gini regional no período). Por outro lado, o *efeito participação* contribuiu para um movimento de forma contrária: as mudanças nas participações das diferentes fontes de renda entre os dois anos implicaram

um movimento de aumento do índice de Gini correspondente a cerca de 78,9% do valor da variação observada para o índice. Esse resultado, semelhante àquele obtido por Silveira Neto e Azzoni (2012), indica que a redução da desigualdade regional de renda no Brasil tem sido conduzida pelo modo como as distribuições das fontes de renda têm evoluído e não devido a alterações na participação de cada fonte na renda total.

Na tabela 3, verifica-se também que a renda do trabalho e a renda dos programas sociais são as principais fontes de renda que contribuem para a redução do índice de Gini regional no período. Por outro lado, a fonte de renda das aposentadorias e pensões e aquela dos aluguéis e de outras fontes de renda apresentaram movimentos no sentido de elevar a desigualdade regional de renda. Mais especificamente, como um todo, a renda do trabalho foi responsável por um movimento de redução do índice de Gini de cerca de 47% da variação observada entre os anos de 1995 e 2011.

Entre as fontes de renda não vinculadas ao trabalho, destacam-se as fontes de renda associadas aos programas sociais de transferência de renda. Apesar de representarem menos de 2% da renda total em 2011 e serem inteiramente focadas nos indivíduos e não no território (*spatial blind*), as fontes de renda do BPC e, principalmente, do BF contribuíram para um movimento de redução do índice de desigualdade equivalente a 63,9% da variação final observada para este índice. Na verdade, das fontes de renda não vinculadas ao trabalho, a renda do programa BF é aquela que apresenta o mais significativo movimento para variação (no caso, redução) da desigualdade regional: um movimento de cerca de 45% da redução observada no índice de Gini. No sentido oposto, a dinâmica de renda das aposentadorias e pensões indica um movimento de maior concentração regional desta fonte: tal dinâmica implicou movimento de elevação do Gini em 14,8% da variação observada.

Entre as fontes de renda associadas ao trabalho, é interessante observar que, entre as cinco categorias de escolaridade, são exatamente para aquelas de renda do trabalho dos menos escolarizados (Escol1) e dos mais escolarizados (Escol5) que se observam movimentos de renda no sentido de elevação da desigualdade regional de renda, ou seja, movimentos que impedem maior redução do índice de Gini entre 1995 e 2011. Note-se, em particular, que o movimento da renda do trabalho do grupo com ensino superior completo impede uma maior redução do índice de Gini, tanto porque tal fonte de renda fica mais concentrada nos estados mais ricos, quanto porque aumenta sua participação na renda *per capita* total, o que se reflete nos *efeitos concentração e participação* desta fonte de renda. Por seu turno, a dinâmica da fonte de renda daqueles que têm apenas o ensino fundamental incompleto impede uma maior queda do índice de Gini pelo fato desta fonte de renda apresentar redução em sua participação no período e apresentar nível

de distribuição pró estados mais pobres (baixo valor para coeficiente de concentração). Em sentido oposto, percebe-se que, isoladamente, a fonte de renda que mais contribui para a redução da desigualdade regional brasileira, entre 1995 e 2010, é a fonte de renda dos trabalhadores com ensino médio completo e ensino superior incompleto (Escol4) e, basicamente, porque tal fonte de renda torna-se regionalmente menos concentrada: tal fonte é responsável por um movimento de redução da desigualdade de 84,9% da redução do índice de Gini, observada entre 1995 e 2011.

Esses resultados indicam que, se por um lado, a redução dos desníveis de escolaridade entre os estados brasileiros contribuiu de forma decisiva, por meio das ocupações e renda associadas aos níveis intermediários de escolaridade, principalmente aquele do ensino médio, por outro, houve no período um movimento no sentido de maior concentração regional de renda associado ao nível mais alto de escolaridade (ensino superior). Este último movimento, por sua vez, é consistente com a elevação da concentração regional de atividades principalmente ligadas ao setor de serviços especializados, como por exemplo, setor financeiro e aqueles ligados à tecnologia da informação.

TABELA 3

Participações, coeficientes de concentração e contribuições das fontes de renda para níveis e variação da desigualdade regional de renda no Brasil entre 1995 e 2011 – Desigualdade medida pelo índice de Gini (Em %)

Fontes de Renda	Participação		Coeficientes de concentração		Contribuição para a desigualdade		Decomposição (1995-2011)		
	1995 (1)	2011 (2)	1995 (3)	2011 (4)	1995 (5)	2011 (6)	EP (7)	EC (8)	ET (9)
Trabalho									
Escol1	13,36	6,93	0,025	0,003	1,60	0,10	-97,79	18,11	-79,68
Escol2	18,23	9,75	0,234	0,171	20,32	8,47	-0,60	70,12	69,52
Escol3	10,69	9,42	0,260	0,195	13,25	9,30	2,45	52,50	54,95
Escol4	22,48	27,65	0,245	0,199	26,23	27,87	-7,62	92,51	84,89
Escol5	18,56	23,96	0,295	0,317	26,09	38,54	-44,51	-38,34	-82,85
Outras									
Aposentadorias e pensões	13,59	18,35	0,141	0,167	9,15	15,57	18,79	-33,64	-14,85
BPC	-	0,73	-	-0,118	-	-0,43	15,32	3,44	18,76
Bolsa Família	-	1,30	-	-0,230	-	-1,51	33,18	11,99	45,17
Capital	0,70	0,50	0,256	0,166	0,85	0,42	0,12	4,35	4,47
Aluguéis e outras	2,40	1,42	0,219	0,232	2,50	1,67	1,72	-2,10	-0,38
Total	100,00	100,00	0,210	0,197	100,00	100,00	-78,92	178,92	100,00

Fonte: elaboração própria com base nos microdados da Pnad/IBGE.

4 SENSIBILIDADE DA DESIGUALDADE REGIONAL DE RENDA A VARIAÇÕES NAS FONTES DE RENDA

As evidências apresentadas nas seções anteriores permitem apreender o papel das diferentes fontes de renda associadas aos diferentes níveis de escolaridade na redução da desigualdade regional de renda, medida pelo índice de Gini, entre os anos de 1995 a 2011. Dadas as estruturas produtivas dos estados nesses dois anos, é possível também obter medidas da influência de variações marginais em cada uma das diferentes fontes de renda sobre o nível de desigualdade regional de renda. Isso permite não só conhecer, para o ano mais recente (2011), quais as perspectivas de reduções adicionais da desigualdade com base em variações em cada fonte de renda, como comparar a eficiência relativa dessas diferentes fontes com respeito à redução da desigualdade. O objetivo desta seção é, assim, obter medidas da sensibilidade da desigualdade regional de renda entre as unidades federativas do país em relação a variações em cada fonte de renda associada aos diferentes níveis de escolaridade. Isto é, obtêm-se estimativas da elasticidade do índice de Gini em relação a cada fonte de renda. Num esforço complementar, aplicando-se uma decomposição de Shapley, discutem-se os papéis de diferentes fatores na variação das elasticidades de cada fonte de renda entre os anos de 1995 a 2011.

Ao utilizar uma decomposição alternativa do índice de Gini, é possível estudar diretamente a influência dos níveis de desigualdade de cada fonte de renda, apreendidos por respectivos índices de Gini, sobre a evolução da desigualdade regional de renda total. Nesse sentido, seguindo Leman e Yitzhaki (1985), o primeiro passo é escrever o índice de Gini na forma:

$$G = \sum_{f=1}^N R_f G_f \alpha_f, \quad (9)$$

em que $R_f = Cov(y_f, F)/Cov(y_f, F_f)$, G_f é o índice de Gini da fonte de renda f e α_f corresponde, como antes, à participação dessa fonte de renda na renda total. O termo R_f que corresponde à razão entre a covariância entre a renda da fonte f e a função de distribuição acumulada total, denotada por F , e a covariância entre a renda da fonte f e sua função de distribuição acumulada, F_f expressa uma medida de correlação entre as distribuições da fonte específica f de renda e da renda total⁶.

Como bem colocam Silveira Neto e Azzoni (2012), na expressão acima, o índice de Gini torna-se uma agregação de índices de Gini individuais para cada fonte de renda, ponderados pela participação na renda total e pelo valor de R_f . Dessa forma, um determinado nível de desigualdade na distribuição de uma fonte de renda f (determinado valor de G_f) terá tanto mais influência no nível de

6. Como se pode perceber, a partir da comparação entre as equações (9) e (4), $R_f = C_f/G_f$, ou seja, o fator R_f corresponde também à razão entre o coeficiente de concentração da fonte f de renda e o índice de Gini de sua distribuição.

desigualdade final quanto maiores forem os valores de sua participação, α_f e da correlação entre a distribuição dessa fonte de renda e a distribuição da renda total, R_f . Note-se, também, que valores positivos (negativos) de R_f indicam que a fonte de renda f contribui para um aumento (redução) do coeficiente de Gini da distribuição da renda total e que, assim, elevações (diminuições) da desigualdade na distribuição da fonte de renda (aumentos (diminuições) de G_f) implicam elevações da desigualdade na distribuição da renda total. Valores extremos de R_f (próximos de um) indicam, além disso, que a desigualdade na distribuição da fonte de renda f (captada pelo valor de G_f) aproxima-se da desigualdade na distribuição da renda total.

Como mostraram, pioneiramente, Leman e Yitzahki (1985), partindo-se da equação (9), é possível obter a elasticidade do índice de Gini para a distribuição da renda total em relação a variações nas fontes de renda da seguinte forma:

$$\varepsilon_{G_yf} = \frac{(\alpha_f G_f R_f)}{G} - \alpha_f. \quad (10)$$

E, como $R_f = \frac{C_f}{G_f}$,

$$\varepsilon_{G_yf} = (\alpha_f C_f G^{-1}) - \alpha_f, \quad (11)$$

em que ε_{G_yf} corresponde à elasticidade do índice de Gini da distribuição da renda total em relação à fonte de renda específica f .

A expressão obtida por estes autores mostra que a elasticidade da desigualdade em relação a cada fonte de renda depende positivamente do quanto uma fonte de renda está distribuída pró estado rico ou pobre (ou seja, do coeficiente de concentração, C_f), da participação desta fonte na renda total e, negativamente, do próprio nível de desigualdade total (G). Assim, por exemplo, fontes de renda com valores negativos para o coeficiente de concentração sempre terão valores negativos para a elasticidade: aumentos específicos dessas fontes de renda sempre diminuem a desigualdade regional, influências que serão maiores quanto menor for a desigualdade da distribuição da renda total. Finalmente, o efeito da participação na renda total, α_f sobre a desigualdade está na dependência do valor do coeficiente de concentração, C_f em relação ao índice de Gini, G . Nesse sentido, para uma fonte de renda altamente concentrada nos estados ricos, por exemplo, a relação entre seus níveis de participação e o valor da elasticidade será positiva.

Nas colunas (1), (2) e (3) da tabela 4, apresentada a seguir, são registradas as elasticidades da desigualdade da distribuição regional da renda total em relação a

cada fonte de renda para os anos de 1995 a 2011, bem como as variações dessas elasticidades nesse período.

Entre as fontes de renda do trabalho, há um claro destaque para a fonte de renda das pessoas com nível superior completo (Escol5). Tal fonte de renda apresenta os maiores valores da elasticidade nos dois anos analisados, bem como maior variação da elasticidade no período. Mais especificamente, o valor obtido para elasticidade dessa fonte de renda em 2011 indica que um aumento de 1% da renda das pessoas com nível superior completo estava associada a um aumento do índice de Gini da distribuição da renda total de cerca de 14,6%. Além disso, a elevação da elasticidade dessa fonte de renda entre 1995 e 2011 (em mais de 100%) indica que, em 2011, a desigualdade regional brasileira passou a ser ainda mais sensível à variação da renda dos mais escolarizados que em 1995.

Um reflexo da expansão dos níveis de escolaridade dos estados mais pobres é interessante também perceber que, em 2011, as únicas fontes de renda do trabalho que apresentavam valores positivos para a elasticidade da desigualdade da distribuição regional de renda entre os estados brasileiros eram as duas fontes correspondentes às rendas dos indivíduos mais escolarizados, Escol4 (indivíduos com ensino médio ou superior incompleto) e a própria fonte Escol5 (indivíduos com superior completo). Ou seja, diferentemente de 1995 quando os níveis de escolaridade intermediários também impactavam positivamente a desigualdade, no ano mais recente, a influência dos níveis de escolaridade sobre a desigualdade regional tende a atuar a partir dos níveis mais altos de escolaridade.

Entre as fontes de renda do trabalho, a única fonte de renda com valores negativos para a elasticidade da desigualdade da distribuição da renda total, tanto em 1995 quanto em 2011, é aquela dos indivíduos que completaram, no máximo, a terceira série do ensino fundamental (Escol1). Isso significa que, a partir da renda do trabalho, somente para tal grupo de indivíduos é que as elevações de renda estavam nos dois anos associadas a diminuições da desigualdade regional de renda no Brasil. Note-se, contudo, que apesar de tal efeito contribuir de modo favorável à redução da desigualdade, sua magnitude é bem menor em 2011 que em 1995: enquanto em 1995 o aumento de 1% desta fonte de renda estava associado a uma redução da desigualdade regional de renda, medida pelo índice de Gini, em torno de 11,8%, em 2011, o mesmo aumento percentual indicava uma redução de 6,8% do valor do mesmo indicador de desigualdade.

Tais resultados indicam que a expansão da renda do trabalho está longe de ser neutra, em relação à distribuição regional da renda no Brasil. Mesmo com o maior progresso escolar relativo dos estados mais pobres observado no período mais recente, os estados mais ricos ainda tendem a se beneficiar relativamente mais com a expansão da renda em setores econômicos que utilizam predominantemente

força de trabalho mais escolarizada e, com isso, os ganhos relativos das unidades mais ricas têm aumentado ao longo do tempo.

No que diz respeito aos valores das elasticidades para as fontes de renda não associadas ao trabalho, os destaques são as fontes associadas às rendas das aposentadorias e pensões e à renda do Programa Bolsa Família (BF), que, por sinal, em 2011, apresentavam valores negativos de mesma magnitude para a elasticidade da desigualdade da distribuição de renda (-2,8%). Aqui, é notável que a fonte de renda associada ao BF, com apenas 1,3% de participação na renda total, em 2011, apresente o terceiro maior valor (absoluto) para a elasticidade entre todas as fontes de renda consideradas neste trabalho.

Da equação (11), é possível perceber também que as variações das elasticidades da desigualdade regional na distribuição da renda total associadas a cada renda estão associadas a variações no próprio nível de desigualdade (por meio do índice de Gini, G), a variações no coeficiente de concentração (C_f) e a variações nas participações das fontes de renda (α_f). Com o objetivo de evidenciar a contribuição de cada um desses potenciais componentes, são apresentados os resultados de uma decomposição de Shapley para a variação da elasticidade de cada fonte de renda, entre 1995 e 2011, a partir da equação (13). Formalmente, isso corresponde a considerar a variação entre 1995 ($t = 0$) e 2011 ($t = t$) de cada elasticidade na forma:

$$\varepsilon_{G_y f t} - \varepsilon_{G_y f 0} = \left(\alpha_{f t} C_{f t} G_t^{-1} \right) - \alpha_{f t} - \left(\alpha_{f 0} C_{f 0} G_{t 0}^{-1} \right) - \alpha_{f 0} \quad (12)$$

A obtenção de medidas da contribuição de cada um dos três fatores expressos na equação (11) para a variação da desigualdade regional brasileira entre 1995 e 2011 pode ser obtida com base na decomposição de Shapley (Shorrocks, 1999; Shorrocks, 2012). O valor de Shapley é um conceito desenvolvido no âmbito da teoria dos jogos cooperativos e sua aplicação para a decomposição de índices de pobreza e desigualdade foi estabelecida por Shorrocks (1999)⁷. Nesse sentido, a variação da elasticidade de cada fonte de renda entre 1995 e 2011 pode ser calculada por meio da seguinte equação:

$$\Delta \varepsilon_{G_y f} = C(\Delta C_f) + C(\Delta G) + C(\Delta \alpha_f) \quad (13)$$

7. Para mais detalhes sobre a decomposição de Shapley, ver Deutsch e Silber (2005), Shorrocks (2012) e Sastre e Trannoy (2000).

em que a contribuição da variação do coeficiente de concentração de cada fonte de renda para a variação na elasticidade global $C(\Delta C_f)$ pode ser calculada como⁸:

$$C(\Delta C_f) = \frac{2!}{3!} [C_f^t(G^0)^{-1} \alpha_f^0 - C_f^t(G^0)^{-1} \alpha_f^t] + \frac{1!}{3!} [C_f^t(G^t)^{-1} \alpha_f^0 - C_f^t(G^t)^{-1} \alpha_f^t] + \frac{1!}{3!} [C_f^t(G^0)^{-1} \alpha_f^t - C_f^t(G^0)^{-1} \alpha_f^t] + \frac{2!}{3!} [C_f^t(G^t)^{-1} \alpha_f^t - C_f^t(G^t)^{-1} \alpha_f^t] \quad (14)$$

Decomposições similares podem ser realizadas para $(G)^{-1}$ e α_f . Portanto, $C(\Delta C_f)$ representa o efeito da variação do coeficiente de concentração da fonte f de renda sobre a variação da elasticidade desta fonte de renda, $C(\Delta G)$ representa o efeito da variação da desigualdade sobre a variação da elasticidade de cada fonte de renda e representa o efeito da variação da participação de cada fonte de renda f sobre a variação da elasticidade.

Embora permita apreender a contribuição dos diferentes fatores associados às mudanças de determinada variável (no caso, a elasticidade do índice de Gini em relação a variações em diferentes fontes de renda), é importante reconhecer que a decomposição de Shapley apresenta algumas limitações. Além, evidentemente, de não permitir interpretações das contribuições como efeitos causais. Uma limitação importante é que tal decomposição não satisfaz o princípio da independência, o que significa que a contribuição de cada fonte de renda sobre a desigualdade depende de como são agregadas as demais fontes (Sastre e Trannoy, 2002; Cowell e Fiorio, 2011). Adicionalmente, como argumentam Chantreuil e Trannoy (2013), quando usado para mensurar, por exemplo, a contribuição das fontes de renda para o nível geral de desigualdade, a decomposição de Shapley gera valores de contribuições diferentes, a depender do índice de desigualdade utilizado⁹. Tais observações, contudo, não invalidam os resultados, aqui, que devem ser considerados condicionados à estrutura de fontes de renda assumida.

Os resultados da aplicação da decomposição de Shapley para a variação da elasticidade da desigualdade da distribuição de renda entre os estados brasileiros em relação a cada fonte de renda são apresentados nas colunas (4), (5) e (6) da tabela 4.

8. Também aqui, a decomposição envolve a soma para fonte de renda de todas as possíveis combinações (alterando-se a ordem de cada elemento) de variações de cada um dos três componentes C , G e α ponderadas pela frequência de ocorrência de cada valor da parcela. Como novamente existem três fatores, o número de todas as possíveis combinações será $3! = 6$. Para a primeira parcela, por exemplo, existem apenas duas possibilidades de combinações dos elementos que envolvem mesmo valor do produto: $C_f^t(G^0)^{-1} \alpha_f^t - C_f^t(G^0)^{-1} \alpha_f^0 = C_f^t \alpha_f^t (G^0)^{-1} - C_f^t \alpha_f^0 (G^0)^{-1}$, o que explica o numerador do termo de ponderação igual a $2 = 2!$.

9. Chantreuil e Trannoy (1999) e Shorrocks (1999) propõem extensões da decomposição de Shapley ao associar o problema ao princípio da independência e ao propor estruturas hierárquicas para as fontes de renda. Contudo, como argumentam Cowell e Fiorio (2011), tais soluções nem sempre têm apelo e interpretações econômicas claras. Estes últimos autores, na verdade, propõem uma compatibilização entre decomposições de desigualdade por meio de fontes ou grupos de renda, e aquelas mais recentemente obtidas por meio de regressões.

Entre as fontes de renda do trabalho, nas referidas colunas da tabela 4, nota-se, por exemplo, que 33,4% da elevação da elasticidade da desigualdade associada à fonte de renda dos indivíduos mais escolarizados entre 1995 e 2011 apontada acima decorre do fato desta fonte estar mais concentrada (efeito da variação do coeficiente de concentração, C_f), sendo o restante decorrente da própria redução do índice de Gini (27,8%) e da elevação de sua participação (38,8%). Por outro lado, o menor valor absoluto da elasticidade da desigualdade associada à fonte de renda dos indivíduos com menor nível de escolaridade (Escol1) decorre quase que exclusivamente do menor peso que esta fonte de renda passa a ter na renda total em 2011. Para as fontes de renda do trabalho associadas aos níveis intermediários de escolaridade, os movimentos de diminuição dos valores das elasticidades entre 1995 e 2011 decorrem, sobretudo, da diminuição do nível de concentração dessas fontes de renda (componente C_f).

Por fim, entre as fontes de renda não associadas ao trabalho, a mudança mais importante entre 1995 e 2011 diz respeito à diminuição da influência das rendas das aposentadorias e pensões sobre a desigualdade regional de renda entre os estados brasileiros. Tal movimento é, sobretudo, explicado pelo aumento do coeficiente de concentração (C_f) desta fonte de renda, ou seja, pelo fato de que a evolução dessa fonte particular de renda ter sido mais favorável aos estados mais ricos do país (ver valores desses coeficientes na tabela 3).

TABELA 4
Elasticidades da desigualdade da distribuição de renda entre os estados brasileiros e decomposição de Shapley para a variação das elasticidades

Fontes de Renda	Elasticidade		Variação da elasticidade	Contribuição (1995-2011) %		
	1995 (1)	2011 (2)	1995-2011 (3)	C_f (4)	G (5)	α (6)
Trabalho						
Escol1	-0,118	-0,068	0,049	-22,43	0,94	121,49
Escol2	0,021	-0,013	-0,034	126,92	-25,68	-1,24
Escol3	0,026	-0,001	-0,027	120,25	-25,83	5,58
Escol4	0,038	0,002	-0,035	160,27	-47,11	-13,15
Escol5	0,075	0,146	0,070	33,37	27,84	38,80
Outras						
Aposentadorias e pensões	-0,044	-0,028	0,017	123,95	44,727	-68,677
BPC	0,000	-0,012	-0,012	18,32	0,74	80,94
BF	0,000	-0,028	-0,028	26,40	1,07	72,54
Capital	0,002	-0,001	-0,002	113,53	-16,50	2,97
Aluguéis e outras	0,001	0,003	0,002	85,31	85,75	-71,06

Fonte: elaboração própria com base nos microdados da Pnad/IBGE.

5 CONSIDERAÇÕES FINAIS

O baixo nível de escolaridade da população e o alto nível de desigualdade de renda pessoal e regional são problemas persistentes no Brasil. No entanto, como mostraram diferentes pesquisadores, na última década, o Brasil apresentou uma redução consistente não só na desigualdade de renda pessoal, como também na desigualdade de renda regional. Há dois movimentos diretamente associados a estas reduções de desigualdade (pessoal e regional), ambos associados a políticas públicas, com foco nos indivíduos e não no território ou nas regiões: a elevação da escolaridade em todas as regiões do país e a introdução e expansão das políticas sociais de transferência de renda. Nesse contexto, o objetivo deste trabalho foi analisar a evolução da disparidade regional de renda *per capita* no Brasil, calculada pelo índice de Gini, verificando especificamente qual a contribuição de cinco fontes de renda do trabalho associadas a diferentes níveis de escolaridade sobre a evolução da desigualdade regional de renda *per capita* no Brasil entre os anos 1995 e 2011 e sua importância relativa em relação às demais fontes de renda.

Nesse sentido, primeiramente, o trabalho mostrou que, em relação à distribuição da renda total, por um lado, as fontes de renda do trabalho associadas a indivíduos com maiores níveis de escolaridade (no mínimo com ensino médio) encontravam-se em 2011 mais concentradas nas unidades federativas mais ricas e, por outro, que as fontes de renda do trabalho de indivíduos com menos escolaridade e as fontes associadas aos programas de transferência de renda (sobretudo) se apresentavam, no mesmo ano, mais favoráveis às unidades federativas mais pobres do país.

Com base no cálculo das contribuições das diferentes fontes de renda para a desigualdade, os principais resultados deste estudo indicam que a renda do trabalho explica a maior parte da desigualdade regional, calculada pelo índice de Gini, bem como grande parte do movimento da recente queda no índice de Gini regional nos últimos anos. Enquanto em 1995 o conjunto das fontes de renda relacionadas ao trabalho era responsável por cerca de 87% da desigualdade regional medida pelo índice de Gini, em 2011 tal responsabilidade era 84%, um reflexo do crescimento da importância de outras fontes de renda. Porém, esse é um valor que ainda torna clara a importância dessa fonte de renda para o entendimento das disparidades regionais no país.

Não menos importante, os resultados obtidos indicam que, em 2011, cerca de 65% da desigualdade regional de renda entre os estados brasileiros decorria dos níveis (elevados) de concentração da renda dos indivíduos que possuíam, no mínimo, o ensino médio completo, um percentual superior àquele observado em 1995 (em torno de 53%), o que decorria de uma distribuição de renda fortemente pró estado rico das fontes de renda associadas à maior escolaridade. Entre as demais fontes de renda, o maior destaque na explicação da desigualdade

regional diz respeito à distribuição das rendas das aposentadorias e pensões, que em 2011 contribuía com cerca de 16% da desigualdade regional brasileira. Ou seja, nos anos mais recentes, a maior parte da desigualdade regional brasileira decorre da distribuição regional bastante desigual da renda do trabalho dos indivíduos mais escolarizados, principalmente daqueles com ensino superior.

No que tange à evolução da desigualdade regional entre 1995 e 2011, as evidências obtidas indicam que as dinâmicas das rendas do trabalho de indivíduos com níveis intermediários de escolaridade (com fundamental completo até o ensino médio completo e superior incompleto) e das rendas dos programas sociais de transferência de renda (Bolsa Família e Benefício de Prestação Continuada) foram as responsáveis pela redução do índice de Gini da distribuição regional de renda no período no Brasil. Especificamente, em seu conjunto, as rendas do trabalho foram responsáveis por, aproximadamente, 47% da variação do índice de Gini regional entre os anos de 1995 e 2011. Um movimento que, seguramente, está associado à expansão da educação nos estados mais pobres do Brasil. É importante salientar, contudo, que, no mesmo período, a dinâmica de renda dos indivíduos mais escolarizados (ensino superior) não apenas ganhou importância ao longo do tempo na renda total dos estados, como apresentou dinâmica no sentido de uma significativa concentração regional de renda entre as unidades federativas do país: na verdade, a redução do índice de Gini observada poderia ter sido quase 83% maior, caso tal fonte de renda não apresentasse evolução em favor das unidades mais ricas.

A importância do movimento da renda do trabalho associado ao nível de escolaridade mais elevado (ensino superior) para a configuração da distribuição regional de renda apresenta-se, aliás, mais significativa que aquelas das fontes de renda dos programas sociais (Bolsa Família e Benefício de Prestação Continuada), que, confirmando resultado obtido por Silveira Neto e Azzoni (2012) explicam parte não desprezível da redução da desigualdade regional de renda no Brasil entre 1995 e 2011.

Ainda que a metodologia utilizada não permita estabelecer rigorosamente relações causais entre as diferentes fontes de renda e a dinâmica de desigualdade regional do conjunto de evidências apresentadas neste trabalho, há, ao menos, dois resultados úteis para a política pública que tenha como objetivo reduzir as disparidades regionais: Primeiro, é importante ter em conta que a redução das disparidades regionais de renda no Brasil é fortemente dependente da equalização dos níveis de escolaridade dos indivíduos entre suas unidades federativas em níveis superiores de escolaridade. Isso parece ainda mais importante quando se considera que o crescimento da renda dos estados mais pobres observados nos últimos anos certamente traz uma elevação da demanda no setor de Serviços, setor de trabalho intensivo, e sobre qual deve recair a maior responsabilidade a respeito das disparidades

regionais brasileiras. Segundo e além disso, embora a experiência recente do país com as políticas de transferência de renda mostre a importância desse tipo de políticas com foco nos indivíduos para as disparidade regionais de renda (Silveira Neto e Azzoni, 2012), as evidências obtidas neste trabalho indicam que as disparidades regionais brasileiras são bem mais sensíveis (absoluta e relativamente) a movimentos na distribuição regional de renda associada aos níveis mais altos de escolaridade.

REFERÊNCIAS

AZZONI, C. R. Concentração regional e dispersão das rendas per capita estaduais: análise a partir de séries históricas estaduais de Pib, 1939-1995. **Estudos Econômicos**. São Paulo, v.27, n. 3, p-341-393. Set-Dez, 1997.

BARROS, R. *et al.* Uma análise das principais causas da queda recente na desigualdade de renda brasileira. **Econômica**, n. 8, 117–147. 2006.

_____. **Determinantes da queda da desigualdade de Renda no Brasil**. Texto para discussão. Ipea, 2010.

CHANTREUIL, F.; TRANNOY, A. Inequality decomposition values: the trade-off between marginality and efficiency, **Journal of Economic Inequality**, 11:83-98, 2013.

CHANTREUIL, F., TRANNOY, A. Inequality decomposition values: the trade-off between marginality and consistency. Working Papers 99-24, **THEMA**, Université de Cergy-Pontoise, 1999.

COWELL, A. F.; FIORIO, C. V. Inequality Decomposition – A Reconciliation, **Journal of Economic Inequality**, December 2011, v. 9, Issue 4, pp 509-528.

DUARTE, A.; FERREIRA, P.C.; SALVATO, M. **Disparidades regionais ou educacionais?** Um exercício com contrafactuais, 2003. Mimeografado.

FEL, J.; RANIS, G.; KUO, S. **Growth with Equity**: The Taiwan Case. New York: Oxford University Press, 1979.

HOFFMANN, Rodolpho. Transferências de renda e a redução da desigualdade no Brasil e cinco regiões entre 1997 e 2004. **Econômica**, v. 8(1), p. 55-81. jun. 2006.

HOFFMANN, R.; NEY, M. G. A recente queda da desigualdade de renda no Brasil: análise de dados da Pnad, do Censo Demográfico e das Contas Nacionais. **Econômica**, v. 10, n. 1, p. 7-39, 2008.

KAKWANI, Nanaki. **Income Inequality and Poverty**: Methods of Estimation and Policy Applications. New York: Oxford University Press, 1980.

LEMAN, R.; YITZHAKI, S. Income Inequality Effects by Income Source: a new approach and applications to the United States. **Review of Economics and Statistics**, v. 67, n. 1, 151-156. 1985.

OSORIO, R. G.; SOUZA, P. H. G. F. **Evolução da pobreza extrema e da desigualdade de renda na Bahia: 1995 a 2009**. Brasília: Ipea, 2012. (Texto para Discussão, n. 1.696).

POSE, A. R.; TSELIOS, V. **Education and Income inequality in the regions of European Union**. v. 49, 411-437, 2009.

PYATT, G., CHEN, C., & FEL, J. The distribution of income by factor components. **Quarterly Journal of Economics**, 95, 451-474, 1980.

SASTRE, M., TRANNOY, A. Shapley inequality decomposition by factor components: some methodological issues. **J. Econom.** Supplement 9: 51-90, 2002.

SHORROCKS, A. F. **Decomposition procedures for distributional analysis**: a unified framework based on the shapley value. University of Essex: mimeo, 1999.

SHORROCKS, A. F. Decomposition Procedures for Distributional Analysis: A Unified framework Based on the Shapley Value. **Jornal of Economic Inequality**, v. 11, n. 1, p. 99-126, 2012.

SILVEIRA, R. M. N.; MENEZES, T. A. Disparidades regionais de renda no Brasil: analisando a importância do capital humano. *In*: **Fundação Konrad-Adeneur**. (Org.), Fórum Brasil-Europa; Fundação Konrad-Adeneur. 2008.

SILVEIRA, R. M. N.; AZZONI, C. R. Non-Spatial Government Policies and Regional Income Inequality in Brazil. **Regional Studies**, v. 45, n. 4, p.453-461, 2011.

_____. Social policy as regional policy: market and non-market factors determining regional inequality. **Journal of Regional Science**, v. 51, p. 1-18, 2012.

SOARES, S. Análise de bem-estar e decomposição por fatores da queda na desigualdade entre 1995 e 2004. **Econômica**, v. 8, n.1, p. 83-115, 2006.

SOARES, S. *et al.* Os impactos do benefício do Programa Bolsa Família sobre a desigualdade e a pobreza. *In*: CASTRO, J. A.; MODESTO, L. (Eds.). **Bolsa Família 2003-2010: avanços e desafios**. Brasília: Ipea, v. 2, p. 366, 2010.

SOUZA, P. H. G. F. **As causas imediatas do crescimento da renda, da redução da desigualdade e da queda da extrema pobreza na Bahia, no Nordeste e no Brasil entre 2003 e 2011**. Brasília: Ipea (Texto para Discussão, n. 1.816), 2013.

(Originais submetidos em setembro de 2014. Última versão recebida em julho de 2015. Aprovada em setembro de 2015.)

DETERMINANTES DAS MIGRAÇÕES INTERESTADUAIS NO BRASIL: EVIDÊNCIAS A PARTIR DE UM MODELO GRAVITACIONAL

Hilton Martins de Brito Ramalho¹

Erik Figueiredo²

José Luis da Silva Netto Júnior³

Este artigo investiga os principais determinantes dos fluxos migratórios interestaduais no Brasil. Para tanto, estimou-se um modelo gravitacional com base em de um painel de dados das Pesquisas Nacionais por Amostra de Domicílios (Pnads) anuais entre 2004 e 2009. O modelo econométrico permite identificar a equação gravitacional, a acomodação de fluxos zero, a robustez à heterocedasticidade/não normalidade e controlar efeitos não observados de atratividade de várias regiões sobre os fluxos bilaterais de migração (resistência multilateral). Os resultados sugerem que as diferenças de rendimento do trabalho esperado por ocupação, a dimensão populacional, a proximidade geográfica e as redes sociais são importantes condicionantes dos recentes movimentos migratórios. Essas variáveis mantêm-se relevantes, mesmo considerando estimações econométricas para trabalhadores migrantes em diferentes faixas de idade e de instrução.

Palavras-chave: migrações interestaduais; modelo gravitacional; Brasil.

A GRAVITY MODEL FOR INTER STATE MIGRATION IN BRAZIL

This paper aims to investigate the main determinants of interstate migration flows in Brazil. A gravity model was used controlling multilateral resistance to migration based on The National Household Sample Survey (PNAD) dataset from 2004 to 2009. The method applied is adequate to identify the gravity equation, handle zero migration flows and heteroskedasticity/non-normal data robust, controlling for unobservables in terms of multilateral attractiveness. The results suggest regional inequalities in expected wage by occupation, population size, short distances and social networks are important determinants of migrations streams. These correlations survive even after we consider estimates based on age and education groups.

Keywords: interstate migrations; gravity model; Brazil.

JEL: O15; R23; C34; C32.

1. Professor do Programa de Pós-graduação em Economia da Universidade Federal da Paraíba (PPGE/UFPB). *E-mail:* <hilton@ccsa.ufpb.br>.

2. Professor do PPGE/UFPB. Pesquisador do CNPq, Brasil. *E-mail:* <eafigueiredo@gmail.com>.

3. Professor do PPGE/UFPB. *E-mail:* <juniorluis@yahoo.com>.

1 INTRODUÇÃO

Durante as últimas décadas, o debate sobre os movimentos populacionais tem se destacado nas esferas acadêmica e política ao redor do mundo, especialmente no tocante à discussão de políticas públicas em países desenvolvidos. A grande população, extensão territorial e elevada desigualdade regional do Brasil são características que tornam o estudo das migrações internas fundamental para a compreensão das dinâmicas econômica, demográfica e social.

É fato estilizado que a dinâmica migratória brasileira, em sua grande parte, tem sido diretamente associada aos diferenciais de riqueza entre os estados das regiões Nordeste e Sudeste (Cançado, 1999; Netto Júnior *et al.*, 2003). Dados dos censos demográficos do Instituto Brasileiro de Geografia e Estatística (IBGE), abrangendo desde a década de 1970 até a década de 1990, revelam que, historicamente, os movimentos migratórios têm se processado no sentido Nordeste – Sudeste. Em especial, destacam-se, por um lado, os estados de São Paulo e Rio de Janeiro, enquanto importantes áreas de destino, e, por outro, os estados nordestinos como principais polos de origem. Contudo, deve ser ressaltada a reversão dessa dinâmica durante as últimas décadas, associada, sobretudo, à redução da taxa de natalidade e à migração de retorno (Oliveira; Ervatti; O’neill, 2011).

O entendimento dos fatores econômicos que direcionam os fluxos migratórios tem sido alvo de investigação crescente entre economistas. No Brasil, os estudos sobre essa temática aumentaram consideravelmente nas últimas décadas. Sahota (1968) foi um dos primeiros a investigar empiricamente se migrações interestaduais no Brasil estão relacionadas com desigualdades regionais de renda, dinâmica urbana e distâncias geográficas. Usando dados do censo demográfico de 1950 e modelos de regressão linear, o referido autor achou evidências de que os movimentos populacionais estão diretamente associados aos diferenciais de renda estadual e inversamente relacionados às distâncias entre capitais. Justo e Silveira Neto (2006) usaram dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) para estimar um modelo linear de dados em painel, tendo a taxa líquida de migração interestadual como variável dependente. Seus resultados sugerem uma relação direta entre a taxa líquida de migração e a renda esperada (renda *per capita* ponderada pela taxa de ocupação) dos estados, achado que se manteve após a ponderação dessa última variável por uma matriz de peso espacial. Arbex, Freguglia e Chein (2013) estimaram a probabilidade de migração de trabalhadores condicionada à habilidade e aos diferenciais salariais. Seus resultados, tendo como base as informações da Pesquisa de Economia Informal Urbana do IBGE, indicam a existência de uma relação inversa entre a probabilidade de migração de trabalhadores informais e seus respectivos níveis educacionais. Por sua vez, Freguglia, Gonçalves e Silva Ribeiro (2014) analisaram os determinantes e a composição dos fluxos migratórios interestaduais entre 1995 e 2006, tendo como base as informações da RAIS-Migra

e do IBGE. Os resultados, obtidos por meio de um modelo *logit* com efeitos fixos, sugerem que os diferenciais salariais exercem papel relevante no tocante à migração de trabalhadores qualificados.

Na literatura internacional, vários estudos documentam evidências sobre a influência de variáveis econômicas na explicação de fluxos migratórios entre países. Vogler e Rotte (2000), por exemplo, investigaram os determinantes dos fluxos migratórios dirigidos de países africanos e asiáticos para a Alemanha durante o período de 1981 a 1995. Seus resultados mostram que as redes sociais (presença de pessoas conhecidas, amigos, familiares, grupos étnicos/religiosos etc) funcionam como importante incentivo para as migrações. Ademais, seus achados sugerem um possível efeito limitado gerado por restrições financeiras e preferência pelo consumo no país de origem. Clark, Hatton e Williamson (2007), em estudo para migrações direcionadas aos Estados Unidos (EUA) entre 1971 e 1998, também indicam que conexões por meio de redes sociais desempenham um papel relevante nas explicações dos contingentes populacionais recebidos pelo último país. Seus achados empíricos também revelaram uma relação inversa entre os fluxos de migração e as distâncias entre os países de origem e os EUA. Já o trabalho de Mayda (2010) foi realizado para migrações entre quatorze países membros da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) durante 1980 e 1995. Suas principais evidências sugerem que países com maior nível de renda tendem a receber mais migrantes e que as distâncias bilaterais parecem atuar negativamente sobre os fluxos migratórios. Afora os trabalhos já citados, cabe também destacar o estudo de Beine, Docquier e Özden (2011) a partir de um painel de dados sobre fluxos migratórios dirigidos de 195 países para trinta países da OCDE entre 1990 e 2000. Os referidos autores analisaram os efeitos das diásporas (concentração de migrantes de mesma origem vivendo em um mesmo país) sobre a estrutura de seletividade em educação dos fluxos migratórios. Os principais resultados sugerem que as diásporas reduzem os custos de migração a partir de conexões por redes sociais, favorecendo a atração de migrantes com baixa qualificação.

Apesar do conjunto de evidências reportadas na literatura especializada, é importante ressaltar que poucos estudos têm se preocupado em considerar empiricamente os efeitos de atratividade, de restrições e/ou de políticas que várias regiões de destino potencial podem exercer sobre fluxos migratórios bilaterais (Beine; Parsons, 2012; Bertoli; Fernández-Huertas Moraga, 2012; Ortega; Peri, 2013). Em estudo recente, Bertoli e Fernández-Huertas Moraga (2013) denominam esse conjunto de efeitos não observados como resistência multilateral, destacando sua relação com pseudo-modelos gravitacionais⁴ derivados de agregações de modelos

4. Os modelos gravitacionais são bastante conhecidos na literatura sobre comércio internacional. Eles fazem adaptação da Lei Gravitacional de Isaac Newton, em que a massa dos corpos é substituída pela massa econômica (nível de renda), incluindo distâncias bilaterais com importante força de atrito. Para mais detalhes, vide Head e Mayer (2014).

de utilidade aleatória para decisão individual de migração do trabalho. Em linhas gerais, os autores em destaque mostram que os efeitos de resistência multilateral podem ser incorporados ao termo de erro probabilístico de modelos gravitacionais considerando uma estrutura complexa de correlações seriais e espaciais, representada por um vetor de fatores específicos às regiões de origem/destino e variantes no tempo. A omissão desses fatores pode gerar vieses nas estimativas dos efeitos de variáveis econômicas relevantes, que são associadas à parte determinística da função utilidade envolvida na decisão de migração.

No Brasil, são poucos os estudos que analisaram os determinantes das migrações internas de acordo com um painel de dados agregados de fluxos populacionais⁵, especialmente explorando a estrutura inerente aos dados para controlar efeitos de variáveis omitidas, a exemplo dos chamados efeitos de múltiplas atratividades regionais (resistência multilateral). Ademais, não há na literatura nacional sobre migrações internas evidências satisfatórias sobre os possíveis efeitos de redes sociais e de diferenciais salariais ponderados por ocupação e risco de desemprego⁶. Portanto, o objetivo deste estudo é investigar os principais determinantes das migrações interestaduais no Brasil durante a década de 2000, a partir da estimação de um modelo gravitacional com fatores de resistência multilateral e robusto à presença de fluxos zero, à log-linearização e a não normalidade dos dados.

Incluindo esta introdução, este estudo está dividido em cinco partes. Na segunda seção é apresentado um breve panorama da dinâmica migratória interestadual brasileira na década de 2000. Na terceira parte, são feitas considerações teóricas associadas à equação gravitacional, estratégia de estimação, base de dados e seleção amostral. Na quarta seção, são apresentados os resultados do estudo. Finalmente, a quinta seção é reservada às considerações finais.

2 MIGRAÇÕES INTERESTADUAIS NO BRASIL: FATOS OBSERVADOS

A dinâmica migratória no Brasil durante as últimas décadas, sobretudo no período de 1950 a 1980, notabilizou-se pela polarização dos movimentos populacionais majoritariamente no sentido Nordeste-Sudeste. Todavia, a partir da década de 1980, os grandes centros de atração de migrantes (Rio de Janeiro e São Paulo) começaram a perder expressão, marcando uma mudança no padrão de migração

5. A esse respeito vide os estudos de Sabbadini e Azzoni (2006) e de Golgher, Rosa e Araújo Júnior (2008). Os primeiros autores usaram dados dos censos demográficos de 1991 a 2000 e um modelo gravitacional para investigar os determinantes da migração interestadual de pessoas com elevada instrução no Brasil. Suas evidências sugerem que a renda média estadual é um importante determinante da fuga de cérebros. Por sua vez, Golgher, Rosa e Araújo Júnior (2008) também analisaram os determinantes das migrações internas no Brasil, contudo, usando dados censitários para fluxos de migração por mesorregiões. Seus achados apontam que os migrantes da região Nordeste estão sujeitos a uma "armadilha de pobreza".

6. Harris e Todaro (1970) contribuíram para a abordagem econômica das migrações ao considerar o papel da taxa de desemprego como fator determinante da mobilidade populacional. No modelo dos referidos autores, a diferença de salário esperado (valor presente do salário de mercado ponderado pela taxa de desemprego) é a variável mais importante para a decisão de migrar.

interestadual no Brasil (Oliveira; Ervatti; O'Neill, 2011). Nesta seção, apresenta-se um breve panorama comparativo dos padrões e das tendências de absorção/emissão líquida de pessoas por parte dos estados brasileiros ao longo das décadas de 1980, 1990 e 2000, procurando-se também identificar as rotas predominantes de migração na década mais recente.

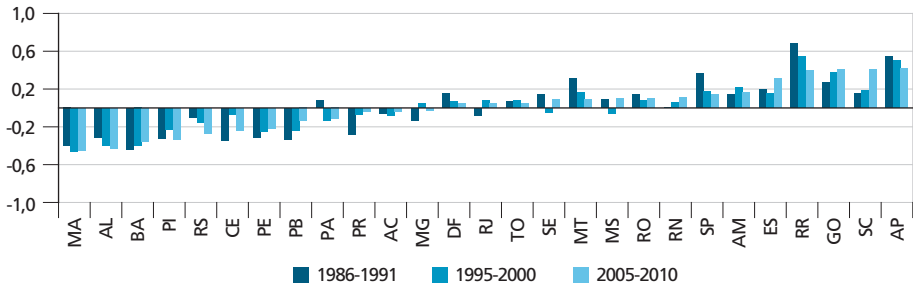
Para caracterizar os fatos observados a respeito das migrações interestaduais durante as últimas três décadas no Brasil, foi calculado o Índice de Eficácia Migratória (IEM), que é a razão entre o saldo migratório (diferença entre o total de imigrantes e emigrantes interestaduais) e a soma do total de emigrantes e imigrantes interestaduais⁷. Ou seja, o IEM foi computado para todos os estados e o Distrito Federal (DF) tendo como base as informações dos censos demográficos de 1991, 2000 e 2010. Ademais, o critério usado para identificação do migrante interestadual foi o de data fixa, isto é, aquele indivíduo que na data do censo demográfico afirmou residir permanentemente em um estado diferente daquele que morava há cinco anos.

A figura 1 apresenta a evolução do IEM por estados brasileiros (inclusive DF) considerando distintas coortes de migração nos períodos 1986-1991, 1995-2000 e 2005-2010. A maioria dos estados brasileiros pode ser caracterizada como áreas de baixa evasão ou de baixa absorção migratória, segundo a classificação proposta por Oliveira, Ervatti e O'Neill (2011)⁸. Considerando dados da coorte mais recente de migração (2005-2010), é possível verificar que entre o grupo de média evasão migratória destacam-se os estados do Maranhão, Alagoas, Bahia e Piauí. Em seguida, Rio Grande do Sul, Ceará, Pernambuco, Paraíba e Pará completam o grupo de áreas de baixa evasão de migrantes. Por outro lado, os estados do Amapá, Santa Catarina, Goiás, Roraima e Espírito Santo podem ser classificados como áreas de média absorção migratória, ao passo que Amazonas, São Paulo, Rio Grande do Norte, Rondônia e Mato Grosso do Sul surgem como polos de baixa absorção de migrantes interestaduais. Ademais, no período de 2005-2010 oito estados podem ser rotulados como áreas de rotatividade migratória, isto é, áreas onde há um equilíbrio entre entradas e saídas de migrantes, são eles: Mato Grosso, Sergipe, Tocantins, Rio de Janeiro, Distrito Federal, Minas Gerais, Acre e Paraná.

7. O IEM foi proposto por Baeninger (2000). Esse indicador permite comparações de trocas populacionais entre regiões no tempo, uma vez que se situa no intervalo -1 e 1. Valores próximos a -1 indicam alta capacidade de evasão populacional; valores próximos a 1 sugerem alta absorção populacional e valores próximos a zero, equilíbrio entre entradas e saídas de pessoas.

8. A classificação considera: (a) Área de forte evasão migratória (IEM entre -1,00 e -0,51); (b) Área de média evasão migratória (IEM entre -0,50 e -0,30); (c) Área de baixa evasão migratória (IEM entre -0,29 e -0,10); (d) Área de rotatividade migratória (IEM entre -0,09 e 0,09); (e) Área de baixa absorção migratória (IEM entre 0,10 e 0,29); (f) Área de média absorção migratória (IEM entre 0,30 e 0,50) e (g) Área de forte absorção migratória (IEM entre 0,51 e 1,00).

FIGURA 1
Índice de Eficácia Migratória (IEM) por estados nos períodos de migração 1986-1991, 1995-2000 e 2005-2010



Fonte: Elaboração própria a partir dos microdados dos Censos Demográficos de 1991, 2000 e 2010. Nota: Dados ordenados pelo IEM do período de migração 2005-2010.

Os dados da figura 1 também chamam a atenção para o aumento da capacidade de absorção de migrantes nos estados de Santa Catarina, Goiás e Espírito Santo durante as últimas três décadas, por um lado, e, por outro, para a redução de atração populacional nos estados de São Paulo, Roraima e Amapá. No tocante aos estados com tendência à evasão de migrantes, com exceção de Alagoas, Rio Grande do Sul, Maranhão e Piauí, que têm aumentado a emissão líquida de pessoas, as demais unidades federativas do grupo em destaque registram uma tendência de redução na força de emissão líquida da população (Bahia, Ceará, Pernambuco e Paraíba).

Vale ressaltar o caso dos estados do Paraná e de Minas Gerais, que no quinquênio 1986-1991 eram áreas de baixa evasão de migrantes, passando para grupo de rotatividade migratória a partir da década de 2000. Por sua vez, Mato Grosso, Sergipe e o Distrito Federal, que entre 1986-1991 eram espaços de baixa absorção migratória, também tornaram-se estados de rotatividade. Por outro lado, Mato Grosso do Sul e Rio Grande do Norte, que se caracterizam com regiões de rotatividade de migrantes no período 1986-1991, transformaram-se em áreas de baixa absorção de migrantes interestaduais entre 2005-2010. Por fim, o estado do Pará, que no período 1986-1991 era classificado como local de rotatividade, passa para área de baixa evasão populacional a partir da década de 2000, conforme os dados censitários.

É incontestável a mudança dos padrões migratórios ocorrida nas últimas décadas. A polarização Nordeste-Sudeste não tem a escala de outrora e algumas razões são apontadas para ocorrência de tal processo. Dois principais fatores são o crescimento da migração de retorno e a redução da taxa de natalidade. Em linhas gerais, a migração de retorno tem se processado com especial intensidade por naturais da região Nordeste antes residentes nos estados do Sudeste (Oliveira; Ervatti; O'neill, 2011; Queiroz; Ramalho, 2011).

Somada à importância da migração de retorno, houve uma maior retenção da população do Nordeste em sua própria região, associada ao aumento da dinâmica migratória intrarregional e à redução da atratividade das demais regiões. Também são apontados o crescimento da importância dos movimentos pendulares, o surgimento de novos polos de atração de migrantes no interior do Brasil (com destaque para as cidades de médio porte) e a redução da taxa de crescimento populacional (Dedecca; Cunha, 2004).

No período entre 2005-2010, os dados censitários apontam que a migração interestadual no Brasil movimentou cerca de 4,6 milhões de pessoas ou o equivalente a 2,4% da população do país em 2010. A t, a seguir, fornece uma visão sintética das rotas da migração interestadual no Brasil ao mostrar a distribuição percentual dos migrantes de data fixa, segundo os estados de residência em 2005 e em 2010.

Os dados mostram a importância dos estados de São Paulo, Minas Gerais, Goiás, Santa Catarina, Paraná e Rio de Janeiro na absorção dos migrantes interestaduais no período de 2005-2010. Esses estados foram, respectivamente, destinos de 21,4%, 8,1%, 7,8%, 6,5%, 5,9% e 5,8% dos migrantes. São Paulo, estado responsável pela maior parcela da produção de riquezas no Brasil, além de registrar peso considerável no processo de absorção de migrantes, destaca-se como área de preferência dos emigrados da região Nordeste, também absorvendo parte considerável de emigrantes procedentes de estados da própria região Sudeste, Sul e Centro-Oeste.

Minas Gerais destaca-se como destino de migrantes oriundos dos estados do Espírito Santo, São Paulo, Rio de Janeiro e Goiás. O estado de Goiás, por seu turno, tem atraído, sobretudo, migrantes procedentes do Tocantins, Distrito Federal, Mato Grosso e Maranhão. Santa Catarina apresenta poder de atratividade mais restrito a sua própria macrorregião, isto é, recebeu principalmente emigrados dos estados do Rio Grande do Sul e do Paraná. Apenas o Paraná, na região Sul, conseguiu atrair grandes contingentes procedentes de outras regiões, como emigrados do Mato Grosso do Sul, São Paulo, Mato Grosso e Rondônia. Já o estado do Rio de Janeiro evidencia-se pela absorção de migrantes que se deslocaram do Espírito Santo, Minas Gerais e de alguns estados do Nordeste (Paraíba, Ceará e Rio Grande do Norte).

O padrão de absorção de migrantes dos estados da região Norte é bem restrito à própria área, mas também caracteriza-se por um elevado número de pessoas procedentes dos estados do Maranhão, Mato Grosso e Goiás. Os estados localizados nas regiões Centro-Oeste e Nordeste também registram maior influência na atração de emigrados das próprias regiões. No caso dos estados da região Centro-Oeste, destaca-se a absorção de pessoas saídas dos estados do Maranhão, Piauí e Rondônia. Na região Nordeste, o estado do Maranhão tem recebido contingentes do Pará e de Roraima, enquanto o da Bahia tem atraído emigrantes de São Paulo e do Espírito Santo.

TABELA 1
Distribuição dos migrantes interestaduais no Brasil segundo o estado de residência em 2005 (saída) e por estado de residência em 2010 (entrada)
 (Em %)

Estado de residência em 2005	Estado de residência em 2010																																				
	Rondônia	Acre	Amazonas	Roraima	Pará	Paráíba	Ceará	Piauí	Maranhão	Tocantins	Amapá	Roraima	Amazonas	Acre	Rondônia	Roraima	Amazonas	Acre	Rondônia	Pará	Paráíba	Ceará	Piauí	Maranhão	Tocantins	Amapá	Roraima	Amazonas	Acre	Rondônia	Total						
	6,3	12,5	1,4	2,5	0,1	0,7	1,0	0,3	1,7	0,5	0,8	1,3	0,3	0,2	1,9	3,4	4,5	1,6	7,8	11,2	2,6	1,2	4,4	25,7	4,5	1,7	100										
	34,0	-	23,8	1,2	1,9	0,1	0,6	0,4	0,1	2,3	0,8	0,3	1,0	0,5	0,1	0,5	3,0	0,7	3,2	4,8	2,4	1,7	1,0	1,5	5,3	5,7	3,3	100									
	12,0	8,0	-	10,5	20,7	1,0	0,5	3,7	0,9	4,0	1,7	1,0	2,6	0,7	0,2	1,3	2,0	8,0	7,3	1,9	1,3	2,1	0,9	1,7	2,4	3,2	100										
	6,1	0,3	26,9	-	9,9	0,4	1,2	13,2	0,8	5,5	2,2	2,2	4,2	0,7	0,0	1,2	3,7	0,9	3,1	4,6	1,7	1,1	2,7	1,5	1,7	3,6	2,5	100									
	1,4	0,2	13,6	3,3	-	14,1	8,7	14,4	1,5	3,2	0,5	0,6	1,2	0,2	0,2	2,1	2,6	0,8	3,5	6,1	1,8	1,3	0,8	0,6	4,3	10,7	2,2	100									
	0,8	0,0	4,5	0,8	55,2	-	0,8	9,0	1,6	4,6	2,6	1,2	1,3	0,0	0,3	0,4	2,8	0,6	1,1	4,0	1,2	0,8	0,7	0,1	0,6	1,9	3,4	100									
	1,4	0,1	0,2	20,7	0,2	-	9,0	1,3	0,8	0,2	0,3	0,8	0,2	0,2	1,7	4,0	0,3	0,7	4,4	1,2	0,7	0,6	1,0	3,2	40,5	5,9	100										
	1,0	0,1	1,5	1,8	23,1	1,2	7,3	-	6,6	1,9	0,2	0,4	0,8	0,1	0,1	0,7	3,6	0,4	4,0	14,5	0,6	0,8	0,3	0,3	5,3	16,2	7,4	100									
	0,3	0,1	0,8	0,2	2,9	0,2	1,7	11,5	-	5,4	0,6	0,5	3,0	0,2	0,2	2,4	2,7	0,4	3,1	37,2	0,7	0,5	0,5	0,5	1,6	10,2	12,7	100									
	0,9	0,2	1,6	0,5	3,0	0,4	0,7	2,6	4,3	-	4,6	2,7	5,2	0,7	0,4	4,1	5,1	0,5	14,3	33,2	1,3	1,8	1,4	0,7	1,0	3,9	4,9	100									
	0,8	0,1	1,3	0,7	1,5	0,5	0,7	0,7	0,8	9,8	-	17,3	6,2	1,2	0,6	3,2	4,4	0,9	10,5	22,1	1,0	1,5	1,7	1,0	0,6	5,8	5,2	100									
	0,7	0,1	0,2	0,3	0,8	0,1	0,3	1,0	0,5	3,6	10,0	-	13,9	0,7	0,6	3,3	2,7	0,4	19,3	30,4	1,2	1,2	0,7	0,5	1,3	3,7	2,7	100									
	0,4	0,1	0,8	0,3	1,3	0,1	0,6	1,2	1,3	4,5	3,1	9,9	-	7,1	1,4	8,0	3,5	0,6	6,6	39,1	1,3	1,6	0,6	1,0	1,7	2,6	1,6	100									
	0,3	0,0	0,3	0,1	0,3	0,1	0,6	0,6	0,4	1,0	0,8	1,0	1,1,9	-	8,6	6,3	11,1	2,1	3,0	38,2	1,4	1,2	0,2	1,6	5,0	3,1	0,9	100									
	0,5	0,0	0,3	0,1	1,1	0,1	0,5	0,7	0,5	1,0	0,9	1,0	2,8	10,1	-	25,5	4,5	1,8	6,1	32,7	1,5	1,7	0,7	0,6	1,0	1,8	2,6	100									
	0,4	0,1	0,2	0,1	0,7	0,0	0,6	0,6	0,6	1,1	0,5	0,7	3,9	0,9	3,8	-	8,2	8,5	5,8	46,1	1,4	1,6	0,6	0,5	0,8	8,0	4,5	100									
	1,0	0,2	0,4	0,1	1,5	0,1	0,9	0,7	0,3	0,8	0,4	0,4	0,9	0,3	0,2	5,3	-	9,6	12,0	41,5	3,0	1,3	1,0	1,2	1,3	9,2	6,5	100									
	3,2	0,1	0,4	0,1	1,1	0,1	0,2	0,6	0,4	0,7	0,4	0,4	0,9	0,2	0,3	15,7	30,8	-	25,4	10,5	1,7	1,3	1,0	0,7	0,8	1,6	1,7	100									
	0,4	0,2	2,2	0,2	1,8	0,2	1,4	0,6	0,8	5,0	3,7	6,7	4,0	0,9	1,3	5,6	18,4	9,3	-	18,8	3,2	2,8	3,4	1,8	0,6	1,9	5,0	100									
	0,8	0,1	0,4	0,1	1,0	0,1	0,6	1,5	2,4	4,5	1,8	3,2	6,5	2,4	1,5	12,2	19,8	1,5	5,6	-	15,3	5,0	2,0	4,6	2,2	3,4	1,8	100									
	1,7	0,2	0,5	0,2	0,7	0,1	0,5	0,4	0,1	0,5	0,2	0,3	0,4	0,1	1,6	3,2	0,4	1,7	27,4	-	40,3	5,8	5,0	6,1	1,6	0,9	100										
	1,5	0,1	0,4	0,2	0,5	0,1	0,3	0,4	0,2	0,6	0,5	0,4	0,8	0,3	0,1	1,2	2,4	0,4	2,9	14,1	38,8	-	24,8	2,2	3,2	1,9	2,0	100									
	0,8	0,1	0,9	0,2	0,9	0,1	0,4	0,4	0,4	1,0	0,4	0,5	1,0	0,3	0,3	1,9	2,1	0,5	4,3	9,8	12,3	51,9	-	2,2	3,4	1,7	2,3	100									
	2,7	0,4	0,4	0,3	1,1	0,1	0,6	0,5	0,4	0,5	0,8	1,0	0,6	0,3	0,3	1,6	4,4	0,7	1,4	28,8	18,1	5,1	3,6	-	16,6	5,7	1,3	100									
	11,4	0,6	1,3	0,3	6,8	0,2	2,4	1,8	0,7	0,6	0,4	0,3	0,7	0,6	0,3	1,8	3,9	0,7	1,1	12,5	13,1	3,1	3,1	11,3	-	17,0	1,8	100									
	1,6	0,1	0,5	0,3	5,9	0,1	12,1	3,5	2,1	1,4	0,9	0,9	0,9	0,2	0,2	5,3	14,4	0,4	1,6	9,4	1,7	1,4	1,2	2,4	10,3	-	21,1	100									
	0,2	0,1	0,7	0,2	1,3	0,1	2,5	3,9	4,8	3,0	1,4	2,1	1,6	0,4	0,7	5,0	7,8	1,1	4,5	6,1	1,5	1,3	1,5	0,4	0,8	47,0	-	100									
	1,4	0,3	1,5	0,6	3,5	0,8	1,9	2,3	1,6	2,4	1,5	2,1	3,2	1,2	1,1	4,9	8,1	2,8	5,8	21,4	5,9	6,5	2,2	2,1	3,1	7,8	4,1	100									

Fonte: Elaboração própria a partir dos microdados dos censos demográficos de 1991, 2000 e 2010. Nota: Dados expandidos para o universo.

Em linhas gerais, os dados da tabela 1 sugerem que os estados de São Paulo e do Rio de Janeiro são relevantes na distribuição espacial dos fluxos migratórios, especialmente na atração de pessoas procedentes dos estados da região Nordeste. Caberia saber, portanto, em que medida as conhecidas desigualdades regionais de renda e de ocupação no Brasil podem condicionar esses movimentos populacionais? Por outro lado, a redistribuição espacial da população brasileira também parece se processar segundo estados vizinhos, sugerindo pistas sobre a importância dos custos de migração (gastos com viagens e hospedagem, gastos para acesso à informação, gastos com procura por trabalho), que podem ser afetados pela presença de conexões entre pessoas conhecidas, amigos, familiares, sindicatos e grupos religiosos entre regiões. Essas questões passam a ser melhor exploradas conforme a estratégia empírica discutida a seguir.

3 ESTRATÉGIA EMPÍRICA

A análise empírica de movimentos populacionais a partir de modelos gravitacionais tem se caracterizado como um fenômeno recente, a despeito do inegável sucesso de tais modelos em achados sobre comércio internacional. Acompanhando essa tendência, é possível encontrar na literatura especializada novos desenvolvimentos teóricos que incorporam fundamentos microeconômicos à conhecida equação gravitacional (Anderson, 1979; Anderson; Wincoop, 2003). Por outro lado, a pouca disponibilidade de dados sobre fluxos migratórios, a presença de muitos fluxos zero em bases de dados disponíveis e os problemas de identificação estrutural da equação gravitacional ainda se apresentam como principais desafios para a produção de novas evidências na área de pesquisa (Santos Silva; Tenreyro, 2006; Head; Mayer, 2014).

Nesta seção, faz-se uma apresentação da fundamentação teórica do modelo gravitacional de migração, estratégias de estimação e tratamento da base de dados empregada para obtenção dos resultados empíricos.

3.1 EQUAÇÃO GRAVITACIONAL: TEORIA, IDENTIFICAÇÃO E ESTIMAÇÃO

3.1.1 Modelo gravitacional de migração com resistência multilateral

Anderson (2011) desenvolveu um modelo estrutural de gravitação aplicado à migração do trabalho, no qual explicita o papel de forças de atratividade/barreiras exercidas por várias regiões no tocante à determinação de fluxos migratórios bilaterais.

Considere um mundo com n regiões, tal que w^i e w^j são os rendimentos médios do trabalho esperados nas regiões i e j , $\forall i \neq j$, respectivamente. Cada trabalhador h que avalia migrar da região j para a região i tem uma função utilidade $u_{ijh} = \frac{w^i}{\delta^{ij} w^j} \varepsilon^{jih}$, em que $\delta^{ij} > 1$ é o custo de migração diretamente relacionado com a distância entre tais regiões (custo de transporte do tipo *iceberg*), $\frac{w^i}{\delta^{ij} w^j}$ é a

diferença de renda esperada entre as regiões líquida do custo de migração (parte observada da função utilidade) e ε^{jib} representa um componente estocástico específico a cada indivíduo. Entre as $n - 1$ regiões de destino potencial, o trabalhador migra para aquela com máximo valor de utilidade esperada $u_{jib}^* = \max(u_{jib})$, com $u_{jib} = \geq 0$ para ao menos uma região i .

A probabilidade de migração de trabalhadores da região j para i , considerando a agregação de todos os potenciais migrantes em j , é equivalente à proporção de trabalhadores de j que escolhem migrar para i . Assim, pode-se prever o fluxo migratório da região j para i por:

$$M^{ji} = G(v^{ji})N^j, \quad (1)$$

em que $G(v^{ji}) = \frac{\exp(v^{ji})}{\sum_{k=1}^n \exp(v^{jk})} = \frac{w^i/\delta^{ji}}{\sum_{k=1}^n w^k/\delta^{jk}}$ é uma função geradora de probabilidade individual de migração, isto é, uma função *Gumbel* de valor extremo do tipo 1; $v^{ji} = \ln(w^i) - \ln(w^j) - \ln(\delta^{ji})$ é a função utilidade esperada de cada trabalhador na forma logarítmica⁹ e N^j a população da região j .

A estrutura do modelo também incorpora condições *market clearing* para assegurar o equilíbrio geral em todos os mercados regionais. Para tanto, considere-se o total da força de trabalho fornecida por todas as regiões de origem para a região :

$$L^i = \sum_j M^{ji}, \quad (2)$$

seja a oferta mundial de trabalho dada por $N \equiv \sum_i N^i = \sum_i L^i$. Então, substituindo (1) em (2), tem-se, em equilíbrio, $L^i = w^i \sum_j ((1/\delta^{ji})/w^j)N^j$, de modo que:

$$w^i = \frac{L^i}{\Omega^i N}, \quad (3)$$

e

$$\Omega^i = \sum_j \frac{1/\delta^{ji} N^j}{W^j N}, \quad (4)$$

em que $w^j \equiv \sum_k w^k / \delta^{jk} = \sum_k \frac{1/\delta^{jk} L^k}{\Omega^k N}$.

9. O modelo adota a hipótese simplificadora, que todos os trabalhadores de enfrentam a mesma renda média e custos de migração, no entanto, são heterogêneos em suas preferências. Ademais, seguindo a literatura especializada, assume-se que a variável aleatória segue uma distribuição de probabilidade Gumbel de valor extremo do tipo 1.

Substituindo (3) e (4) em (1), chega-se a uma equação estrutural de migração generalizada¹⁰. Todavia, para obter uma forma de fácil manipulação e interpretação, Anderson (2011) supõe que u_{jih} é uma função de utilidade esperada com coeficiente constante de aversão ao risco (linearizada por logaritmo)¹¹. Portanto, a equação gravitacional de migração pode ser expressa como:

$$M^{ij} = \frac{L^i N^j}{N} \left(\frac{\delta^{ji}}{\bar{\Omega}^i \bar{W}^j} \right)^{1-\theta}, \quad (5)$$

em que $\bar{\Omega}^i = \left[\sum_j \frac{(\delta^{ji})^{1-\theta} N^j}{\bar{W}^j} \right]^{\frac{1}{1-\theta}}$ e $\bar{W}^j = \left[\sum_i \frac{(\delta^{ji})^{1-\theta} L^i}{\bar{\Omega}^i} \right]^{\frac{1}{1-\theta}}$.

Os termos $\bar{\Omega}^i$ e \bar{W}^j são determinados simultaneamente em equilíbrio geral e representam forças endógenas de fricção imputadas a várias alternativas de destino. Note-se que pode ser interpretado como uma média ponderada dos custos bilaterais de migração considerando fluxos migratórios recebidos pela região e oriundos de qualquer região, enquanto representa uma média de custos de migração de potenciais migrantes emitidos por j e dirigidos para qualquer região. Portanto, de forma análoga aos modelos gravitacionais aplicados ao comércio, esses últimos são considerados termos de resistência multilateral (TRM), que, embora não observados, podem ser computados mediante normalizações (Anderson; Wincoop, 2003; Anderson; Yotov, 2010). A equação (5) ainda pode ser decomposta em duas partes: $\frac{L^i N^j}{N}$, que significa a parcela da população originária de j em qualquer região de destino na ausência de custos de migração, e $\left(\frac{\delta^{ji}}{\bar{\Omega}^i \bar{W}^j} \right)^{1-\theta}$, que resume o efeito de fricções multilaterais exercidas por outras regiões sobre o fluxo migratório da região j para a região i .

3.2 ABORDAGEM ECONOMÉTRICA

3.2.1 Especificação do modelo

Seguindo a literatura especializada¹² considere que os custos de migração são diretamente associados às distâncias geográficas entre regiões, inversamente relacionados com o compartilhamento de fronteiras (vizinhança) e com o estoque de pessoas na região de destino que no passado procederam da mesma região de origem dos migrantes recentes (*proximity* para efeitos de redes sociais), conforme a seguinte equação:

$$\delta^{ij} = d_{ij}^{\beta_1} s_{ij}^{\beta_2} \exp(\beta_3 f_{ij}), \quad (6)$$

10. A equação gravitacional geral é dada por $M^{ij} = (L^i N^j / N) (1/\delta^{ij}) / (\Omega^i W^j)$.

11. Uma função de utilidade com coeficiente constante de aversão ao risco θ assume a forma $u(c) = c^{1-\theta} / (1-\theta) > 0$ e $\theta \neq 1$.

12. Vide Clark, Hatton e Williamson (2007), Mayda (2010), Beine e Parsons (2012) entre outros.

em que d^{ij} representa um custo fixo de migração (custo de transporte, por exemplo); s_{ij} é o estoque acumulado de pessoas na região i que nasceram na região j ; f_{ij} é uma variável binária que assume o valor 1, se as regiões de origem e de destino do migrante compartilham fronteiras, e 0 caso contrário; $\beta_1 > 0$, $\beta_2 > 0$ e $\beta_3 < 0$ e são parâmetros¹³.

Conforme pode ser observado nas equações (3) e (4), a oferta de trabalho recebida por cada região de destino L^i é determinada endogenamente em equilíbrio geral, a qual é função da diferença de renda esperada entre i e as demais regiões de origem ponderada por uma média de custos bilaterais entre todas as regiões do mundo (termos não observados de fricção). Portanto, para simplificar, pode-se aproximar L^i em (5) como uma função parcial e direta da razão de rendimento esperado do trabalho entre as regiões i e j e da população da região i :

$$L^i \approx \exp\left(\beta_4 \frac{w^i}{w^j}\right) (N^i)^{\beta_5}, \quad (7)$$

em que é a população da região de destino, e são parâmetros.

Seguindo Santos Silva e Tenreyro (2006), considere uma versão estocástica da equação (5) na forma exponencial com coeficientes de elasticidade constante $M_{ij} = \exp(\ln(M^{ij}))\varepsilon_{ij}$, em que ε_{ij} é uma variável aleatória não negativa. Considerando-se (6) e (7) e após algumas manipulações:

$$M_{ij} = \exp\left(\beta_0 + \beta_1 \ln(d_{ij}) + \beta_2 \ln(s_{ij}) + \beta_3 f_{ij} + \beta_4 \left(\frac{w^i}{w^j}\right) + \beta_5 \ln(N^i) + \beta_6 \ln(N^j) + r_{ij}\right) \varepsilon_{ij}, \quad (8)$$

em que $\beta_0 = \ln(1/N)$ e $\beta_6 > 0$ e são parâmetros e $r_{ij} = \ln\left(\left[\sum_j \frac{(\delta^{ij})^{1-\theta}}{w^j} \left(\frac{N^j}{N}\right)\right]^{\frac{1}{\theta-1}} \left[\sum_i \frac{(\delta^{ij})^{1-\theta}}{w^i} \left(\frac{L^i}{N}\right)\right]^{\frac{1}{\theta-1}}\right)^{1-\theta}$.

Cabe ressaltar que r_{ij} captura a influência de todas as alternativas de destino em termos de atratividade/obstáculos sobre a migração da região j para a região i em cada período, isto é, representa termos de resistência multilateral (TRM), que embora sejam não observados, podem ser inferidos (Anderson, 2011). Por outro lado, conforme observam Bertoli e Fernández-Huertas Moraga (2013), r_{ij} apresenta uma estrutura complexa de autocorrelação serial e espacial¹⁴ e correlaciona-se com

13. Observe-se que, se $\beta_3 < 0 \rightarrow \exp(\beta_3) < 1$, a migração entre estados vizinhos registra custo relativamente menor.

14. Bertoli e Fernández-Huertas Moraga (2013) adotam funções de geração de probabilidades Gumbel generalizadas e mostram suas relações com a estrutura de correlação em termos não observados na equação gravitacional.

as demais covariadas do modelo. Por exemplo, se por um lado o fluxo de migração oriundo da região j e dirigido para duas regiões distintas k e l pode ser afetado pela influência atrativa de outras regiões de destino potencial (r_{kj} correlacionado com r_{lj}), por outro, fluxos migratórios procedentes de duas regiões de origem j e h e direcionados para uma mesma região de destino k também são afetados pela resistência multilateral (r_{kj} e r_{kh} correlacionados). Ademais, a média de renda esperada da região i pode correlacionar-se com rendimentos de trabalho em outras regiões, sobretudo se essas regiões estiverem expostas à mesma estrutura de investimentos, infraestrutura, instituições, comércio e choques econômicos. Destarte, ignorar a influência das condições econômicas de outras regiões pode superestimar o efeito que a diferença de renda entre as regiões i e j tem sobre o fluxo migratório bilateral de j para i (Beine; Bertoli; Fernández-Huertas Moraga, 2014).

Outro aspecto que deve ser observado é que o modelo empírico (8) é não linear nos parâmetros e não pode ser estimado por Mínimos Quadrados Ordinários (MQO). Uma alternativa seria considerar uma transformação linear por logaritmos:

$$\ln(M_{ij}) = \beta_0 + \beta_1 \ln(d_{ij}) + \beta_2 \ln(s_{ij}) + \beta_3 f_{ij} + \beta_4 \left(\frac{w^i}{w^j} \right) + \beta_5 \ln(N^i) + \beta_6 \ln(N^j) + r_{ij} + \ln(\varepsilon_{ij}). \quad (9)$$

No entanto, Santos Silva e Tenreyro (2006) argumentam que, sob heterocedasticidade, não se pode identificar ambas as equações, (8) e (9). A fonte desse problema é a desigualdade de *Jensen*, isto é, as esperanças condicionais são diferentes $E(\ln(\varepsilon_{ij}) | d_{ij}, s_{ij}, f_{ij}, \frac{w^i}{w^j}, N^i, N^j, r_{ij}) \neq E(\varepsilon_{ij} | d_{ij}, s_{ij}, f_{ij}, \frac{w^i}{w^j}, N^i, N^j, r_{ij})$.

3.2.2 Estimação

A estimação do modelo gravitacional na forma log-linear (9) envolve várias dificuldades. Primeiro, o problema da desigualdade de *Jensen*, presente nas versões (8) e (9) do modelo gravitacional, implica, em viés de parâmetros produzidos pela log-linearização de um modelo estruturalmente não linear nos parâmetros. Segundo, a não incorporação da estrutura de correlação entre fatores não observados resultante de resistência multilateral pode produzir parâmetros tendenciosos¹⁵.

Nesse trabalho, procura-se contornar a primeira questão a partir do uso de regressões *quantílicas*. Em estudo recente, Figueiredo, Lima e Schaur (2014) demonstram que tanto o modelo multiplicativo não linear (8), quanto a versão log-linear (9), podem ser identificadas a partir de *quantis*. Essa estratégia é assegurada

15. Conforme ressaltado por Bertoli e Fernández-Huertas Moraga (2013), o potencial problema de endogeneidade entre e as covariadas observadas no modelo gravitacional não pode ser contornado com o uso de variáveis instrumentais, uma vez que a estrutura de correlação entre fatores observados e não observados envolve mais de um regressor.

pela propriedade de equivariância nos *quantis* da distribuição da variável dependente do modelo diante das transformações monótonas da função de regressão. Então, ao contrário dos modelos fundamentados na média condicionada, a identificação dos *quantis* no modelo exponencial é equivalente à sua identificação no modelo log-linear e vice-versa.

No tocante ao problema de identificação gerado pela presença de TRM, procura-se controlar a variabilidade de fatores específicos às regiões de origem e de destino e variantes no tempo. Assim, conforme proposto por Ortega e Peri (2013), r_{ij} em (9) é aproximado por um conjunto de variáveis de controle para efeitos fixos nas seguintes dimensões: (a) variáveis binárias por local de origem do fluxo migratório D_i , (b) *dummies* por local de destino D_j , (c) *dummies* por ano de migração t , (d) variáveis binárias interadas entre local de origem-ano $D_i * t$ e (e) *dummies* interadas entre local de destino-ano $D_j * t$. Essa estrutura de controles permite a otimização de funções *quantílicas* em painéis de dados de curta dimensão temporal (inferiores a 15 períodos) sem incorrer em viés gerado por parâmetros incidentais¹⁶ (Figueiredo, Lima e Orefice, 2016).

Outro problema não menos relevante que aqueles referentes à identificação do modelo gravitacional (9) é a presença de grande quantidade de fluxos migratórios iguais a zero¹⁷ na maioria das bases de dados sobre movimentos populacionais (Ramos; Surinach, 2013; Figueiredo, Lima e Orefice, 2016). Para superar tal dificuldade, este estudo segue a estratégia de Figueiredo, Lima e Orefice (2016), que consideram a grande quantidade de fluxos migratórios iguais a zero como resultado de aproximações estatísticas. Por exemplo, se os dados são coletados em milhões, então será comum atribuir valor zero para todos os fluxos inferiores a 1¹⁸. Destarte, assumam-se que uma variável $z_{ij} = 0$ se $M_{ij} < 1$ e $Z_{ij} = M_{ij}$ se $M_{ij} \geq 1$. Logo, considerando-se o modelo log-linear (9), a definição anterior é equivalente a $\ln(z_{ij}) = \max(0, \ln(M_{ij}))$ ¹⁹. Quando $\ln(z_{ij})$ é igual a zero, então a observação original está sujeita a aproximações estatísticas, ou, $M_{ij} < 1$.

Desde que $\ln(z_{ij}) = \max(0, \ln(M_{ij}))$, a propriedade da equivariância nos *quantis* garante que $Q_\tau[\ln(z_{ij}) | X_{ij}] = \max(0, Q_\tau[\ln(M_{ij}) | X_{ij}]) = \max(0, X_{ij}\beta(\tau))$, em que Q_τ é o *quantil* τ da distribuição de fluxos bilaterais de migração; X_{ij} representa

16. Bertoli e Fernández-Huertias Moraga (2013) adotam o método de estimação para dados em painel com estrutura multifatorial de erros (*common correlated effects estimator* – CCE). Apesar de esse estimador apresentar uma estrutura mais rica para controle de efeitos fixos (*dummies* por par de origem/destino), seus resultados apenas são robustos em painéis de dados com grande dimensão temporal, isto é, ao menos quinze períodos.

17. Esse problema é discutido na literatura de comércio internacional desde o início da década de 1980. Veja, por exemplo, Head e Mayer (2014), seção 5.2.

18. Argumento similar pode ser encontrado em Lull (2013) e Ramos e Surinach (2013), os quais consideram que todos os fluxos migratórios são positivos, porém, alguns podem apresentar valores muito pequenos. Nesse caso, em populações finitas, as migrações iguais a zero podem ocorrer se a probabilidade de um fluxo bilateral entre dois estados for muito pequena.

19. Outros pontos de censura diferentes de $M_{ij} = 1$ podem ser adotados com $\ln z_{ij} = \max(0, \ln(M_{ij}) - \ln(c))$. Neste trabalho, assume-se $c = 1$.

a matriz de covariadas observadas e descritas na equação (9) e $\beta(\tau)$ é o vetor de parâmetros *location-scale*, isto é, as covariadas X_{ij} afetam não somente a localização da distribuição condicionada de $\ln(z_i)$, mas também a escala de sua dispersão²⁰. Portanto, será considerado o seguinte modelo de regressão *quantílica* censurada com controles para efeitos fixos variantes e invariantes no tempo (TRM):

$$Q_{\tau}[\ln(M_{ij}) | X_{ij}, D_b, D_j, t] = X_{ij}\beta(\tau) + \alpha_1' t + \alpha_2' D_i + \alpha_3' D_j + \alpha_4'(D_i * t) + \alpha_5'(D_j * t) \quad (10)$$

em que $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ e α_5 são vetores de parâmetros.

De acordo com Powell (1984, 1986), o método de regressão *quantílica* censurada fornece um caminho para inferir os modelos *Tobin-Amemiya* sem suposições relativas à distribuição do termo aleatório ou sobre uma possível forma funcional da heterocedasticidade. O estimador de *Powell* é definido a partir da maximização da função objetivo:

$$\begin{aligned} Ln(\beta) = & - \sum_{i,j=1} \omega_{ij} \rho_{\tau}[\ln(M_{ij}) - \max(0, X_{ij}\beta(\tau) + \alpha_1' t + \alpha_2' D_i + \alpha_3' D_j \\ & + \alpha_4'(D_i * t) + \alpha_5'(D_j * t))], \end{aligned} \quad (11)$$

em que ρ_{τ} representa a tradicional função perda da regressão *quantílica* e ω_{ij} é um ponderador. Chernozhukov e Hong (2003) demonstram que o estimador extremo representado por (11) possui uma série de problemas de otimização oriundos da sua não convexidade. Uma solução robusta para otimizar essa função também é fornecida por Chernozhukov e Hong (2003). Em suma, os autores demonstram que o método de *Markov Chain Monte Carlo* (MCMC) pode ser aplicado a vários problemas de inferência estatística, inclusive aqueles que possuam funções objetivo pseudo-quadráticas²¹, conforme o caso em destaque.

Em linhas gerais, como mostrado em Figueiredo, Lima e Schaur (2014), o modelo *quantílico* censurado possui, ao menos, quatro vantagens: *i*) ele acomoda muitos dados de fluxos bilaterais iguais a zero; *ii*) os parâmetros estimados podem ser interpretados como elasticidades e, quando analisados ao longo dos *quantis* condicionados, revelam a heterogeneidade do impacto das covariadas sobre a variável dependente; *iii*) é robusto frente à heterocedasticidade; e *iv*) não necessita de hipóteses relacionadas à distribuição do erro aleatório.

20. Para detalhes, ver Koenker (2005).

21. Alternativamente, seria possível considerar o estimador de três estágios proposto por Chernozhukov e Hong (2002). Para detalhes sobre o método de estimação MCMC, vide Baker (2013).

3.3 Base de dados e seleção amostral

Os dados de migrações empregados neste estudo são oriundos da Pesquisa Nacional por Amostra de Domicílios (Pnad) abrangendo o período de 2004 a 2009. A Pnad é uma pesquisa anual realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), que engloba todas as unidades federativas brasileiras (incluindo o Distrito Federal) a partir de uma amostragem complexa²². Seus quesitos abrangem não apenas características pessoais e domiciliares dos entrevistados, como também migração, ocupação e rendimento. Não obstante, é preciso reconhecer que os dados das PNADs, assim como os dados dos censos demográficos brasileiros, registram maior precisão quando se pretende produzir estimativas de estoques de migrantes ou de taxas de migração por naturalidade ou data fixa.

No intuito de se obter uma estimativa razoável dos fluxos migratórios interestaduais (variável dependente no modelo empírico (9)), a estratégia adotada neste trabalho foi o cruzamento de informações sobre o estado de residência do indivíduo na data de entrevista, tempo de residência e estado de residência anterior. Primeiro, foram consideradas apenas pessoas entre 20 e 65 anos de idade, chefes de domicílio e que não frequentavam nenhum curso. Tal recorte teve o intuito de levar em conta apenas as pessoas que efetivamente teriam condições de decidir sobre a possibilidade de migração tendo em vista oportunidades de trabalho, excluindo-se, portanto, agregados e estudantes (Fiess; Verner, 2003). Em seguida, o migrante interestadual foi identificado como o indivíduo que residia há no máximo um ano no estado informado na data da entrevista, mas que anteriormente morava em outro estado. Por fim, para cada estado e período (2005 a 2009), foi contabilizado o total migrantes segundo a unidade federativa de residência anterior e o estado de residência na data da pesquisa, considerando a ponderação por pesos amostrais²³.

A tabela 2 registra as estatísticas descritivas das variáveis selecionadas para análise empírica. Os dados mostram que, de um total de 3.510 pares de fluxos migratórios de origem-destino identificados no período de 2005 a 2009 (702 pares por ano), 1.905 (54,3%) são fluxos zero. Em comparação com outros estudos internacionais, o percentual de fluxos de migração zero pode ser considerado relativamente baixo (Simpson; Sparber, 2012), porém, ainda pode significar uma fonte importante de viés para análises empíricas. Portanto, cabe ressaltar a importância do método de regressão

22. Apenas a partir de 2004, a Pnad passou a entrevistar moradores da zona rural na região Norte do Brasil. Destarte, para assegurar o maior número de informações possíveis na contagem dos fluxos migratórios interestaduais e comparações consistentes no tempo, este estudo desconsiderou as Pnads de 2001, 2002 e 2003 referentes à década de 2000.

23. Foram estimados fluxos anuais de migração interestadual para cada par de estado de origem (estado de residência anterior) e estado de destino (estado que o indivíduo residia há no máximo um ano da data da entrevista). Note-se que ao se controlar para um curto tempo de residência (até um ano), essa aproximação por fluxos tende a eliminar da amostra possíveis migrantes sobreviventes entre entrevistas de distintas Pnads. Destarte, para cada período de entrevista (2005, 2006, 2007, 2008 e 2009), contabilizaram-se distintas coortes correspondentes aos seguintes períodos de migração 2004-2005, 2005-2006, 2006-2007, 2007-2008 e 2008-2009. Os pesos amostrais foram recentemente atualizados pelo IBGE e consideram estimativas da população dos municípios, utilizando a tendência de crescimento entre 2000-2010.

quantilica censurada empregado neste trabalho, sobretudo, porque o mesmo permite obter estimativas robustas à presença de grande número de fluxos zero conforme já reportado.

Os fluxos de migração interestadual também foram contabilizados considerando duas faixas etárias, migrantes com idade entre 20 e 35 anos (jovens) e migrantes entre 36 e 65 anos (adultos) e três faixas de instrução, migrantes entre 0 e 4 anos de estudo, migrantes entre 5 e 10 anos e migrantes com 11 ou mais anos completos de estudo. Essa estratégia busca a realização de teses de robustez, uma vez que na literatura econômica sobre migrações a questão da autosseleção amostral dos migrantes é bem documentada (Borjas, 1987; Chiswick, 1999; Dos Santos Júnior; Menezes-Filho; Ferreira, 2005). A abordagem do capital humano considera a migração como uma forma de investimento em habilidades e conhecimentos (Schultz, 1961; Sjaastad, 1962). Assim, espera-se que pessoas mais jovens e mais escolarizadas sejam mais sensíveis às diferenças de oportunidades econômicas entre regiões, uma vez que podem antecipar o fluxo de retorno financeiro do investimento em migração e são mais predispostas ao risco.

TABELA 2
Estatísticas descritivas da amostra por pares de origem-destino

	Observações	Média	Desvio-padrão	Mínimo	Máximo
Fluxos migratórios bilaterais					
Total (amostra selecionada)*	1.605	1.741,73	2.759,85	74	30.621
Total (amostra selecionada)	3.510	796,43	2.057,86	0	30.621
De 20 a 35 anos de idade*	1.174	1.265,60	1.682,31	74	16.333
De 20 a 35 anos de idade	3.510	423,31	1.141,37	0	16.333
De 36 a 65 anos de idade*	1.141	1.147,81	1.551,52	79	15.205
De 36 a 65 anos de idade	3.510	373,12	1.034,98	0	15.205
De 0 a 4 anos de estudo*	767	1.061,70	1.292,17	112	11.345
De 0 a 4 anos de estudo	3.510	232,00	746,35	0	11.345
De 5 a 10 anos de estudo*	823	1.023,08	1.154,66	74	9.190
De 5 a 10 anos de estudo	3.510	239,88	707,28	0	9.190
11 ou mais anos de estudo*	1.135	1.003,66	1.344,18	79	13.440
11 ou mais anos de estudo	3.510	324,54	896,87	0	13.440
Covariadas					
População	3.510	6.954.787,00	8.201.516,00	395.180	41.742.009
Distância (Km)	3.510	1.545,31	777,15	80	3.730
Fronteira	3.510	0,1453	0,3525	0	1
Estoque de migrantes (fluxo bilateral) – 2004	3.510	40.960,79	145.452,30	0	2.048.417
Estoque de migrantes (demais fluxos) – 2004	3.510	1.024.020,00	1.002.157,00	15.302	4.065.888
Razão de salário esperado <i>ji</i>	3.510	1,0600	0,2862	0,5221	3,0227

(Continua)

(Continuação)

	Observações	Média	Desvio-padrão	Mínimo	Máximo
Componentes de rendimentos esperados					
Rendimento do trabalho (R\$ de 2008)					
Dirigentes	3.510	1.684,37	293,93	1.068,97	2.533,29
Profissionais das ciências e das artes	3.510	1.319,52	233,76	801,88	2.139,50
Técnicos de nível médio	3.510	758,83	147,12	481,28	1.072,51
Trabalhadores de serviços administrativos	3.510	544,93	66,89	400,61	706,77
Trabalhadores dos serviços	3.510	347,24	62,48	208,10	483,96
Vendedores/prestadores de serviço/comércio	3.510	393,10	89,23	203,64	583,09
Trabalhadores agrícolas	3.510	174,61	70,61	56,91	423,02
Trab. da produção de bens e serviços	3.510	482,57	102,09	277,06	718,74
Membros das forças armadas e auxiliares	3.510	1.395,86	269,91	687,40	2.203,69
Proporção de trabalhadores					
Dirigentes	3.510	0,0454	0,0161	0,0170	0,0969
Profissionais das ciências e das artes	3.510	0,0581	0,0232	0,0253	0,1552
Técnicos de nível médio	3.510	0,0694	0,0154	0,0429	0,1227
Trabalhadores de serviços administrativos	3.510	0,0737	0,0218	0,0354	0,1477
Trabalhadores dos serviços	3.510	0,1936	0,0307	0,1243	0,2704
Vendedores/prestadores de serviço/comércio	3.510	0,1006	0,0178	0,0706	0,1610
Trabalhadores agrícolas	3.510	0,2384	0,1115	0,0071	0,5010
Trab. da produção de bens e serviços	3.510	0,2113	0,0407	0,1331	0,3002
Membros das forças armadas e auxiliares	3.510	0,0097	0,0072	0,0009	0,0386

Fonte: Elaboração própria.

Nota: *Amostra não censurada (fluxos migratórios positivos). A variável rendimento do trabalho refere-se à remuneração do trabalho principal por grupo de ocupação (média por estado). Valores atualizados para preços constantes de 2008 pelo Índice de Preços ao Consumidor Amplo (IPCA).

Entre os fluxos migratórios envolvendo pessoas mais jovens (20 a 35 anos de idade), 66,5% representam fluxos zero, contra 67,5% observado entre os adultos (pessoas entre 36 e 65 anos). Considerando os fluxos positivos (não censurados), os dados da tabela 2 mostram movimento médio anual de 1.266 jovens e de 1.148 adultos por pares de estados entre 2005-2009. Observando os dados de migrações por faixa de instrução, a maior presença de fluxos zero se acha entre os migrantes de baixa escolaridade 78,1% (de 0 a 4 anos de estudo), seguida por 76,5% entre os migrantes na faixa de 5 a 10 anos de estudo e 67,7% entre aqueles com onze ou mais anos completos de estudo. As médias de migrantes observados por pares de estado de origem e destino e, segundo a instrução, são bastante próximas, 1.062 migrantes entre 0 e 4 anos de estudo, 1.023 entre 5 e 10 anos e 1.004 com onze ou mais anos de estudo.

Outras variáveis foram coletadas e calculadas na tentativa de discriminar os fluxos populacionais entre os estados brasileiros com base na especificação do modelo empírico (9). A escolha destas também teve respaldo em outros estudos empíricos na literatura especializada (Clark; Hatton; Williamson, 2007; Lewer; Van den Berg, 2008; Mayda, 2010; Simpson; Sparber, 2012). O quadro 1 do apêndice, apresenta uma descrição detalhada das variáveis empregadas na análise empírica, assim como, o padrão esperado de correlação dos fluxos migratórios (variável dependente) com as covariadas selecionadas. Por outro lado, a tabela 2 também registra as estatísticas descritivas das referidas variáveis.

No conjunto de variáveis explicativas selecionadas, pode-se destacar, inicialmente, aquelas invariantes no tempo e que representam importantes forças de atrito relacionadas aos custos de migração (custo de viagens, custo de hospedagem, custo de aquisição de informações, custo de procura por trabalho):

- *distância entre capitais*. Trata-se da distância geodésica em quilômetros (considerando a curvatura da Terra) entre os pontos centrais das capitais dos estados de origem e de destino dos fluxos migratórios. Essa variável é tradicionalmente empregada em modelos gravitacionais como uma *proxy* para custos de transporte. Em estudos aplicados à migração do trabalho, geralmente representa custos observados: custos de viagens e/ou de acesso à informação e custos não observados: custos psicológicos por separação de entes familiares. Quanto maior a distância entre os locais, maior deve ser o custo de deslocamento, e, por conseguinte, menor o total de migrantes (Sjaastad, 1962; Sahota, 1968; Clark; Hatton; Williamson, 2007);
- *compartilhamento de fronteiras*. É uma variável binária que recebe o valor um, se o estado de origem e de destino de fluxos migratórios compartilham fronteiras, e zero, caso contrário. Essa variável é uma *proxy* para efeitos de vizinhança/proximidade e reforça o controle para custos associados às migrações a partir de uma métrica de compartilhamento de fronteiras. Em estados vizinhos, menor deve ser o custo de aquisição de informações sobre postos de trabalho, custos de aluguel e custos de viagens, por exemplo. Logo, espera-se que os movimentos populacionais sejam mais intensos entre regiões que compartilham fronteiras;
- *estoque de migrantes acumulados, segundo fluxos bilaterais*. É uma variável que considera o total de migrantes observados em 2004 na unidade federativa de destino dos fluxos migratórios contabilizados entre 2005-2009, mas que nasceram na unidade federativa de partida desses últimos fluxos. Essa variável é uma *proxy* para capturar efeitos de redes sociais na orientação dos fluxos de migração e redução de custos envolvidos

na mobilidade geográfica. Espera-se que a conexão entre o histórico de migração passada e os fluxos correntes por regiões possa favorecer o acesso à informação, suporte logístico e financeiro, facilitação de moradia e procura por emprego, estimulando os movimentos populacionais entre estados com maior presença de redes migratórias (Vogler; Rotte, 2000; Clark; Hatton; Williamson, 2007; Mayda, 2010);

- *estoque de migrantes acumulados segundo fluxos multilaterais*. No intuito de assegurar maior precisão na captura de efeitos de redes sociais, foi calculado o total de migrantes, observados em 2004, que nasceram na mesma unidade federativa de partida dos fluxos migratórios contabilizados entre 2005-2009, mas que se dirigiram a todos os estados potenciais de destino (exceto o estado de destino do fluxo bilateral). Essa variável procura controlar efeitos passados de tendências de emissão de migrantes para quaisquer lugares; característica que poderia mascarar possíveis efeitos de redes sociais caso fosse omitida da análise empírica.

Os dados da amostra selecionada na tabela 2 indicam que, em média, em 14,5% dos fluxos migratórios observados entre 2005-2009, os estados de origem e de destino compartilham fronteiras e registram uma distância de 1.545,3 quilômetros. Por outro lado, em média, cerca de 40.961 migrantes acumulados em 2004 estão associados às mesmas rotas bilaterais dos fluxos migratórios analisados entre 2005 e 2009, sendo que pouco mais de um milhão de migrantes nascidos na mesma unidade federativa de partida dos fluxos mais recentes, dirigiram-se para outros estados que não o estado de destino dos fluxos bilaterais.

Outras três covariadas foram selecionadas/construídas considerando um ano de defasagem em relação à contagem dos fluxos migratórios bilaterais²⁴. Elas buscam ponderar condições econômicas das unidades federativas de origem e de destino dos migrantes interestaduais:

- *população segundo o estado de partida*. Refere-se ao total da população residente no estado de moradia anterior dos migrantes interestaduais. Essa variável procura capturar a capacidade de emissão populacional e espera-se uma relação direta entre ela e o total de migrantes direcionados para outros estados;
- *população segundo o estado de residência*. É o total da população residente no estado de destino dos fluxos migratórios. Procura ponderar possíveis

24. Essa estratégia foi adotada na tentativa de minimizar possíveis problemas de causalidade reversa entre as covariadas em destaque e a variável dependente na análise empírica: fluxos de migração bilateral nos períodos de migração 2004-2005, 2005-2006, 2007-2008 e 2008-2009. Assim, adota-se o pressuposto de exogeneidade fraca com base em covariadas predeterminadas diante da observação dos fluxos migratórios.

efeitos de forças de aglomerações que podem atuar tanto na atração, como na dispersão de migrantes;

- *razão de renda esperada (renda relativa)*. Considerando a literatura econômica sobre migrações, essa é uma das principais variáveis determinantes da decisão de migrar (SJAASTAD, 1962; BORJAS, 1994). Conforme discuto na seção 3.1, a decisão individual de migrar envolve a comparação de rendimentos esperados do trabalho entre o lugar de residência e o local de destino. No entanto, esse diferencial é previamente condicionado a importantes determinantes salariais, isto é, as diferenças por ocupação/setor e a probabilidade de emprego. Neste trabalho, a referida variável foi calculada de forma a condicionar a diferença interestadual de renda aos hiatos observados por grupos ocupacionais e ao risco relativo de desemprego por ocupação. Apesar de Justo e Silveira Neto (2006) terem incorporado a variável renda esperada no estudo dos determinantes das migrações interestaduais no Brasil, o condicionamento das desigualdades de rendimentos e do risco de desemprego aos grupos de ocupação não têm sido uma estratégia comum na literatura especializada. De forma mais específica, neste estudo foram considerados nove grupos ocupacionais conforme categorias de trabalho principal na semana de referência elencadas pelo IBGE: *i)* dirigentes em geral; *ii)* profissionais das ciências e das artes; *iii)* técnicos de nível médio; *iv)* trabalhadores de serviços administrativos; *v)* trabalhadores dos serviços; *vi)* vendedores e prestadores de serviço do comércio; *vii)* trabalhadores da produção de bens e serviços e de reparação e manutenção; e *viii)* membros das forças armadas e auxiliares²⁵. Após considerar a atualização das rendas individuais (rendimento do trabalho principal) para preços constantes de 2008²⁶, foi calculada a média estadual de rendimento do trabalho²⁷ e a proporção de trabalhadores por grupo ocupacional. Em seguida, computou-se a razão de renda esperada (rendimento médio do trabalho ponderado pela probabilidade de desemprego²⁸) segundo o estado de destino e de partida dos fluxos migratórios e por categoria de ocupação. A expressão

25. Foram desconsideradas pessoas com ocupações mal-definidas ou não declaradas.

26. Essa atualização foi feita com base no índice de preços ao consumidor amplo (IPCA) do IBGE. Em razão da indisponibilidade de índices de custo de vida específicos a cada estado brasileiro, não foi realizado nenhum ajuste para diferenças regionais de custo de vida.

27. Antes do cálculo das médias de renda por estado e ocupação, foram desconsideradas pessoas com informações aberrantes (*outliers*). Os dados aberrantes foram identificados pela regra do *boxplot*, isto é, valores fora do intervalo $(q_1 - 1,5 [q_3 - q_1], q_3 + 1,5 [q_3 - q_1])$, em que q_1 representa o valor do primeiro quartil da distribuição de rendimentos entre indivíduos, q_3 é o terceiro quartil e $q_3 - q_1$ mensura a distância interquartilica.

28. Seguindo estratégia semelhante a de Justo e Silveira Neto (2006), a probabilidade de desemprego foi estimada por um menos a proporção de trabalhadores por categoria de ocupação.

a seguir resume o cálculo da razão de renda esperada entre o estado de destino e o estado de origem dos fluxos migratórios. Ela é uma média aritmética das razões de renda esperada por grupo de ocupação considerando cada par de estado de origem e de destino de migrantes interestaduais:

$$RRE_{ji} = \sum_{k=1}^9 \frac{\omega_{jk}(1 - p_{jk})}{\omega_{ik}(1 - p_{ik})}, \quad (12)$$

em que ω_{jk} é a média de rendimento do trabalho da ocupação k no estado de destino j ; ω_{ik} é a média de renda da ocupação k no estado de origem i ; p_{jk} e p_{ik} são as proporções de trabalhadores ocupados na categoria nos estados i e j .

A tabela 2 apresenta médias da população residente e das razões de renda esperada por pares de estado de origem e de destino dos fluxos migratórios interestaduais entre 2005-2009. Cada par de fluxo bilateral está associado a uma média de 6,9 milhões de habitantes por estados de origem e de destino. Em média, os estados de destino apresentaram um rendimento esperado do trabalho 0,6% superior às unidades federativas de origem dos migrantes, sugerindo que a decisão de migrar no Brasil parece ser consistente com a busca de melhores oportunidades de emprego e renda. Vale ainda observar que entre as ocupações consideradas no cálculo da razão esperada de renda entre estados, os dados mostram que apesar de os trabalhadores dirigentes, membros das forças armadas e profissionais das ciências e das artes, registrarem médias salariais bastante superiores àquelas observadas para as demais categorias ocupacionais, também apresentam menor probabilidade de ocupação. Os grupos de ocupação que envolvem atividades agrícolas, produção de bens e serviços, serviços e comércio são aquelas com maiores chances de emprego. No entanto, os trabalhadores agrícolas têm menor média salarial R\$ 174,6, contra R\$ 482,6 dos empregados na produção de bens e serviços.

A figura 2 apresenta gráficos de dispersão entre fluxos bilaterais de migração estadual de 2005-2009 e covariadas selecionadas. Em geral, os padrões de dispersão observados corroboram as expectativas teóricas iniciais. Observando-se as figuras 2A e 2B, por exemplo, parece haver uma associação direta entre os fluxos de migração e as populações dos estados de partida e de destino. Na figura 2C, os dados indicam uma relação inversa entre fluxos bilaterais e distâncias entre os estados de origem e de destino. Os dados da figura 2D não registram nenhum padrão regular de relação linear ou não linear que permita conjecturar sobre alguma associação prévia entre fluxos de migrantes e diferenças estaduais de renda. Contudo, é preciso reconhecer que os rendimentos médios de ocupações por estados são determinados por um conjunto de fatores que, em geral, não são diretamente observados e que, até esse

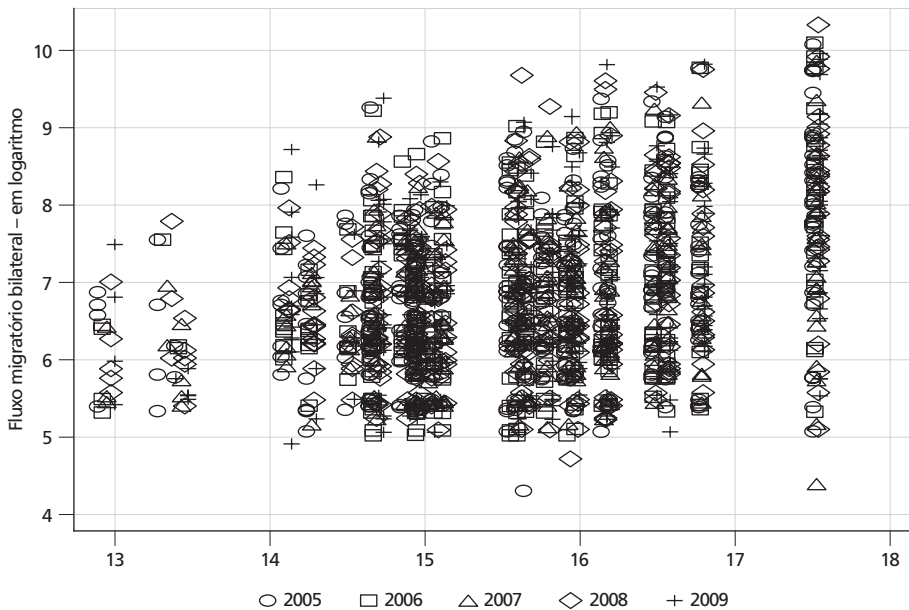
estágio, não foram ponderados na análise estatística. Já a figura 2E permite uma primeira exploração da relação entre fluxos migratórios interestaduais e estoque de migrantes acumulados até o ano de 2004 (período precedente aos dados de migração explorados neste estudo). Apesar de não se controlar a emissão total de migrantes de um determinado estado para todos os demais, os dados indicam uma relação direta entre os fluxos anuais de migração interestadual e o estoque inicial de migrantes acumulados. Ou seja, estados com residentes naturais de outras unidades federativas tendem a receber mais imigrantes partidos dessas últimas com o passar do tempo. Esse padrão é consistente com os efeitos esperados das redes sociais sobre o direcionamento dos migrantes e redução de custos envolvidos na arbitragem (Mckenzie; Rapoport, 2007; Yamauchi; Tanabe, 2008).

Na próxima seção, são apresentados os resultados econométricos deste trabalho, em que se estima a equação gravitacional (10) a partir de diferentes especificações e controles para efeitos fixos (termos de resistência multilateral). O intuito é verificar se as correlações discutidas nesta seção são válidas ou não enquanto evidências empíricas para o caso das migrações interestaduais no Brasil.

GRÁFICO 2

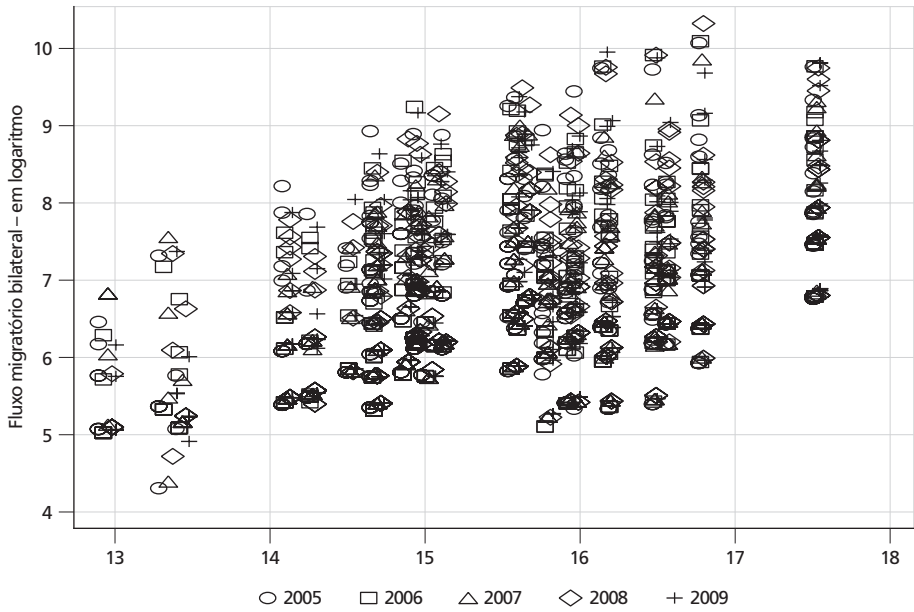
População do estado de origem (t-1) – em logaritmo

2A



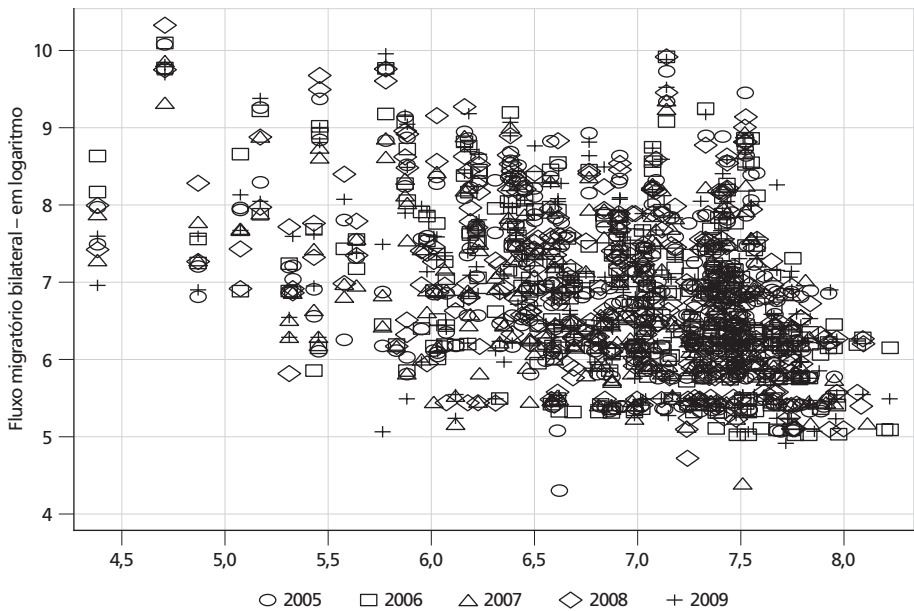
População do estado de destino (t-1) – em logaritmo

2B



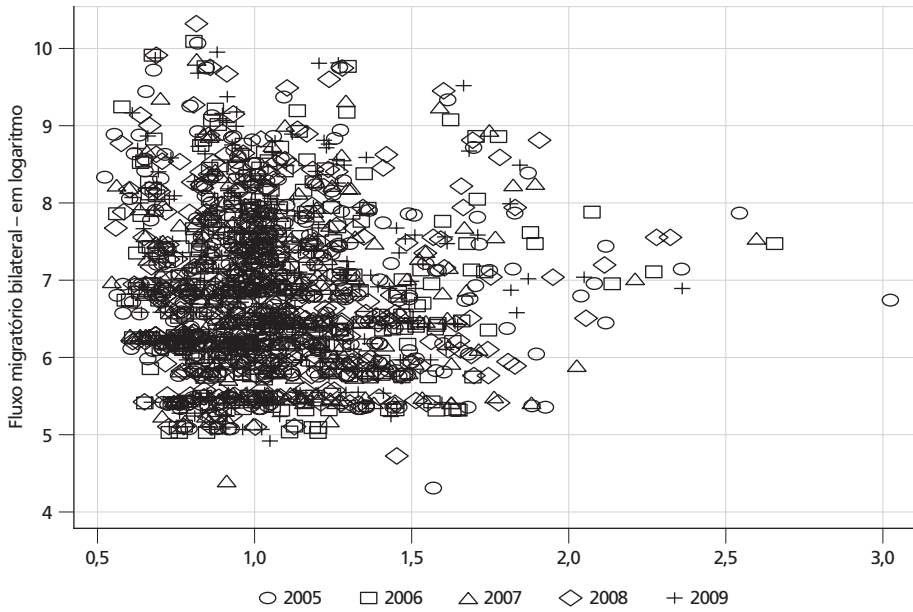
Distância entre estados de origem e destino – em logaritmo

2C



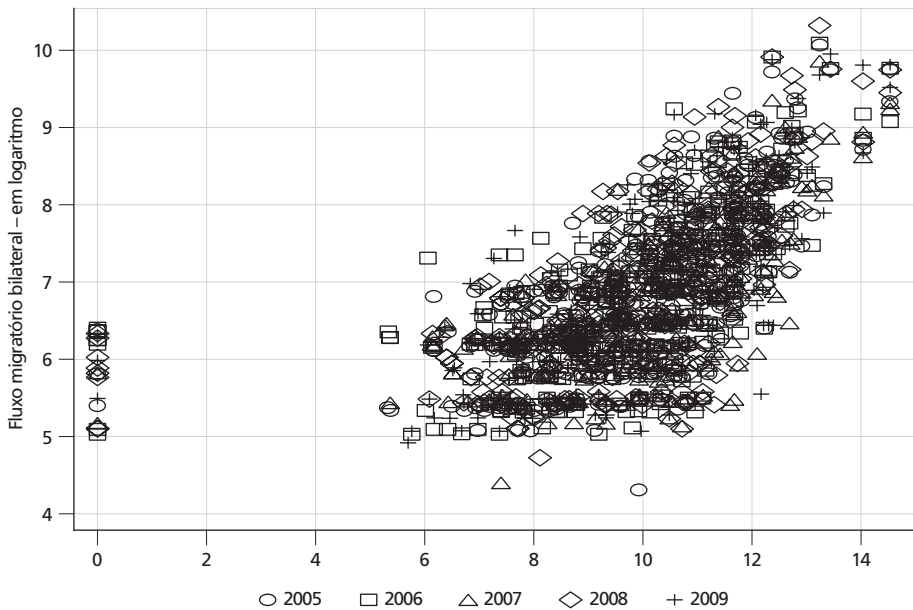
Razão de renda esperada entre estado de destino e estado de origem (t-1) – em logaritmo

2D



Estoque de migrantes no estado de destino nascidos no estado de origem – 2004 – em logaritmo

2E



4 RESULTADOS

4.1 Determinantes dos fluxos migratórios interestaduais

A tabela 3 registra evidências sobre os principais determinantes dos fluxos migratórios interestaduais no Brasil. São apresentados os coeficientes estimados a partir de regressões *quantílicas* censuradas (na mediana) para o modelo (9), considerando diferentes controles para efeitos fixos variantes e invariantes no tempo, respectivamente nas colunas 1, 2, 3, 4 e 5. A regressão presente na coluna 1 inclui apenas variáveis binárias para cada período do painel de dados (2005-2009). Na regressão da coluna 2, foram acrescentadas variáveis binárias para cada unidade federativa de origem dos fluxos migratórios. A regressão registrada na coluna 3 inclui *dummies* para os estados de destino, enquanto as regressões das colunas 4 e 5 foram adicionadas variáveis binárias de interação entre estado de origem e tempo e estado de destino e tempo, respectivamente. Conforme discutido por Ortega e Peri (2013) e Beine, Bertoli e Fernández-Huertas Moraga (2014), tais variáveis de controle buscam capturar efeitos de fatores específicos às regiões e variantes no tempo, os quais representam forças de atratividade/obstrução que várias localidades podem exercer sobre os fluxos bilaterais de migração (*proxies* para efeitos não observados de resistência multilateral).

Os dados da tabela 3 permitem observar que os coeficientes estimados são consoantes com a maioria das correlações entre variáveis selecionadas já apresentadas na seção anterior. Primeiro, um aumento da população do estado de origem tende a aumentar o fluxo de emigrantes para qualquer região de forma mais que proporcional. Por exemplo, um aumento de 1% da população residente em estado típico tende a elevar o total de emigrantes em cerca de 4,5% no próximo ano (coluna 5). Essa evidência é bastante regular, uma vez que o coeficiente de elasticidade da variável população do estado de origem mostrou-se estatisticamente significativo e com sinal positivo em todas as regressões em destaque.

A relação entre fluxos migratórios bilaterais e população dos estados de destino apenas revela-se positiva quando se controlam características não observadas por local de partida (coluna 2). No entanto, nas regressões apresentadas nas colunas 3 e 4, há variáveis binárias por estado de destino e de interação entre estado de origem e ano. Nesses casos, a associação entre fluxos de migração e população do local de destino torna-se inversa, indicando que um incremento de 1% da população de um estado típico de destino pode reduzir o total de imigrantes oriundos de quaisquer estados em cerca de 0,8% no período seguinte (coluna 4). Vale ainda observar que essa relação não é estatisticamente significativa quando se considera um controle mais amplo para possíveis efeitos de resistência multilateral (coluna 5).

As evidências sobre a associação dos fluxos migratórios com as distâncias estaduais no Brasil mostram-se bem sugestivas e estão em linha com os achados de estudos anteriores (Sahota, 1968). Em todas as regressões registradas na tabela 3, o coeficiente de elasticidade da variável distância apresenta sinal negativo. Após aumentar o leque de variáveis binárias de controle, observa-se um aumento em valor absoluto do referido coeficiente. Por exemplo, um aumento de 10% na distância entre duas capitais estaduais pode diminuir o fluxo de migrantes entre os mesmos estados em cerca de 12,1%. Ou seja, considerando dados observados de fluxos migratórios e de distâncias na mediana de suas distribuições, cerca de 114 quilômetros a mais de distância entre dois estados pode resultar em 99 migrantes/ano a menos no fluxo de migração bilateral.

A despeito de a distância bilateral ser uma *proxy* bastante usada para custos de transporte, os possíveis efeitos dos diversos custos associados à migração também podem ser ponderados com base em outra métrica de proximidade, como é o caso do compartilhamento de fronteiras (efeito vizinhança). Os resultados deste estudo corroboram aqueles obtidos por Justo e Silveira Neto (2006) ao mostrarem que estados com fronteiras em comum apresentam maiores fluxos bilaterais de migrantes. Os dados da tabela 3 revelam que em todas as regressões o coeficiente da variável fronteira mostra-se positivo. Considerando a regressão da coluna 5, que inclui maior controle para fatores não observados aos estados, percebe-se que entre estados vizinhos o fluxo bilateral de migração é cerca de cinco vezes maior²⁹ que aqueles observados entre unidades federativas sem compartilhamento de fronteiras. Ou seja, na mediana, pode-se estimar que estados vizinhos têm fluxos bilaterais/anuais com cerca de 4.279 migrantes a mais, relativamente às unidades federativas não vizinhas.

TABELA 3

Regressões *quantílicas* censuradas (equações gravitacionais estimadas na mediana): variável dependente – fluxo de migração interestadual entre estados de origem e destino (Em logaritmo)

	(1)	(2)	(3)	(4)	(5)
Log(população) <i>i</i> (t-1)	1,5887*** (0,1641)	4,7772*** (0,1148)	3,5583*** (0,0966)	4,8350*** (0,0655)	4,5387*** (0,0789)
Log(população) <i>j</i> (t-1)	0,1123 (0,1254)	0,2601** (0,1320)	-1,3463*** (0,1681)	-0,8799*** (0,0832)	-0,1164 (0,0928)
Log(distância) <i>ij</i>	-0,9912*** (0,2088)	-0,9556*** (0,1267)	-1,1877*** (0,1144)	-1,0293*** (0,1036)	-1,2121*** (0,1210)

(Continua)

29. Considerar o cálculo do coeficiente de elasticidade por exponencial natural $(1, 8264) - 1$ na forma log-linear do modelo de regressão.

(Continuação)

	(1)	(2)	(3)	(4)	(5)
Fronteira ij	1,2159** (0,5305)	1,4097*** (0,1526)	1,3331*** (0,1458)	1,6696*** (0,1513)	1,8264*** (0,1062)
Renda relativa ji (t-1)	0,5437*** (0,2016)	0,4004*** (0,1521)	3,5882*** (0,1000)	4,4476*** (0,0742)	6,2554*** (0,0593)
Log(migrantes) ij 2004	1,2743*** (0,2376)	1,0293*** (0,2575)	0,8681*** (0,1831)	0,8736*** (0,1155)	0,8125*** (0,1833)
Log(migrantes) ik 2004	-1,0913*** (0,0825)	-3,3860*** (0,0840)	-4,3741*** (0,0433)	-4,2442*** (0,0409)	-3,8795*** (0,0354)
Intercepto	-13,7358*** (0,3073)	-31,4545*** (0,0740)	20,7731*** (0,0479)	-7,3700*** (0,0397)	-18,7154*** (0,0622)
Observações	3.510	3.510	3.510	3.510	3.510
Efeitos fixos					
Ano	sim	sim	sim	sim	sim
UF de origem	não	sim	sim	sim	sim
UF de destino	não	não	sim	sim	sim
UF de origem/ano	não	não	não	sim	sim
UF de destino/ano	não	não	não	não	sim

Fonte: Elaboração própria.

Notas: i – indexa o estado de partida do fluxo migratório; j indexa o estado de destino dos migrantes; k indexa todos os estados de destino, exceto j ; $t-1$ indica que a variável refere-se ao período anterior à contagem dos fluxos migratórios. *** Estatisticamente significativo a 1%. ** Estatisticamente significativo a 5%. * Estatisticamente significativo a 10%. Erros-padrão robustos à heterocedasticidade entre parênteses. Resultados estimados a partir do algoritmo MCMC desenvolvido por Baker (2013).

Os resultados econométricos também mostram a importância das diferenças estaduais de renda condicionadas a diferentes ocupações e riscos de desemprego. Essas dimensões são incorporadas na variável razão de renda esperada, que, conforme discutido na seção 3.3, é uma média das razões de renda esperadas (rendimento do trabalho principal ponderado pelo risco de desemprego) por ocupação entre os estados de partida e de destino dos fluxos migratórios anuais. É possível observar que em todas as regressões constantes em análise o coeficiente associado à variável renda relativa é positivo e estatisticamente significativo. Tal evidência indica que na mediana os fluxos migratórios interestaduais no Brasil são dirigidos para estados onde a diferença de renda e o risco de desemprego por ocupação são relativamente menores. As alterações na magnitude do coeficiente da variável renda relativa, a partir da inclusão de controles mais amplos para efeitos fixos, revela a importância de se incorporar os efeitos de atratividade de outros estados sobre a migração bilateral (resistência multilateral), sob pena de forte viés nas estimativas. Por exemplo, enquanto na regressão da coluna 1, que ignora a correlação da renda esperada com fatores não observados por estados, o efeito de uma diferença de um aumento de 10% na razão de renda entre duas unidades federativas pode gerar uma elevação de 5% no fluxo bilateral de migração do ano posterior, na regressão da coluna 5,

que inclui termos de controle para resistência multilateral, esse efeito é de 63% na mediana, ou o equivalente a um incremento 517 migrantes por fluxo bilateral/ano, contra um aumento de apenas 41 na primeira estimativa³⁰. Em linhas gerais, essas evidências corroboram aquelas encontradas por Justo e Silveira Neto (2006).

A presença de muitas pessoas em um estado de residência de migrantes recentes e que no passado emigraram dos mesmos estados desses últimos parece formar conexões importantes para acesso à informação, apoio financeiro, apoio na procura por postos de trabalho, compartilhamento de moradia etc., que, em geral, podem direcionar as migrações interestaduais a partir da redução de custos de migração. Os achados econométricos presentes na tabela 3 indicam que as redes sociais podem desempenhar um papel relevante na determinação da mobilidade geográfica do trabalho no Brasil. Mesmo controlando o estoque de imigrantes de cada estado a partir de pessoas nascidas em um mesmo estado i (pessoas que nasceram em i , mas que em 2004 estavam em qualquer estado exceto j), a relação entre fluxo bilateral/ano de migrantes (pessoas que partiram do estado i para o estado j entre 2005 e 2009) e o estoque de migrantes acumulados nascidos no mesmo estado de partida dos migrantes recentes (pessoas que nasceram no estado i e que em 2004 estavam em j) é direta em todas as regressões. Vale observar que o coeficiente de elasticidade da variável de migração acumulada de i para j em 2004 é sobrestimado nas regressões que não incluem controles para resistência multilateral. Os resultados da regressão presente na coluna 5 com controles para efeitos fixos, por exemplo, sugerem que um incremento de 10% no total de migrantes acumulados no estado j , que nasceram em i , eleva o fluxo bilateral corrente em cerca de 8,1% na mediana. Ou seja, estima-se que para uma rede com cerca de 2.441 migrantes acumulados a mais, pode haver um aumento de 66 migrantes por fluxo de bilateral anual.

É importante ressaltar que os resultados discutidos até esse estágio também podem ser comparados com regressões feitas por MQO e presentes na tabela A.1 do apêndice. Apesar de os sinais dos coeficientes estimados não serem diferentes daqueles já apresentados, notam-se diferenças consideráveis nas magnitudes dos parâmetros quando comparados àqueles constantes na tabela 3. Com efeito, as estimativas por MQO são tendenciosas em razão da grande presença de fluxos zero e da log-linearização do modelo gravitacional (problema de identificação). Ademais, as regressões constantes nas tabelas A.2 e A.3 do apêndice mostram como os coeficientes estimados comportam-se frente a diferentes especificações da equação gravitacional, método de estimação e controles para efeitos fixos. Em linhas gerais, há diferenças na magnitude dos coeficientes à medida que se aumenta o número de covariadas no modelo gravitacional. Conforme os dados da tabela A.2, mesmo sem controlar efeitos fixos, quando se considera o estoque de migrantes acumulados em 2004 (colunas 5-8), a magnitude dos coeficientes associados às

30. Em ambos os casos, considere o coeficiente de elasticidade calculado por βx_m , onde β é o coeficiente estimado e x_m o valor da variável na mediana. No presente caso, a mediana é 821 migrantes.

demais variáveis *proxies* para custos de migração (distância e compartilhamento de fronteiras) diminui em termos absolutos. O coeficiente da variável renda relativa não apresenta um padrão de sinal e nem regularidade estatística ao longo das diferentes especificações na tabela em foco. No entanto, ao se incorporar variáveis binárias para controle da atratividade não observada de todas as unidades federativas (ver tabela A.3), o coeficiente da última variável registra sinal positivo e significância estatística sob várias especificações do modelo de regressão³¹.

A figura 3 resume resultados obtidos com base em várias regressões, considerando diferentes *quantis* da distribuição dos fluxos de migrantes interestaduais e o mesmo conjunto de covariadas e controles para efeitos fixos presentes na coluna 5 da tabela 3. Ou seja, esses resultados permitem avaliar as correlações já discutidas não apenas na mediana da distribuição das migrações, mas ao longo de toda a distribuição.

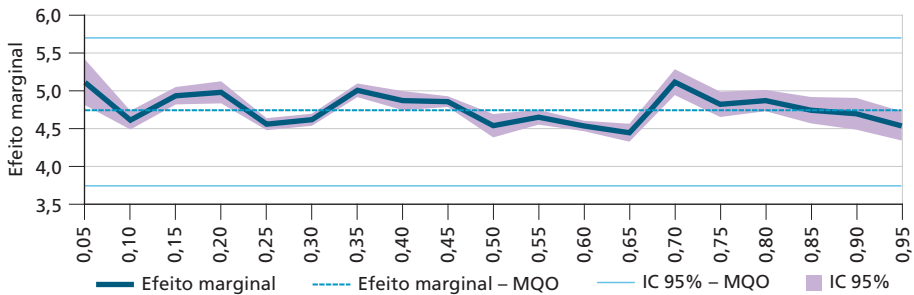
Os resultados mostram que os efeitos marginais/elasticidades têm sinais regulares ao longo de vários *quantis* da distribuição dos fluxos bilaterais de migração, sugerindo que, em geral, as correlações estimadas para a mediana, permanecem válidas em partes extremas da distribuição de migrantes. Por outro lado, é possível constatar algumas diferenças de magnitude dos efeitos marginais em partes extremas da distribuição em destaque. Por exemplo, na figura 3B, a relação entre a população do estado de destino e os fluxos de migrantes é negativa em torno da mediana e, sobretudo, quando se consideram estados que emitem poucos migrantes (*quantis* inferiores). Destarte, o aumento da população de um estado típico de destino tende a diminuir a absorção de imigrantes oriundos de quaisquer outros estados. Acima do *quantil* 0,8, isto é, levando em conta estados com grandes emissões de migrantes, não se pode inferir uma relação positiva em virtude da maior imprecisão no intervalo de confiança.

A relação entre distâncias estaduais e migrações permanece inversa em várias partes da distribuição de fluxos bilaterais de migração (figura 3C). Também percebe-se que a estimativa do coeficiente de elasticidade da variável distância é superestimada pelo método de MQO. Apenas no *quantil* 0,95 verifica-se um efeito marginal menor em termos absolutos, o que provavelmente pode relacionar-se com efeitos de formação de novas redes/fluxo de informações entre estados com elevados fluxos de migrantes. Por sua vez, no tocante ao efeito de vizinhança entre estados, as estimativas para os *quantis* inferiores e ligeiramente superiores à mediana (0,55 a 0,80) revelaram-se menores que aquela obtida por MQO (Figura 3D). No geral, estados que compartilham fronteiras têm fluxos migratórios maiores que aqueles não vizinhos.

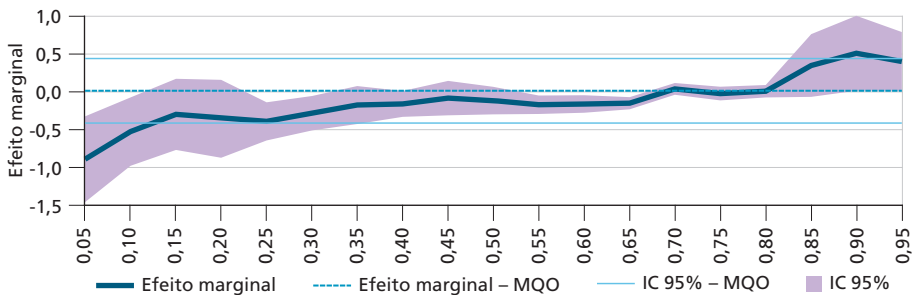
31. Vale ressaltar que outras especificações foram testadas, incluindo taxas de ocupação por setor, razão de renda esperada por ocupação, taxas de homicídios, taxas de desocupação e variáveis climáticas. Devido à pouca variabilidade de muitas dessas variáveis (pouca discriminação dos fluxos de migração) e problemas de multicolinearidade com variáveis teoricamente mais relevantes, os resultados não se mostraram satisfatórios. Portanto, para evitar problemas de precisão na análise dos coeficientes de elasticidades, optou-se por uma especificação parcimoniosa e em linha com estudos recentes na literatura internacional (Mayda, 2010; Ortega, Peri, 2013).

Não obstante a grande amplitude do intervalo de confiança, as estimativas do coeficiente de elasticidade da variável estoque de migrantes acumulados em 2004 por *quantis* são positivas e superam aquela produzida pelo modelo gravitacional estimado por MQO (figura 3F). Os resultados sugerem que os possíveis efeitos de redes sociais ou de lugares parecem ser mais expressivos entre estados com baixos fluxos anuais de migração (*quantis* inferiores) ou próximos à mediana dos fluxos. Entre estados com grandes fluxos bilaterais por ano, o estoque de migrantes acumulados no passado pode favorecer as migrações correntes, mais em menor magnitude.

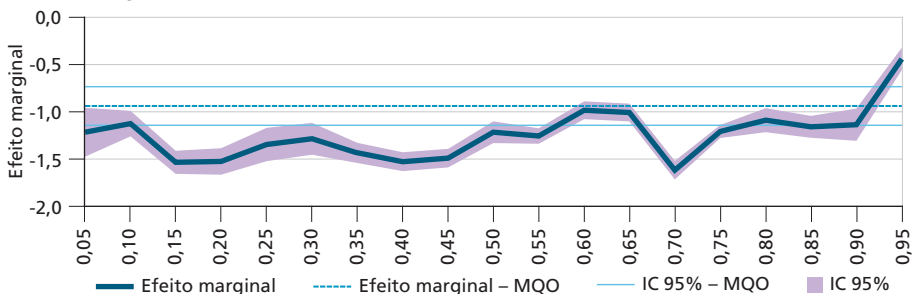
GRÁFICO 3
Quantil da distribuição de fluxos migratórios
 3A – População do estado de origem
 (Em logaritmo)

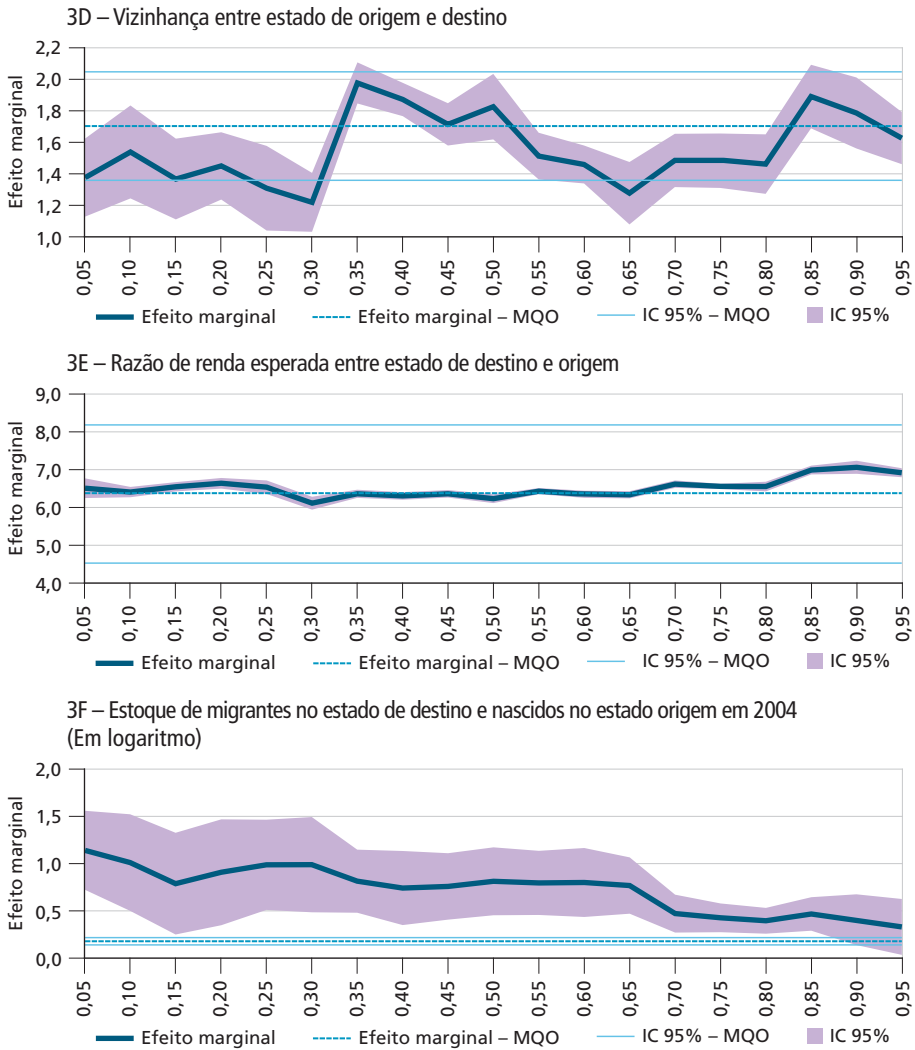


3B – População do estado de destino
 (Em logaritmo)



3C – Distância entre estado de origem e destino
 (Em logaritmo)





As estimativas mais regulares ao longo dos *quantis* foram aquelas obtidas para os coeficientes das variáveis população do estado de origem e renda relativa, respectivamente, nas figuras 3A e 3E. No primeiro caso, os resultados indicam que estados com maior população tendem a emitir mais trabalhadores para outras unidades federativas, a despeito do volume dos fluxos bilaterais de migração. Ainda de acordo com os resultados, as diferenças interestaduais de rendimento esperado por ocupação estão diretamente relacionadas com as migrações internas no Brasil. Essa evidência é robusta em toda a distribuição dos fluxos bilaterais de migração e consistente com a expectativa teórica.

4.2 Seletividade e Robustez

No Brasil, há evidências de que os migrantes interestaduais são trabalhadores positivamente selecionados, isto é, são pessoas diferenciadas em termos de maior predisposição ao risco, maior perseverança, empreendedorismo e entusiasmo (características produtivas não observadas), quando comparados aos não migrantes (Dos Santos Júnior; Menezes-Filho; Ferreira, 2005; Queiroz; Ramalho, 2011; Freguglia; Gonçalves; Silva Ribeiro, 2014). Essas evidências estão em linha com a abordagem teórica do capital humano e com vários estudos internacionais (Borjas, 1987; Chiswick, 1999), que associa tais atributos produtivos a trabalhadores mais jovens e/ou mais instruídos (Sjaastad, 1962).

Seguindo a intuição de Sahota (1968), nessa subseção procura-se verificar se os achados reportados são robustos diante de grupos de migrantes por faixa etária e escolaridade. Para tanto, a tabela 4 registra regressões *quantílicas* para o modelo gravitacional (9) em que a variável dependente (fluxos bilaterais de migração) foi computada considerando diferentes faixas etárias e faixas de instrução. Nas regressões apresentadas nas colunas 1 e 2, os fluxos bilaterais de migração foram calculados considerando trabalhadores entre 20 e 35 anos de idade (jovens) e entre 36 e 65 anos (adultos), respectivamente. Nas colunas 3, 4 e 5, encontram-se os resultados de regressões para migrantes em três faixas de escolaridade, na seguinte sequência: *i*) trabalhadores de 0 a 4 anos completos de estudo; *ii*) trabalhadores de 5 a 10 anos completos de estudo e *iii*) trabalhadores com onze ou mais anos de escolaridade³².

Os resultados mostram diferenças sugestivas entre os coeficientes estimados por grupos de idade e de instrução, indicando que evidências mais críveis sobre os determinantes dos movimentos migratórios devem considerar efeitos de seletividade (não aleatoriedade amostral). Por outro lado, a despeito das diferenças de magnitude imputadas à seletividade dos migrantes interestaduais, os resultados revelam que os sinais dos coeficientes estimados são, em geral, consistentes com aqueles obtidos para a amostra geral de fluxos migratórios.

Em conformidade com a teoria do capital humano, os dados da tabela em destaque apontam que os migrantes jovens são mais sensíveis às diferenças interestaduais de renda por ocupação. Ao se cotejar as regressões das colunas 1 e 2, observa-se que o coeficiente da variável renda relativa para o grupo de migrantes entre 20 e 35 anos de idade é mais que o dobro daquele estimado para os migrantes adultos. Ademais, os migrantes jovens parecem responder com maior intensidade

32. Essa é uma estratégia parcimoniosa, uma vez que o controle dos fluxos migratórios por seleção em variáveis observadas (idade e instrução) pode ser uma condição necessária, mas não suficiente para controlar a heterogeneidade não observada dos migrantes. O método mais apropriado para tratar essa questão seria o uso de instrumentais à la Heckman (1979), considerando um modelo estrutural de determinação conjunta da probabilidade de ocorrência de fluxos migratórios (inclusive fluxos zero) e dos volumes de migração. No entanto, vale observar que tal procedimento torna-se inviável na maior parte dos casos, haja vista a dificuldade de se achar uma variável de exclusão ("instrumento") que determine a ocorrência de fluxos de migração, mas não determine o volume desses fluxos.

ao crescimento do número de migrantes acumulados, ao aumento da população do estado de origem e à proximidade (vizinhança), quando comparados aos migrantes adultos.

TABELA 4

Regressões *quantílicas* censuradas (equações gravitacionais estimadas na mediana): variável dependente – fluxo de migração interestadual entre estados de origem e destino por faixa etária e faixa de instrução
(Em logaritmo)

	(1) Idade: 20 a 35	(2) Idade: 36 a 65	(3) Instrução: 0 a 4	(4) Instrução: 5 a 10	(5) Instrução: 11 anos+
Log(população) <i>i</i> (t-1)	5,2441*** (0,0522)	4,6423*** (0,0528)	6,3015*** (0,0625)	5,7373*** (0,0470)	4,4006*** (0,0534)
Log(população) <i>j</i> (t-1)	-0,1286 (0,1392)	-0,3052** (0,1345)	-0,9845*** (0,2041)	-0,8373*** (0,2080)	0,4132*** (0,1098)
Log(distância) <i>ij</i>	-1,2206*** (0,1188)	-1,2379*** (0,1123)	-0,9098*** (0,1203)	-0,8968*** (0,1227)	-1,3957*** (0,1268)
Fronteira <i>ij</i>	2,3006*** (0,0806)	2,0067*** (0,0543)	2,2250*** (0,1284)	2,1841*** (0,1226)	1,5854*** (0,0649)
Renda relativa <i>ji</i> (t-1)	6,9206*** (0,0568)	3,3447*** (0,0616)	6,5001*** (0,0867)	6,1995*** (0,0731)	0,9711*** (0,0855)
Log(migrantes) <i>ij</i> 2004	0,8600*** (0,2681)	0,7965*** (0,2458)	0,8395** (0,3309)	0,8203** (0,3256)	0,6960*** (0,2077)
Log(migrantes) <i>ik</i> 2004	-4,5348*** (0,0603)	-4,0206*** (0,0436)	-6,2676*** (0,0364)	-5,3420*** (0,0501)	-3,7932*** (0,0685)
Intercepto	-23,5517*** (0,0455)	-13,9258*** (0,0466)	-8,0747*** (0,0620)	-12,3156*** (0,0506)	-20,3705*** (0,0550)
Observações	3.510	3.510	3.510	3.510	3.510
Efeitos fixos					
Ano	sim	sim	sim	sim	sim
UF de origem	sim	sim	sim	sim	sim
UF de destino	sim	sim	sim	sim	sim
UF de origem/ano	sim	sim	sim	sim	sim
UF de destino/ano	sim	sim	sim	sim	sim

Fonte: Elaboração própria.

Notas: *i* – indexa o estado de partida do fluxo migratório; *j* indexa o estado de destino dos migrantes; *k* indexa todos os estados de destino, exceto *j*; *t-1* indica que a variável refere-se ao período anterior à contagem dos fluxos migratórios. *** Estatisticamente significativo a 1%. ** Estatisticamente significativo a 5%. * Estatisticamente significativo a 10%. Erros-padrão robustos à heterocedasticidade entre parênteses. Resultados estimados a partir do algoritmo MCMC desenvolvido por Baker (2013).

Ao se comparar as regressões feitas para grupos de migrantes por faixa de instrução (colunas 3, 4 e 5), pode-se verificar que quanto maior a população do estado de partida, maior tende a ser a emigração no ano posterior, sobretudo, a emigração de trabalhadores de baixa escolaridade. O efeito da população do estado de destino é assimétrico. Para os migrantes de 0 a 4 anos de estudo e de 5 a 10 anos, o aumento da população reduz a absorção de migrantes procedentes de qualquer estado no período posterior. No entanto, para migrantes mais instruídos (11 anos de estudo ou mais) essa correlação é direta. Tal evidência sugere que estados com maior população/aglomeração podem produzir externalidades positivas (maior acesso à cultura, serviços, diversificação de consumo e de oportunidades de trabalho) para trabalhadores mais qualificados.

O coeficiente de elasticidade da variável distância bilateral indica que os migrantes com maior instrução parecem ser mais sensíveis às distâncias bilaterais quando da tomada de decisão de migrar em relação aos migrantes com menor escolaridade. Uma provável explicação para tal achado é que as pessoas mais instruídas podem ter maior discernimento sobre vários custos associados à mobilidade geográfica, especialmente, custos monetários de viagens diretamente afetados pela distância entre os locais de origem e destino. Também se observa que estados vizinhos apresentam maior fluxo bilateral de migração em relação aos não vizinhos, independente da faixa etária ou de instrução dos migrantes. Cabe notar que entre os migrantes de baixa instrução (de 0 a 4 anos de escolaridade), o efeito vizinhança é relativamente mais forte de acordo com migrantes com maior instrução. Esse resultado pode relacionar-se com a maior facilidade de se detectar custos de transporte e/ou de acesso a informações sobre postos de trabalho entre estados que compartilham fronteiras.

A hipótese de que os migrantes mais instruídos conseguem assimilar melhor os custos relacionados à migração pode ser reforçada com as evidências sobre a elasticidade dos fluxos migratórios interestaduais relacionados ao estoque acumulado de migrantes. Os resultados mostram que, mesmo controlando a emissão passada de migrantes do estado de partida para quaisquer outros estados, há uma relação direta entre os movimentos de migração bilateral e o estoque de migrantes acumulados. Em particular, o coeficiente de elasticidade da variável migrantes acumulado em 2004 no estado que nasceram no estado é relativamente menor quando se consideram fluxos de migrantes com onze ou mais anos de estudo (coluna 5). Ou seja, as conexões geradas por redes sociais parecem ter menor efeito entre os trabalhadores mais escolarizados, possivelmente em razão desses últimos absorverem melhor vários custos associados à migração (custos de moradia, procura por trabalho, busca de informações etc.).

O processo de seleção dos migrantes interestaduais por instrução também produz diferenças importantes na resposta dos fluxos migratórios ao hiato de rendimento esperado por ocupação. Conforme resultados apresentados na tabela 4, o coeficiente da variável renda relativa é bem menor para migrantes com onze ou mais anos de estudo comparado às estimativas para os demais grupos. Tal achado favorece a hipótese de maior importância de diferenças de amenidades naturais e/ou sociais³³ na função de bem-estar dos trabalhadores mais instruídos, uma vez controlados efeitos fixos por estados de partida e de destino.

Para fins comparativos, os modelos de regressão constantes na tabela 4 também foram estimados por MQO e se acham na tabela A.4 do apêndice. Em linhas gerais, os resultados mostram que, mesmo ponderando efeitos de seletividade dos fluxos migratórios, as estimativas por MQO produzem vieses relevantes em relação ao grande número de fluxos zero e à identificação log-linear do modelo gravitacional. Há diferenças relevantes de magnitude entre os coeficientes estimados na tabela 4 por regressão *quantílica* (mediana) e aqueles obtidos por MQO (tabela A.4), sobretudo no tocante às covariadas de distância, compartilhamento de fronteira e migrantes acumulados.

5 CONSIDERAÇÕES FINAIS

Este artigo teve por objetivo investigar os principais determinantes dos fluxos migratórios interestaduais no Brasil durante a década de 2000. Para tanto, foram produzidas evidências com base em um modelo gravitacional aplicado a um painel de dados de Pnads anuais. O estudo contribui para a literatura especializada ao estimar um modelo gravitacional com resistência multilateral (Anderson, 2011) pelo método de regressão *quantílica* censurada, o qual permite superar problemas bastantes conhecidos, tais como o de identificação estrutural (modelo log-linear), acomodação de grande quantidade de fluxos zero, robustez à heterocedasticidade e vies por omissão de variáveis relacionadas às atratividades exercidas por várias regiões sobre os movimentos bilaterais (Figueiredo, Lima e Orefice, 2016). Por outro lado, além de atualizar evidências para o caso do Brasil, esse trabalho agrega novas evidências à literatura nacional, sobretudo no esforço de capturar possíveis efeitos de redes sociais e de heterogeneidade dos migrantes levando em consideração estimativas econométricas por faixas de idade e de instrução.

Os resultados mostraram que os fluxos de migração interestadual no Brasil estão associados às diferenças de renda esperada por ocupação, à dimensão populacional, às distâncias bilaterais, ao compartilhamento de fronteiras (vizinhança) e à presença

33. Em razão de vários tipos de amenidades locais serem relativamente constantes no tempo (dotação local de parques, praias, oferta de cultura e lazer, preferência por consumo no local de origem), o conjunto de variáveis binárias para efeitos fixos por estado de origem e de destino nas regressões em destaque também podem capturar tais diferenças de amenidades.

de migrantes acumulados de mesma procedência. Em linhas gerais, os achados empíricos sugerem que: *i*) estados com maior renda relativa (considerando o risco de desemprego por grupos de ocupação) apresentam maior atratividade de migrantes, enquanto maiores distâncias bilaterais parecem reduzir os movimentos populacionais; *ii*) o número de migrantes acumulados no passado e oriundo da mesma região de partida de migrantes recentes, se relacionada diretamente às migrações bilaterais, favorecendo possíveis efeitos de redes sociais sobre custos associados à migração do trabalho; *iii*) estados vizinhos também apresentam maiores fluxos de migrantes; e *iv*) o crescimento da população nos estados típicos de origem tende a acelerar o processo de emigração. A maioria dessas correlações revelaram-se robustas mediante diferentes especificações do modelo econométrico em covariadas observadas, considerando a amostra geral de fluxos bilaterais de migrantes e controles para efeitos fixos para resistência multilateral (efeitos não observados do conjunto de atratividades regionais sobre os fluxos bilaterais).

Os efeitos da seletividade dos fluxos migratórios não devem ser ignorados mesmo em estudos envolvendo dados agregados. As estimações econométricas, considerando migrantes por diferentes faixas de idade e de escolaridade, revelaram coeficientes de elasticidade bastante diferenciados entre grupos, especialmente entre os migrantes jovens e migrantes com alta instrução. Em geral, os fluxos bilaterais de migrantes jovens (entre 25 e 35 anos de idade) respondem melhor às diferenças de renda esperada por ocupação que de migrantes adultos (entre 36 e 65 anos). Por sua vez, os migrantes mais escolarizados (com 11 ou mais anos completos de estudo) registram menores coeficientes de elasticidade no tocante à diferença relativa de rendimento do trabalho e ao número de migrantes acumulados no passado. Tais evidências sugerem que os migrantes interestaduais mais qualificados podem valorizar diferenças de amenidades entre estados de origem e de destino e lidar melhor com custos envolvidos no processo de migração.

É importante reconhecer a grande complexidade dos movimentos migratórios em um país de grandes dimensões geográficas, desigualdades sociais e população como o Brasil. Os achados deste estudo fornecem suporte ao papel das diferenças regionais de renda e de emprego, dos custos relacionados às distâncias bilaterais e vizinhança e das redes sociais no condicionamento das migrações interestaduais. As respostas dos movimentos migratórios às covariadas que representam diferentes conjuntos de oportunidades (grau de atratividade por estrutura de incentivos e custos) dependem da seletividade dos trabalhadores em idade e instrução. Nesse contexto, o desenho de políticas públicas voltadas à redução das desigualdades regionais deve considerar o impacto das migrações interestaduais nesse processo, sobretudo quando mudanças nos padrões demográficos e/ou nas estruturas das economias regionais estão associadas à redução de riscos e custos envolvidos na mobilidade geográfica do trabalho.

REFERÊNCIAS

- ANDERSON, J. E. A Theoretical Foundation for the Gravity Equation. **American Economic Review**, v. 69, n. 1, p. 106-116, 1979.
- ANDERSON, J. E. The Gravity Model. **Annual Review of Economics**, v. 3, n. 1, p. 133-160, 2011. ISSN 1941-1383.
- ANDERSON, J. E.; WINCOOP, E. V. Gravity with Gravitas: A Solution to the Border Puzzle. **The American Economic Review**, American Economic Association, v. 93, n. 1, p. 170-192, 2003. ISSN 00028282.
- ANDERSON, J. E.; YOTOV, Y. V. The Changing Incidence of Geography. **American Economic Review**, v. 100, n. 5, p. 2157-2186, 2010.
- ARBEX, M.; FREGUGLIA, R.; CHEIN, F. Informal economy and spatial mobility: are informal workers economic refugees? **Journal of Economic Studies**, v. 40, n. 5, p. 671-685, 2013.
- BAENINGER, R. A. Região, metrópole e interior: espaços ganhadores e espaços perdedores nas migrações recentes: Brasil, 1980-1996. *In: Redistribuição da população e meio ambiente: São Paulo e Centro-Oeste*. Campinas: Universidade Estadual de Campinas - Unicamp, Núcleo de Estudos de População - NEPO: [s.n.], 2000. cap. Textos NEP, p. 1-200.
- BAKER, M. J. **Adaptive Markov chain Monte Carlo sampling and estimation in Mata**. [S.l.], 2013.
- BEINE, M.; BERTOLI, S.; FERNÁNDEZ-HUERTAS MORAGA, J. A practitioners' guide to gravity models of international migration. **mimeo**, 2014.
- BEINE, M.; DOCQUIER, F.; ÖZDEN, C. Diasporas. **Journal of Development Economics**, v. 95, n. 1, p. 30-41, maio 2011. ISSN 03043878.
- BEINE, M.; PARSONS, C. **Climatic factors as determinants of International Migration**. 2012.
- BERTOLI, S.; FERNÁNDEZ-HUERTAS MORAGA, J. **Visa Policies, Networks and the Cliff at the Border**. 2012.
- BERTOLI, S.; FERNÁNDEZ-HUERTAS MORAGA, J. Multilateral resistance to migration. **Journal of Development Economics**, v. 102, p. 79-100, maio, 2013. ISSN 03043878.
- BORJAS, G. J. Self-Selection and the Earnings of Immigrants. **American Economic Review**, JSTOR, v. 77, n. 4, p. 531-553, 1987. ISSN 00028282.
- BORJAS, G. J. The Economics of Immigration. **Journal of Economic Literature**, American Economic Association, v. 32, n. 4, p. 1667-1717, 1994. ISSN 00220515.

CANÇADO, R. P. Migrações e Convergência no Brasil: 1960-91. **Revista Brasileira de Economia**, v. 53, n. 2, p. 211-236, 1999.

CHERNOZHUKOV, V.; HONG, H. Three-Step Censored Quantile Regression and Extramarital Affairs. **Journal of the American Statistical Association**, American Statistical Association, v. 97, n. 459, p. 872-882, 2002. ISSN 01621459.

CHERNOZHUKOV, V.; HONG, H. An MCMC approach to classical estimation. **Journal of Econometrics**, v. 115, n. 2, p. 293-346, 2003. ISSN 0304-4076.

CHISWICK, B. R. Are Immigrants Favorably Self-Selected? **The American Economic Review**, American Economic Association, v. 89, n. 2, p. 181-185, 1999. ISSN 00028282.

CLARK, X.; HATTON, T. J.; WILLIAMSON, J. G. Explaining U.S. Immigration, 1971-1998. **The Review of Economics and Statistics**, The MIT Press, v. 89, n. 2, p. 359-373, 2007. ISSN 00346535.

DEDECCA, C. S.; CUNHA, J. M. P. Migração, trabalho e renda nos anos 90: o caso da Região Metropolitana de São Paulo. **Revista Brasileira de Estudos de População**, v. 21, n. 1, p. 49-66, 2004.

DOS SANTOS JÚNIOR, E. d. R.; MENEZES-FILHO, N.; FERREIRA, P. C. Migração, seleção e diferenças regionais de renda no Brasil. **Pesquisa e Planejamento Econômico**, v. 35, n. 3, p. 299-331, 2005.

FIESS, N. M.; VERNER, D. **Migration and human capital in Brazil during the 1990s**. [S.l.], 2003. (World Bank Policy Research, 3093).

FIGUEIREDO, E.; LIMA, L. R.; OREFICE, G. **Migration and Regional Trade Agreement: a (new) Gravity Estimation**. *Review of International Economics*, vol. 24, n. 1, p. 99-125, 2016.

FIGUEIREDO, E.; LIMA, L. R.; SCHAUR, G. Robust estimation of gravity equations and the WTO impact on trade inequality. *In: CESIFO CONFERENCE CENTRE. CESifo Conference Centre*. Munich, 2014. p. 1-27.

FREGUGLIA, R.; GONÇALVES, E.; SILVA RIBEIRO, E. Composition and determinants of the skilled out-migration in the Brazilian formal labor market: A panel data analysis from 1995 to 2006. **Economia**, v. 15, n. 1, p. 100-117, 2014. ISSN 1517-7580.

GOLGHER, A. B.; ROSA, C. H.; ARAÚJO JÚNIOR, A. F. Determinants of migration in Brazil: regional polarization and poverty traps. **Papeles de Población**, v. 14, n. 56, p. 135-171, 2008.

HARRIS, J. R.; TODARO, M. P. Migration, Unemployment and Development: A Two-Sector Analysis. **American Economic Review**, v. 60, n. 1, p. 126-142, 1970.

- HEAD, K.; MAYER, T. Gravity Equations: Workhorse, Toolkit, and Cookbook. In: G, E. H.; ROGOFF, K. (Ed.). **The Handbook of International Economics**. [S.l.]: Elsevier, 2014, (The Handbook of International Economics, v. 4). cap. 3, p. 131-195.
- HECKMAN, J. Sample Selection Bias as Specification Error. **Econometrica**, v. 47, n. 1, p. 153-162, 1979.
- JUSTO, W. R.; SILVEIRA NETO, R. Migração inter-regional no Brasil: Evidências a partir de um Modelo Espacial. **EconomiA**, v. 7, n. 163-187, 2006.
- KOENKER, R. **Quantile regression**. Cambridge: Cambridge University Press, 2005.
- LEWER, J. J.; VAN DEN BERG, H. A gravity model of immigration. **Economics Letters**, v. 99, n. 1, p. 164-167, 2008. ISSN 0165-1765.
- LLULL, J. **Understanding international migration**: evidence from a new dataset of bilateral stocks (1960-2000). Barcelona Graduate School of Economics, 2013. Working Papers 715.
- MAYDA, A. M. International migration: a panel data analysis of the determinants of bilateral flows. **Journal of Population Economics**, v. 23, n. 4, p. 1249-1274, 2010.
- MCKENZIE, D.; RAPOPORT, H. Network effects and the dynamics of migration and inequality: Theory and evidence from Mexico. **Journal of Development Economics**, Elsevier, v. 84, n. 1, p. 1-24, 2007. ISSN 03043878.
- NETTO JÚNIOR, J. L. S. *et al.* Fluxos migratórios e dispersão das rendas per capita estaduais: uma análise por dados em painel no período de 1950-2000. **Revista Econômica do Nordeste**, v. 34, n. 3, p. 379-404, 2003.
- OLIVEIRA, A. T. R.; ERVATTI, L. R.; O'NEILL, M. M. V. C. O panorama dos deslocamentos populacionais no Brasil: PNADs e Censos Demográficos. In: OLIVEIRA, L. A. P.; OLIVEIRA, A. T. R. (Ed.). **Reflexões sobre os Deslocamentos Populacionais no Brasil**. IBGE, cap. 2. 2011. ISBN 978-85-240-4191-4.
- ORTEGA, F.; PERI, G. The Role of Income and Immigration Policies in Attracting International Migrants. **Migration Studies**, v. 1, n. 1, p. 47-74, 2013.
- POWELL, J. Least Squares Absolute Deviations Estimation for the Censored Regression Model. **Journal of Econometrics**, v. 25, p. 303-325, 1984.
- POWELL, J. Censored Regression Quantiles. **Journal of Econometrics**, v. 32, p. 143-155, 1986.
- QUEIROZ, V. S.; RAMALHO, H. M. B. Migração interestadual de retorno e autosseleção: Evidências para o Brasil. **Pesquisa e Planejamento Econômico**, v. 42, p. 731-771, 2011.

RAMOS, R.; SURINACH, J. **A gravity model of migration between ENC and EU**. 2013.

SABBADINI, R.; AZZONI, C. R. **Migração interestadual de pessoal altamente educado: evidências sobre a fuga de cérebros**. In: XXXIV Encontro Nacional de Economia. Salvador: Anpec, 2006.

SAHOTA, G. S. An Economic Analysis of Internal Migration in Brazil. **Journal of Political Economy**, The University of Chicago Press, v. 76, n. 2, p. 218-245, 1968. ISSN 00223808.

SANTOS SILVA, J. M. C.; TENREYRO, S. The Log of Gravity. **Review of Economics and Statistics**, MIT Press, v. 88, n. 4, p. 641-658, nov. 2006. ISSN 0034-6535.

SCHULTZ, T. W. Investment in Human Capital. **American Economic Review**, JSTOR, v. 51, n. 1, p. 1-17, 1961. ISSN 00028282.

SIMPSON, N. B.; SPARBER, C. **The Short- and Long-Run Determinants of Less-Educated Immigration into U.S. States**. 2012.

SJAASTAD, L. A. The costs and returns of human migration. **Journal of Political Economy**, JSTOR, v. 70, n. 5, p. 80-93, 1962. ISSN 00223808.

VOGLER, M.; ROTTE, R. The effects of development on migration: Theoretical issues and new empirical evidence. **Journal of Population Economics**, v. 13, n. 3, p. 485-508, 2000.

YAMAUCHI, F.; TANABE, S. Nonmarket networks among migrants: evidence from metropolitan Bangkok, Thailand. **Journal of Population Economics**, v. 21, n. 00003, p. 649-664, 2008.

(Originais submetidos em maio de 2014. Última versão recebida em julho de 2015. Aprovada em julho de 2015.)

APÊNDICE

QUADRO A.1

Descrição das variáveis selecionadas para a análise empírica

Variável	Tipo	Correlação	Descrição
Log(migrantes) ij	Variável contínua		Logaritmo do total de migrantes que na data de entrevista residiam no estado j há no máximo um ano, mas que anteriormente se achavam no estado i . Fonte: Pnads de 2005, 2006, 2007, 2008 e 2009. Nota: calculada a partir do cruzamento dos quesitos sobre unidade federativa de residência, tempo de residência no estado e unidade federativa de residência anterior. Dados ponderados pelos pesos da amostra.
Renda relativa $ji(t-1)$	Variável contínua	Positiva	Média de razão de rendimentos esperados do trabalho por grupos de ocupação entre o estado de destino j e o estado de origem i (residência anterior) no período de partida dos fluxos migratórios, isto é, em $t-1$. Nesse cálculo, foram considerados nove grupos de ocupação conforme classificação do IBGE: (a) Dirigentes, (b) Profissionais das Ciências e das Artes, (c) Técnicos de nível médio, (d) Trabalhadores de serviços administrativos, (e) Trabalhadores dos serviços, (f) Vendedores e prestadores de serviço do comércio, (g) Trabalhadores agrícolas, (h) Trabalhadores da produção de bens e serviços e (i) Membros das forças armadas e auxiliares. A renda esperada da ocupação k no estado l foi calculada pela média de rendimento dos trabalhadores alocados em k no estado l ponderada pela probabilidade de desemprego em k (um menos a proporção de trabalhadores ocupados na categoria k). O rendimento relativo da ocupação k foi calculado pela razão de renda esperada da referida ocupação, considerado o estado de destino do fluxo de migrantes e tomando por base o estado de partida. Finalmente, a renda relativa esperada é uma média dos rendimentos relativos do trabalho por ocupação considerado cada par de origem-destino dos fluxos migratórios. Fonte: Pnads de 2004, 2005, 2006, 2007 e 2008. Nota: valores corrigidos para preços constantes de 2008 a partir do Índice de Preços ao Consumidor Amplo (IPCA) para preços constantes de 2008 (ipeadata). Médias estaduais calculadas considerando o peso de cada pessoa na amostra.
Log(população) $i(t-1)$	Variável contínua	Positiva	Logaritmo da população do estado i (estado de origem) no período de partida $t-1$ dos fluxos migratórios. Fonte: Pnads de 2004, 2005, 2006, 2007 e 2008. Nota: totais por estados calculados considerando o peso de cada pessoa na amostra.
Log(população) $j(t-1)$	Variável contínua	Positiva ou negativa	Logaritmo da população do estado j (estado de destino) no período de partida $t-1$ dos fluxos migratórios. Fonte: Pnads de 2004, 2005, 2006, 2007 e 2008. Nota: totais por estados calculados considerando o peso de cada pessoa na amostra.
Log(distância) ij	Variável contínua	Negativa	Logaritmo da distância geodésica entre a capital de i (estado de origem dos fluxos migratórios) e a capital de j (estado de destino). Fonte: Malhas digitais (BGE). Nota: calculada a partir da distância entre os centroides das capitais estaduais, considerando a curvatura do planeta Terra a partir do seu raio de 6.371 quilômetros aplicado ao cálculo trigonométrico em esferas.
Fronteira ij	Variável binária	Positiva	Assume o valor um se o estado i (estado de origem) compartilha fronteira com o estado j (estado de destino) e zero, caso contrário. Fonte: Malhas digitais (BGE).
Ln estoque migração ij (2004)	Variável contínua	Positiva	Logaritmo do total de migrantes que na data de entrevista residiam no estado j , mas que nasceram no estado i . Ano da entrevista – 2004. Fonte: Pnad de 2004. Calculada a partir do cruzamento dos quesitos sobre unidade federativa de residência e unidade federativa de nascimento. Dados ponderados pelos pesos da amostra.
		Negativa	Logaritmo do total de migrantes que na data de entrevista residiam em todos os estados k $6=j$, mas que nasceram no estado i . Neste caso, o estado k é qualquer outro que não o estado j . Ano da entrevista – 2004. Fonte: Pnad de 2004. Calculada a partir do cruzamento dos quesitos sobre unidade federativa de residência e unidade federativa de nascimento. Dados ponderados pelos pesos da amostra.
TRM	Variáveis binárias	Positiva ou negativa	Conjunto de variáveis binárias por estado de origem dos fluxos migratórios, estado de destino, período de migração, binárias de interação entre estado de origem e período e binárias de interação entre estado de destino e período.

Fonte: Pnads de 2005, 2006, 2007, 2008 e 2009.

Nota: Calculada a partir do cruzamento dos quesitos sobre unidade federativa de residência, tempo de residência no estado e unidade federativa de residência anterior. Dados ponderados pelos pesos da amostra.

TABELA A.1

Regressões por MQO: variável dependente – fluxo de migração interestadual entre estados de origem e destino
(Em logaritmo)

	(1)	(2)	(3)	(4)	(5)
Log(população) <i>i</i> (t-1)	1,3864*** (0,1042)	4,9368 (4,9150)	3,4755 (4,9125)	4,7962*** (0,4950)	4,7392*** (0,4940)
Log(população) <i>j</i> (t-1)	0,5481*** (0,0411)	0,5043*** (0,0415)	-1,1162 (4,6876)	-0,5643 (4,7945)	0,0136 (0,2175)
Log(distância) <i>ij</i>	-0,7873*** (0,0860)	-0,7217*** (0,0890)	-0,8115*** (0,0995)	-0,8537*** (0,1010)	-0,9326*** (0,1050)
Fronteira <i>ij</i>	1,8094*** (0,1563)	1,6793*** (0,1620)	1,6400*** (0,1692)	1,6621*** (0,1720)	1,7034*** (0,1757)
Renda relativa <i>ji</i> (t-1)	-0,0169 (0,1810)	0,6337** (0,2486)	3,5008*** (0,6142)	4,5111*** (0,7145)	6,4012*** (0,9141)
Log(migrantes) <i>ij</i> 2004	0,3014*** (0,0160)	0,2489*** (0,0164)	0,1890*** (0,0185)	0,1854*** (0,0191)	0,1787*** (0,0196)
Log(migrantes) <i>ik</i> 2004	-0,6819*** (0,0896)	-3,4053*** (0,4514)	-4,2327*** (0,5458)	-4,1337*** (0,5510)	-3,9485*** (0,5498)
Intercepto	-14,0852***	-31,3376	20,9545	-7,3807	-18,4844***
R2	0,4435	0,4686	0,4894	0,4906	0,4915
Observações	3.510	3.510	3.510	3.510	3.510
Efeitos fixos					
Ano	sim	sim	sim	sim	sim
UF de origem	não	sim	sim	sim	sim
UF de destino	não	não	sim	sim	sim
UF de origem/ano	não	não	não	sim	sim
UF de destino/ano	não	não	não	não	sim

Fonte: Elaboração própria.

Notas: *i* – indexa o estado de partida do fluxo migratório; *j* indexa o estado de destino dos migrantes; *k* indexa todos os estados de destino, exceto *j*; *t*-1 indica que a variável refere-se ao período anterior à contagem dos fluxos migratórios. *** Estatisticamente significativo a 1%. ** Estatisticamente significativo a 5%. * Estatisticamente significativo a 10%. Erros-padrão robustos à heterocedasticidade entre parênteses.

TABELA A.2

Regressões por MQO e MCMC (mediana) sem controles para efeitos fixos: variável dependente – fluxo de migração interestadual entre estados de origem e destino (Em logaritmo)

	(1) MQO	(2) MCMC	(3) MQO	(4) MCMC	(5) MQO	(6) MCMC	(7) MQO	(8) MCMC
Log(população) <i>i</i> (t-1)	1,0979*** (0,0402)	1,5061*** (0,1229)	1,1011*** (0,0403)	1,5249*** (0,1105)	0,6800*** (0,0456)	0,6128*** (0,0841)	1,3816*** (0,1043)	1,2419*** (0,1137)
Log(população) <i>j</i> (t-1)	0,8034*** (0,0414)	0,9335*** (0,1096)	0,7956*** (0,0412)	0,9120*** (0,1057)	0,5801*** (0,0408)	0,1204 (0,0979)	0,5469*** (0,0412)	0,0998 (0,1437)
Log(distância) <i>ij</i>	-1,0816*** (0,0872)	-2,3478*** (0,4226)	-1,0855*** (0,0872)	-2,4199*** (0,3541)	-0,7047*** (0,0862)	-0,9462*** (0,1141)	-0,7873*** (0,0861)	-0,6776*** (0,1548)
Fronteira <i>ij</i>	2,4386*** (0,1617)	2,7953*** (0,3620)	2,4459*** (0,1616)	2,7992*** (0,1214)	1,8947*** (0,1581)	0,7967 (0,5050)	1,8094*** (0,1569)	1,3345*** (0,3289)
Renda relativa <i>ji</i> (t-1)			0,2602 (0,1788)	0,1472* (0,0797)	-0,4277** (0,1719)	0,2557 (0,1822)	-0,0142 (0,1824)	-0,1088 (0,0937)
Log(migrantes) <i>ij</i> 2004					0,3075*** (0,0160)	1,2992*** (0,2488)	0,3016*** (0,0160)	1,3253*** (0,3749)
Log(migrantes) <i>ik</i> 2004							-0,6784*** (0,0897)	-0,8013*** (0,0894)
Intercepto	-18,4210*** (1,2153)	-18,7958*** (0,2678)	-18,5997*** (1,2232)	-18,4594*** (0,1112)	-13,2771*** (1,2120)	-13,8841*** (0,2153)	-14,2208*** (1,2130)	-14,3410*** (0,1323)
R2 ajustado	0,3866		0,3869		0,4300		0,4388	
Observações	3.510	3.510	3.510	3.510	3.510			
Efeitos fixos								
Ano	não	não	não	não	não	não	não	não
UF de origem	não	não	não	não	não	não	não	não
UF de destino	não	não	não	não	não	não	não	não
UF de origem/ano	não	não	não	não	não	não	não	não
UF de destino/ano	não	não	não	não	não	não	não	não

Fonte: Elaboração própria.

Nota: *i* – indexa o estado de partida do fluxo migratório; *j* indexa o estado de destino dos migrantes; *k* indexa todos os estados de destino, exceto *j*; *t*-1 indica que a variável refere-se ao período anterior à contagem dos fluxos migratórios. *** Estatisticamente significativo a 1%. ** Estatisticamente significativo a 5%. * Estatisticamente significativo a 10%. Erros-padrão robustos à heterocedasticidade entre parênteses.

TABELA A.3

Regressões por MQO e MCMC (mediana) com controles para efeitos fixos: variável dependente – fluxo de migração interestadual entre estados de origem e destino (Em logaritmo)

	(1) MQO	(2) MCMC	(3) MQO	(4) MCMC	(5) MQO	(6) MCMC	(7) MQO	(8) MCMC
Log(população) <i>i</i> (t-1)	1,4052*** (0,2269)	1,6363*** (0,0623)	1,7746*** (0,2279)	2,0125*** (0,1384)	1,5475*** (0,2161)	1,2628*** (0,0610)	4,7392*** (0,4940)	4,5387*** (0,0789)
Log(população) <i>j</i> (t-1)	1,0103*** (0,2188)	1,0360*** (0,0691)	0,1197 (0,2278)	0,1650*** (0,0291)	0,1687 (0,2204)	-0,0159 (0,0904)	0,0136 (0,2175)	-0,1164 (0,0928)
Log(distância) <i>ij</i>	-1,0973*** (0,0973)	-1,7830*** (0,2431)	-1,3946*** (0,0998)	-2,0825*** (0,2975)	-1,0360*** (0,1040)	-1,0926*** (0,0406)	-0,9326*** (0,1050)	-1,2121*** (0,1210)
Fronteira <i>ij</i>	2,3284*** (0,1742)	2,5205*** (0,0697)	2,3850*** (0,1697)	2,7315*** (0,0396)	2,1244*** (0,1676)	1,8828*** (0,1183)	1,7034*** (0,1757)	1,8264*** (0,1062)
Renda relativa <i>ji</i> (t-1)			8,8434*** (0,9810)	8,8293*** (0,1049)	7,4586*** (0,9497)	7,3548*** (0,1736)	6,4012*** (0,9141)	6,2554*** (0,0593)
Log(migrantes) <i>ij</i> 2004					0,2100*** (0,0199)	0,9855*** (0,1800)	0,1787*** (0,0196)	0,8125*** (0,1833)
Log(migrantes) <i>ik</i> 2004							-3,9485*** (0,5498)	-3,8795*** (0,0354)
Intercepto	-23,9536*** (5,4452)	-23,8434*** (0,0703)	-23,1843*** (5,2852)	-23,4712*** (0,0413)	-23,8950*** (5,0437)	-23,7741*** (0,0359)	-18,4844*** (5,0088)	-18,7154*** (0,0622)
R2 ajustado	0,4499		0,4678		0,4823		0,4915	
Observações	3.510	3.510	3.510	3.510	3.510			
Efeitos fixos								
Ano	sim	sim	sim	sim	sim	sim	sim	sim
UF de origem	sim	sim	sim	sim	sim	sim	sim	sim
UF de destino	sim	sim	sim	sim	sim	sim	sim	sim
UF de origem/ano	sim	sim	sim	sim	sim	sim	sim	sim
UF de destino/ano	sim	sim	sim	sim	sim	sim	sim	sim

Fonte: Elaboração própria.

Nota: *i* – indexa o estado de partida do fluxo migratório; *j* indexa o estado de destino dos migrantes; *k* indexa todos os estados de destino, exceto *j*; *t*-1 indica que a variável refere-se ao período anterior à contagem dos fluxos migratórios. *** Estatisticamente significativo a 1%. ** Estatisticamente significativo a 5%. * Estatisticamente significativo a 10%. Erros-padrão robustos à heterocedasticidade entre parênteses.

TABELA A.4

Regressões por MQO: variável dependente – fluxo de migração interestadual entre estados de origem e destino por faixa etária e faixa de instrução (em logaritmo)

	(1) Idade: 20 a 35	(2) Idade: 36 a 65	(3) Instrução: 0 a 4	(4) Instrução: 5 a 10	(5) Instrução: 11 anos+
Log(população) <i>i</i> (t-1)	5,0333*** (0,5457)	4,4464*** (0,5682)	6,2538*** (0,5166)	5,6209*** (0,5491)	4,2058*** (0,5037)
Log(população) <i>j</i> (t-1)	0,1392 (0,2465)	-0,0744 (0,2255)	-0,7094*** (0,2366)	-0,5229** (0,2121)	0,6176** (0,2501)
Log(distância) <i>ij</i>	-0,7393*** (0,1080)	-0,6424*** (0,1065)	-0,3761*** (0,0943)	-0,4354*** (0,1052)	-0,7697*** (0,1087)
Fronteira <i>ij</i>	2,0137*** (0,1901)	1,8524*** (0,1899)	1,8276*** (0,1910)	1,8327*** (0,1935)	1,3760*** (0,1867)
Renda relativa <i>ij</i> (t-1)	6,6623*** (0,9512)	3,4100*** (0,9097)	6,4100*** (1,0335)	6,0696*** (1,0946)	1,0262 (0,9605)
Log(migrantes) <i>ij</i> 2004	0,1293*** (0,0176)	0,1127*** (0,0163)	0,1108*** (0,0145)	0,0881*** (0,0146)	0,0575*** (0,0172)
Log(migrantes) <i>ik</i> 2004	-4,2282*** (0,6190)	-3,7962*** (0,6425)	-6,1544*** (0,5788)	-5,1671*** (0,6264)	-3,5891*** (0,5576)
Intercepto	-23,6114*** (5,2451)	-14,0100*** (5,1177)	-8,0050 (5,1613)	-12,2417** (4,9127)	-20,4755*** (5,2685)
R2	0,4447	0,4149	0,4072	0,4055	0,3926
Observações	3.510	3.510	3.510	3.510	3.510
Efeitos fixos					
Ano	sim	sim	sim	sim	sim
UF de origem	sim	sim	sim	sim	sim
UF de destino	sim	sim	sim	sim	sim
UF de origem/ano	sim	sim	sim	sim	sim
UF de destino/ano	sim	sim	sim	não	sim

Fonte: Elaboração própria.

Notas: *i* – indexa o estado de partida do fluxo migratório; *j* indexa o estado de destino dos migrantes; *k* indexa todos os estados de destino, exceto *j*; *t*-1 indica que a variável refere-se ao período anterior à contagem dos fluxos migratórios.

*** Estatisticamente significativo a 1%. ** Estatisticamente significativo a 5%. * Estatisticamente significativo a 10%.
Eros-padrão robustos à heterocedasticidade entre parênteses.

CONCORRÊNCIA NO SETOR BANCÁRIO BRASILEIRO: BANCOS INDIVIDUAIS *VERSUS* CONGLOMERADOS BANCÁRIOS

Marcelo Cardoso¹

Paulo F. Azevedo²

Klenio Barbosa³

O conceito de grupo econômico é central em análises de concorrência. Entretanto, a maioria dos estudos sobre competição bancária no Brasil utiliza o relatório 4010 do Banco Central, que fornece informações de bancos identificados pelo seu Cadastro Nacional de Pessoas Jurídicas (CNPJ), desconsiderando, portanto, possíveis efeitos de coordenação entre empresas pertencentes a um mesmo conglomerado financeiro. Este artigo busca investigar empiricamente se essa desconsideração do grupo econômico produz viés nas estimativas de medidas de competição no mercado bancário brasileiro. Neste artigo são estimados os parâmetros usuais de competição no setor bancário, sugeridos por Bresnahan e Lau (1982), Panzar e Rosse (1987) e Boone (2008), utilizando dados financeiro-contábeis de bancos individuais (relatório 4010) e de conglomerados bancários (relatório 4040). As estimativas mostram que, a partir de 2004, há evidências de diferenças significantes entre as duas bases de dados, o que revela que estudos que desconsideram efeitos de coordenação entre empresas do mesmo grupo econômico estão sujeitos a viés de superestimação da intensidade de concorrência bancária.

Palavras-chave: concorrência bancária; grupo econômico; conglomerados financeiros.

COMPETITION IN THE BRAZILIAN BANKING SECTOR: INDIVIDUALIZED BANKS VERSUS BANKING CONGLOMERATES

The concept of economic group is central in competition analysis. However, most studies on banking competition in Brazil uses the Central Bank 4010 Report, which provides information to banks identified by their tax-identification number (CNPJ), disregarding therefore possible coordination between companies belonging to the same financial conglomerate. This paper seeks to empirically investigate whether this disregard of the economic group produces bias in the estimates of competition measures in the Brazilian banking market. In this article we estimate usual parameters of banking competition, suggested by Bresnahan and Lau (1982), Panzar and Rosse (1987) and Boone (2008) using financial-accounting data of individual banks (Report 4010) and banking conglomerates (Report 4040). Our estimates show that, since 2004, there is evidence of a significant difference between the two databases, which reveals that studies that disregard coordination between companies of the same group are subject to overestimation bias of banking competition intensity.

Keywords: banking competition; economic group; financial conglomerates.

JEL: L1; L4; L80.

1. Doutor em Economia pela EESP/FGV. Professor de Economia no Mackenzie e na PUC-SP. *E-mail:* <m.ranieri@uol.com.br>.

2. Professor Insper. *E-mail:* <PauloFA1@insper.edu.br>.

3. Professor da EESP/FGV. *E-mail:* <klenio.barbosa@fgv.br>.

1 INTRODUÇÃO

Bancos mobilizam, alocam e investem a poupança da sociedade. Por tal papel, a conduta dos bancos e o desempenho da indústria bancária têm substancial repercussão sobre a alocação do capital, crescimento das firmas, expansão industrial e desenvolvimento econômico (Berger *et al.*, 2004). Com a reestruturação dessa indústria no Brasil, em que se sobressaíram importantes fusões, mensurar concentração e concorrência ganhou especial atenção, o que estimulou o desenvolvimento de diversos trabalhos empíricos que buscam responder essas questões (Araujo e Jorge Neto, 2007; Lucinda, 2010; Nakane, 2002; e Nakane e Rocha, 2010). Há, contudo, algo comum a essa literatura que pode gerar eventual distorção nas medidas de competição: a utilização de bancos individuais como unidade de análise e não o conceito de conglomerado financeiro.

Por escolhas internas de governança, um grupo econômico pode se organizar na forma de uma única pessoa jurídica ou, mais comumente, em várias pessoas jurídicas, controladas por uma *holding* ou por arranjos societários ainda mais complexos. Essa escolha não retira a característica essencial do grupo econômico, qual seja, a sua capacidade de coordenar estratégias empresariais, em geral, e concorrenciais, em particular. Por este motivo, é já consensual na jurisprudência antitruste a utilização do conceito de grupo econômico como unidade de análise em uma avaliação das condições de concorrência em determinado mercado (Williamson, 2009). Esse consenso manifesta-se na própria Lei nº 12.529/2011, que dispõe sobre o Sistema Brasileiro de Defesa da Concorrência, seja em seus artigos referentes ao controle de condutas (Arts. 33, 36, 37 e 86), seja naquele referente ao controle de estruturas (Art. 88).

O conceito de grupo econômico é definido pelo controle societário, podendo este incluir também participações em empresas inferiores a 50%, desde que seja assegurado o controle de fato ou pela via de acordo de acionistas. Mesmo no caso de participações societárias minoritárias que não implicam controle, a literatura econômica reconhece que pode haver efeito de alinhamento de incentivos e, portanto, indicadores alternativos para os índices de concentração, a fim de contemplar tais efeitos (O'Brian e Salop, 2000). É digno de nota, portanto, os estudos sobre competição bancária no Brasil utilizam o relatório 4010 do Banco Central, que fornece informações de bancos identificados pelo seu CNPJ, desconsiderando os possíveis efeitos de coordenação entre empresas pertencentes a um mesmo conglomerado financeiro.

Este artigo busca investigar empiricamente se essa desconsideração do grupo econômico produz viés nas estimativas de medidas de competição no mercado bancário brasileiro. Neste artigo são estimados os parâmetros usuais de competição no setor bancário, sugeridos por Bresnahan e Lau (1982), Panzar e Rosse (1987) e

Boone (2008), os quais são calculados considerando duas possíveis definições de unidade econômica, a instituição bancária individual e o conglomerado financeiro para o período de dezembro de 1999 a junho de 2010. O artigo traz também uma contribuição adicional, por se tratar da primeira aplicação do modelo de Boone (2008) à indústria bancária brasileira.

As duas bases de dados utilizadas neste trabalho são providas pelo Banco Central, e dadas por: *i*) a base de dados do relatório 4010, que fornece informações dos balancetes mensais dos bancos, e assim representada por CNPJ diferente da unidade bancária; e *ii*) a base de dados do relatório 4040, que fornece informações dos conglomerados financeiros. Essas duas bases de dados servirão para a realização de um teste de diferença de médias dos parâmetros estimados para conduta no setor bancário. Esse teste de diferença de médias servirá, então, para informar se há diferença estatisticamente significativa da estimação entre os parâmetros de competição bancária estimados com as duas bases de dados dos parâmetros em três diferentes modelos de conduta (Bresnahan e Lau, 1982; Panzar e Rosse, 1987; e Boone, 2008). Conceitualmente a diferença está na unidade de decisão. Empiricamente a desconsideração do conglomerado financeiro pode vir a omitir importantes efeitos relacionados à economia de escopo e escala (Barbosa *et al.*, 2015) e a decisões coordenadas de cada grupo financeiro. Este artigo dedica-se, em particular, a investigar este último efeito, os quais podem ser captados por medidas de intensidade de concorrência.

As estimações do modelo de Boone (2008) mostram a existência de diferença estatística entre os parâmetros de conduta estimados com as duas bases de dados para o período de 1999 a 2010, indicando uma superestimação do grau de concorrência no mercado bancário brasileiro quando a análise é realizada com a base de dados de bancos individuais. Os resultados das estimações da estatística H , de Panzar e Rosse (1987), também revelam outro resultado semelhante. Para a estatística H , para a qual foi possível realizar estimações para cada ano, bem como testar as diferenças para os períodos antes e após 2004, identificamos que, a partir de 2004, há evidências de diferença significativa da estatística H computada com base nas duas bases de dados. Tais resultados revelam que estudos baseados nos modelos de Panzar e Rosse e de Boone e que desconsideram efeitos de coordenação entre empresas do mesmo grupo econômico estão sujeitos a viés de superestimação da intensidade de concorrência bancária brasileira.

Por outro lado, as estimações para o modelo de Bresnahan e Lau (1982), as quais foram realizadas também para o período de 1999 a 2010, não têm apresentado diferenças significantes entre as duas bases de dados, tendo também confirmado resultados anteriores sobre o grau de concorrência na indústria.

Além desta seção introdutória, este artigo conta com a descrição dos modelos de concorrência de Bresnahan e Lau (1982), Panzar e Rosse (1987) e Boone (2008) na seção 2; seguidos da apresentação das estatísticas descritivas na seção 3; da apresentação dos resultados das estimações dos modelos na seção 4; e das considerações finais na seção 5. A seguir, nas duas próximas seções, temos as descrições dos principais modelos e medidas utilizados no setor bancário, e que serão utilizados ao longo deste artigo.

2 INDICADORES DE CONCORRÊNCIA

Não há uma metodologia predominante, nem tampouco um modelo claramente superior quando a variável de interesse é nível de concorrência. Em trabalhos voltados à análise de concorrência bancária, a literatura, conforme apontado por Nakane e Rocha (2010), usualmente apoia-se em três métodos principais: *i*) rotação da curva de demanda – Bresnahan e Lau (1982); *ii*) soma das elasticidades-preço dos insumos produtivos em uma equação de receitas – Panzar e Rosse (1987); e *iii*) identificação de padrões pelas condições de primeira ordem de diferentes soluções de jogos – como exemplos Bertrand-Nash, Cartel, Monopólio, ou Liderança-Preço de Stackelberg.

Este artigo utiliza os modelos *i*) e *ii*), mas não os modelos do tipo *iii*), pois estes estão mais sujeitos a escolhas arbitrárias, decorrentes da grande variedade de especificações e formas funcionais que podem ser adotadas para soluções de jogos. Assim, optou-se por especificações mais consensuais, de modo a focar a discussão nas diferenças de estimações pela utilização das duas diferentes bases de dados. Além dos dois modelos citados, este artigo também aplica o modelo proposto por Boone (2008), que, conforme detalhado mais à frente nesta seção, inspira-se nas ideias iniciais de Panzar e Rosse (1987), mas provê uma mensuração da intensidade de concorrência mais informativa e completa do que a do modelo original. Este artigo traz a primeira aplicação do modelo de Boone (2008) para a indústria bancária brasileira, sendo esta uma de suas contribuições originais.

2.1 Modelo Panzar-Rosse (1987)

Este modelo propõe que no longo prazo, o produto e o número de bancos são determinados, em equilíbrio, pela maximização de lucro ao nível do banco e ao nível da indústria.

A hipótese clássica de maximização de lucro implica a igualdade entre receita marginal e custo marginal. Desse modo, tem-se a seguinte condição de maximização para um banco *i*:

$$RMg_i(x_i, n, x_i) - CMg_i(x_i, w_i, t_i) = 0, \quad (1)$$

em que:

$$\left\{ \begin{array}{l} x_i: \text{produto do banco } i. \\ n: \text{número de bancos} \\ z_i: \text{vetor de variáveis exógenas que alteram a função receita do banco } i. \\ w_i: \text{vetor de } m \text{ preços de fatores insumos do banco } i. \\ t_i: \text{vetor de variáveis exógenas que alteram a função custo do banco } i. \end{array} \right.$$

O grande avanço que as deduções de Panzar e Rosse (1987) trazem está na definição de uma medida pela qual possam ser identificados os casos de conluio perfeito, concorrência monopolística ou concorrência perfeita. Esta medida é definida como estatística H e dada por:

$$H = \sum_{k=1}^m \frac{\partial R_i}{\partial w_{k_i}} \cdot \frac{w_{k_i}}{R_i}. \quad (2)$$

A expressão da equação (2) representa o somatório das elasticidades das receitas em relação aos preços dos insumos. Em outras palavras, a estatística H pode ser descrita como a soma das elasticidades receitas-custos.

O quadro 1 foi construído de acordo com as implicações testáveis de Panzar e Rosse (1987) e relaciona o valor estimado da estatística H ao grau e tipo de concorrência numa indústria/mercado. Tais autores mostram que existem três intervalos de valores distintos para a estatística H, que tipificam a forma de concorrência que o setor em análise se encontra. Em particular, Panzar e Rosse mostram que se o valor estimado de H é menor ou igual a 0, então a indústria opera num equilíbrio de monopólio (exercício pleno de poder de mercado), ou concorrência monopolística (algum exercício de poder de mercado). Por outro lado, se o valor estimado de H é igual a 1, então a indústria opera em concorrência perfeita. Por fim, se o valor estimado de H estiver entre 0 e 1, então os bancos operam num equilíbrio de concorrência monopolística.

As derivações originais de Panzar e Rosse (1987), entretanto, não lhes permitem obter uma relação monotônica entre o valor do parâmetro H e o grau de concorrência de um mercado. Todavia, Shaffer (1982, 1983) e Vesala (1995) estendem o modelo Panzar e Rosse e mostram que, sob certas condições, o parâmetro H cresce com o grau de concorrência. Em particular, Vesala (1995) mostra a existência de uma relação negativa entre o parâmetro H e o poder de mercado no caso de concorrência monopolística com livre entrada.

Assim, com base em Shaffer (1982, 1983) e Vesala (1995), podemos concluir que um maior valor de H está associado a um aumento do grau de concorrência.

QUADRO 1
Intervalos de Valores de H (Panzar-Rosse)

Valores de H	Tipo de concorrência
$H \leq 0$	Equilíbrio de monopólio (exercício pleno de poder de mercado), ou concorrência monopolística (algum exercício de poder de mercado)
$H \leq 1$	Concorrência monopolística (algum exercício de poder de mercado)
$H = 1$	Concorrência perfeita (Não há exercício do poder de mercado)

Fonte: Elaboração própria a partir de Panzar e Rosse (1987).

Uma forma usual para obter a estatística H é dada pela estimação da seguinte equação função receita:⁴

$$\underbrace{\text{Ln}(RT_{it})}_{\text{Receita operacional e não operacional}} = \alpha + \underbrace{\text{Ln}(w_{it})' \cdot \beta}_{\text{Preços dos insumos da atividade bancária}} + \underbrace{\text{Ln}(z_{it})' \cdot \gamma}_{\text{Variáveis de controle}} + \mu_i + \varepsilon_{it}. \quad (3)$$

Do lado esquerdo dessa equação, a variável $RT_{i,t}$ representa receitas totais (operacionais e não operacionais), correspondendo à soma das receitas operacionais (atividades principais) do banco i , no instante t , mais as receitas não operacionais que advêm de eventos econômicos aditivos ao patrimônio líquido não associados às atividades principais do banco, como, por exemplo, ganhos de capital, transações com imobilizados ou com investimentos de natureza permanente.

A variável $w_{i,t}$ corresponde aos fatores tratados como insumos da atividade bancária, sendo que a soma dos coeficientes estimados para esses fatores gera a estatística H de interesse. Uma forma de comum aceitação, com a maior referência em Bikker e Haaf (2002), é definir três principais fatores $w_{i,t}$ como sendo relacionados a despesas administrativas, despesas operacionais e despesas de captação. O termo μ_i na equação (3) é o efeito fixo do banco, e ε_{it} é o termo errático. Essa especificação será detalhada na seção 4.1.

A estimação da equação função receita em (3) é relativamente simples, mas requer cautela na construção das variáveis e na estimação dos coeficientes. Caso contrário, estimadores tendenciosos de H podem ser obtidos. Bikker, Shaffer e Spierdijk (2012), por exemplo, mostram que a inclusão da receita normalizada por uma medida de escala (por exemplo, ativo) como variável dependente na

4. Essa forma de estimação do modelo Panzar-Rosse (1987) tem como principais referências os trabalhos de Bikker e Haaf (2002), e Belaisch (2003), e no âmbito nacional Araujo e Jorge Neto (2007), Lucinda (2010), e Nakane e Rocha (2010).

equação função receita em (3) e a inclusão de variáveis de escala na regressão do modelo de Panzar e Rosse podem superestimar o parâmetro H, gerando assim uma medida incorreta do grau de competição na indústria, e diferente da medida de concorrência originalmente proposta por Panzar e Rosse (1987). Este artigo segue as sugestões de Bikker, Shaffer e Spierdijk e estima o parâmetro H utilizando a receita não normalizada do banco, para não ocorrer o problema de viés de estimação destacado anteriormente.

2.2 Modelo Boone (2008)

Boone (2008) introduz uma forma direta de mensuração de concorrência com o conceito de elasticidade do lucro, primeiramente desenvolvido em Boone (2000). Não existe um parâmetro fechado em intervalos como em demais modelos, mas sim o conceito de magnitude de variação no lucro, dadas as variações nos custos marginais dos bancos.

Como referido em Boone, Ours e Wierl (2007), a mensuração da elasticidade do lucro é inspirada nas ideias de Panzar e Rosse (1987). Essa mensuração é baseada na elasticidade preço dos fatores, e a forma funcional direta de mensurar competitividade, segundo esse modelo, é dada por:

$$\text{Ln}(\pi_{it}) = \alpha - \beta \cdot \text{Ln}(c_{it}) + \mu_i + \varepsilon_{it}. \quad (4)$$

As variáveis π_{it} e c_{it} representam, respectivamente, o lucro e o custo marginal do banco i no período t . Este modelo pode ser estimado com dados em painel⁵, e as especificações empíricas incluem variáveis de controle na especificação. O termo μ_i na equação (4) é o efeito fixo do banco e ε_{it} é o termo errático. Custo marginal, nesse caso, pode ser obtido de diversas formas e, talvez, a mais usual seja estimar uma função custo total translog para então encontrar o custo marginal.⁶

A inclinação de β , segundo uma aproximação de primeira ordem de Taylor, pode ser interpretada como uma elasticidade. Segundo Leuvensteijn *et al.* (2007), este indicador de Boone deriva de duas noções gerais: *i*) primeiro, que firmas mais eficientes tendem a ter maior *market share* ou lucro, justamente por observarem um menor custo marginal em suas operações; e, *ii*) segundo, que os efeitos devem ser mais fortes em mercados que tenham níveis de concorrência maior.⁷

5. Como a forma proposta é do tipo log-log, então β assume interpretação de elasticidade, e, por relacionar lucro e custo marginal, recebe a denominação de elasticidade lucro.

6. A forma que adotaremos tanto para o modelo Boone como para o modelo Bresnahan-Lau será a mesma especificação adotada em Shaffer (1993).

7. Um amplo conjunto de modelos e conceitos é desenvolvido por Boone nos seguintes trabalhos: Boone (2000), Boone (2004), e Boone *et al.* (2004).

De acordo com o modelo, os diferentes níveis de elasticidade do lucro estão associados a níveis diferentes de competição via dois efeitos: *i*) a mobilidade dos consumidores de serviços bancários entre diferentes bancos ser mais fácil, dando um sentido de produtos mais próximos; e *ii*) maior ou menor custo de entrada de concorrentes potenciais.

2.3 Modelo Bresnahan-Lau (1982)

O modelo de Bresnahan e Lau (1982) estima equações de demanda e oferta para o total de empréstimos do mercado bancário, adicionando um parâmetro de conjectura do qual se infere o poder de mercado e, assim, o grau de competição da indústria. Essa metodologia foi utilizada em diversos países, tendo sido aplicada no Brasil no estudo pioneiro de Nakane (2002).

Talvez a principal crítica a esta literatura seja circunscrever a análise de competição da indústria bancária à competição por empréstimos, sem a adequada separação de segmentos, como varejo e atacado e de pessoa física e pessoa jurídica. Infelizmente, os dados disponíveis não permitem essas separações. Ainda assim, o argumento de utilizar empréstimo tem, pelo menos, uma boa justificativa. Quando observados os dados agregados dos bancos, cerca de um terço de todos os ativos destinam-se a empréstimos – a atividade de intermediação financeira presume-se ser a principal atividade de um banco e os dados assim corroboram.

O mecanismo de identificação do grau de concorrência é dado pela separação de variáveis de rotação e variáveis de deslocamento de oferta e demanda. O preço de equilíbrio de oligopólio é então caracterizado pela seguinte condição de primeira ordem:

$$P(Q, Y; \alpha) + \lambda \cdot Q \cdot P(Q, Y; \alpha) = C(Q, Z; \beta), \quad (5)$$

na qual:

$$\left\{ \begin{array}{l} P: \text{preço de equilíbrio de mercado} \\ P(Q, Y; \alpha): \text{é a função demanda inversa} \\ Q: \text{é o nível de quantidade de produto bancário (p.ex., empréstimos)} \\ C(Q, Z; \beta): \text{é o custo marginal} \\ \alpha \text{ e } \beta: \text{valores de parâmetros desconhecidos associados com demanda e custos} \\ Y \text{ e } Z: \text{são vetores de variáveis que afetam demanda e custos} \\ \lambda: \text{é o parâmetro de variação conjectural} \end{array} \right.$$

Os valores de λ estão associados, por exemplo, a mudanças na oferta de um banco como uma função resposta de todos os demais bancos. Assim, essa metodologia mensura o coeficiente que sintetiza o poder de mercado e suas variações, analisando as variáveis de rotação da demanda.

As formas funcionais adotadas, neste trabalho, para as curvas de demanda, oferta e custos, são as mesmas utilizadas em Shaffer (1993). Essas formas são descritas na seção de resultados estimados.

3 ESTATÍSTICAS DESCRITIVAS

Esta seção é dedicada à apresentação das principais medidas que serão utilizadas para medir concentração e concorrência utilizando os modelos descritos na seção anterior para as bases de dados do relatório 4010 (CNPJ) e do relatório 4040 (conglomerado).

Cabe aqui uma apresentação das diferenças entre essas duas bases de dados e uma explicação sobre a construção destas em sua origem, pelo Bacen. A diferença essencial entre o relatório 4010 e o relatório 4040 é que no primeiro a unidade de agregação é o CNPJ de cada instituição financeira, de tal modo que as informações correspondem fielmente ao balancete como entidade contábil que o Bacen fiscaliza. No segundo, por sua vez, a unidade de agregação é o conglomerado financeiro, que reúne os CNPJs que estão sob o controle de um mesmo grupo econômico. Para tanto, o Bacen utiliza-se do conceito de controle societário, dado pela estrutura societária de cada instituição financeira, de modo a identificar os CNPJs que pertencem a um mesmo grupo, agregando-os sob o título de Conglomerado Financeiro.

Assim, o relatório 4040 não é uma mera soma dos registros contábeis de todas as empresas coligadas e cointegradas dos conglomerados financeiros, mas é uma construção do Bacen, que incorpora todos os ajustes e compensações reportados à autoridade monetária. Como os dados a respeito desses ajustes e compensações não são disponíveis ao público, não é possível ao pesquisador reconstruir a base de conglomerados financeiros com base no relatório 4010. Por isso, é necessária a utilização do relatório 4040 como a base de conglomerados financeiros, cujos dados já consideram esses ajustes e compensações.

As tabelas 1 e 2 apresentam as estatísticas descritivas de empréstimos totais, ativos totais, depósitos totais e à vista, respectivamente, para as bases de dados do relatório 4010 (CNPJ) e do relatório 4040 (conglomerado).

TABELA 1
Estatística descritiva – relatório 4010

	Empréstimos	Ativos	Depósitos	Depósitos à vista
Mínimo	-	1,26	-	-
Média	335.741.429,47	13.122.204.602,58	5.146.931.097,41	585.777.805,42
Mediana	22.032.333,48	1.264.482.074,86	212.186.553,99	1.549.192,62
Máxima	4.964.200.000,00	651.760.000.000,00	313.595.477.681,66	55.808.858.955,96
Desvio-padrão	13.769.590.256,12	49.653.024.532,06	22.695.680.789,22	3.350.918.864,40

Fonte: Elaboração própria.

Obs.: Todos os valores (em R\$) foram deflacionados pelo IPCA. Período: dez./1999 a jun./2010 com dados trimestrais.

TABELA 2
Estatística descritiva – relatório 4040

	Empréstimos	Ativos	Depósitos	Depósitos à vista
Mínimo	-	6.197.506,45	-	-
Média	213.596.103,26	14.831.523.792,46	5.897.160.067,42	824.239.153,20
Mediana	5.023.414,19	1.191.760.896,33	250.793.826,42	4.162.637,96
Máxima	26.329.096.126,84	629.250.680.000,00	348.279.547.863,14	57.776.755.335,13
Desvio-padrão	1.063.407,53	53.029.912,09	24.642.948,64	4.136.014,16

Fonte: Elaboração própria.

Obs.: Todos os valores (em R\$) foram deflacionados pelo IPCA. Período: dez./1999 a jun./2010 com dados trimestrais.

Tomando como referência os valores médios e máximos de ativos ou de empréstimos nas tabelas 1 e 2, observam-se casos em que a base CNPJ (4010) tem valores maiores que a base do conglomerado (4040), como, por exemplo, a média de empréstimos ou o máximo de ativos. Ao mesmo tempo, existem casos em que a base conglomerado tem valores maiores que a base CNPJ (4010), como, por exemplo, o máximo de empréstimos ou a média de ativos.

Tais diferenças podem ser explicadas por dois principais motivos: *i)* no relatório 4010, bancos de um mesmo grupo têm seus balancetes lançados mensalmente, o que pode ocasionar problema de dupla contagem, por exemplo, após um processo de fusão quando uma das instituições não passará mais a emitir balancete. Isso porque pode existir um período após a fusão em que ainda fique sendo lançado o balancete do banco incorporado, e não existe uma regra única para o período de ajuste dos balancetes; *ii)* existem ajustes entre as empresas coligadas e integradas que apenas o Bacen tem acesso e realiza assim, o relatório 4040 representa essa consolidação dos ajustes.

Para exemplificar, observemos o caso da fusão do Banco do Brasil (CNPJ número 00.000.000) com o Banco Nossa Caixa (CNPJ 43.073.394). O anúncio da fusão foi realizado em 20/11/2008, sendo que até outubro/2009

os balancetes do Nossa Caixa continuaram a ser emitidos. A tabela 3 apresenta as contas para o ativo do Nossa Caixa a partir do relatório 4010, em comparação com os dados disponíveis no relatório 4040.

TABELA 3
Estatística descritiva para ativos totais – relatórios 4010 e 4040

Relatório 4010			Relatório 4040	
Data	Compensação 30000001	Total geral ativos 39999993	Ativo (R\$)	Ativo (em R\$ mil)
Dez./1999	25.760.821.708,64	41.208.714.255,30	15.447.892.546,66	15.447.893,00
Mar./2000	62.649.777.538,67	78.317.768.617,02	15.667.991.078,35	15.667.991,00
Jun./2000	40.537.057.735,64	56.927.368.221,91	16.390.310.486,27	16.390.310,00

Fonte: Elaboração própria. Todos os valores (em R\$) foram deflacionados pelo IPCA.

Em que: 30000001 e 39999993 são respectivamente os números das contas do Cosif (relatório contábil padronizado pelo Bacen) referentes às contas de compensação e total geral ativos.⁸ Período: dez./1999 a jun./2010 com dados trimestrais.

Pode-se notar que, ao realizar a conta total geral ativo (39999993) menos compensação (30000001), obtém-se o mesmo resultado para ativo total com valores do relatório 4010 e 4040. Mas isso só ocorre porque o Banco Nossa Caixa até nov./2008 tinha sua contabilidade independente de qualquer outro banco coligado ou integrado.

Ainda nesse exemplo, se fosse construída a conta para ativo total para o Banco do Brasil com os dados a partir do relatório 4010, não seria possível reproduzir os mesmos valores encontrados no relatório 4040. Isso ocorre para qualquer banco pertencente a algum conglomerado financeiro.

Apesar dessas questões contábeis, a utilização dos dados dos relatórios 4010 e 4040 não necessariamente geraram diferença relevante em medidas de intensidade de concorrência, seja porque não há propriamente diferenças substanciais entre as duas bases de dados, seja porque as entidades jurídicas independentes, definidas pelos seus CNPJs, podem se comportar de modo autônomo. Não haveria, portanto, nesse caso, prejuízo à análise considerar que diferentes instituições financeiras, definidas pelos seus CNPJs, concorrem independentemente. Esta é a hipótese que é testada por este artigo, por meio da comparação entre os três modelos de competição com o uso das duas bases de dados.

8. Lucinda (2010), e Nakane e Rocha (2010) utilizam outras contas do Cosif para fazer o cálculo do ativo, tendo ainda uma divergência entre si quanto ao sinal. Segundo Lucinda (2010) o ativo total pode ser obtido com a conta circulante e realizável a longo prazo (10000007) mais permanente (20000004), o que para o exemplo desta tabela confere o mesmo resultado. Enquanto Nakane e Rocha (2010) descrevem como (10000007) (+) e (20000004) (-) a conta para obtenção do Ativo Total, o que diverge do exemplo desta tabela.

As tabelas 4 e 5, a seguir, apresentam estatísticas descritivas a respeito das variáveis razão de concentração (C_n)⁹, em relação a ativos totais e depósitos à vista.

TABELA 4
Estatística descritiva para ativos totais – relatórios 4010 e 4040

	4010			4040		
	Top 3	Top 5	Top 10	Top 3	Top 5	Top 10
Mínimo	0,3522	0,5189	0,6852	0,356	0,522	0,679
Média	0,3907	0,5744	0,7439	0,407	0,593	0,775
Mediana	0,3808	0,5708	0,7366	0,395	0,573	0,779
Máxima	0,4738	0,6532	0,8216	0,527	0,754	0,866
Desvio-padrão	0,0275	0,0338	0,0369	0,052	0,069	0,050

Fonte: Elaboração própria.
Período: Dez./1999 a jun./2010 com dados trimestrais.

TABELA 5
Estatística descritiva para depósitos à vista – relatórios 4010 e 4040

	4010			4040		
	Top 3	Top 5	Top 10	Top 3	Top 5	Top 10
Mínimo	0,558	0,692	0,831	0,526	0,654	0,846
Média	0,618	0,752	0,900	0,623	0,760	0,909
Mediana	0,615	0,737	0,896	0,625	0,751	0,913
Máxima	0,677	0,876	0,967	0,710	0,886	0,968
Desvio-padrão	0,029	0,051	0,036	0,042	0,058	0,033

Fonte: Elaboração própria.
Período: Dez./1999 a jun./2010 com dados trimestrais.

Seguindo a ótica do conglomerado financeiro, podemos esperar que os dados da média ou valor máximo, representados nas tabelas acima, oferecessem sempre valores maiores ou iguais para o relatório 4040, visto que sua unidade de análise é uma agregação daquela utilizada no relatório 4010. Isso se verifica em geral, mas para os valores mínimos encontramos medidas ligeiramente menores para top 3 e top 5 dos depósitos à vista dos conglomerados financeiros. Esses valores mínimos menores na base 4040, em comparação com a base 4010, ocorrem no terceiro trimestre de 2000. Não são os únicos, mas são pequenas diferenças que surgem nos dados do início da década e apenas para alguns trimestres.¹⁰

9. A razão de concentração (C_n) mede de forma direta a soma de participações de mercado das n maiores firmas no setor. Assim, o índice C_4 corresponde à soma das participações de mercado das quatro maiores firmas no setor analisado, assim como o C_5 é a soma das cinco maiores participações de mercado e assim por diante. Formalmente este índice pode ser expresso por $C_n = \sum_{i=1}^n \beta_i$, na qual β_i é a participação de mercado do banco i .

10. Vale observar que, durante todo o período analisado, em ambos os relatórios, os três maiores bancos foram Banco do Brasil, Bradesco e Itaú.

A seguir tem-se a comparação descritiva para índice Herfindahl–Hirschman¹¹ (HHI) em relação aos ativos, empréstimos, depósitos totais e à vista.

TABELA 6
Estatística descritiva HHI – relatório 4010

	HHI_ativo	HHI_empréstimos	HHI_depósitos	HHI_depósitos à vista
Mínimo	0,056	0,056	0,081	0,122
Média	0,067	0,077	0,089	0,159
Mediana	0,065	0,075	0,085	0,162
Máxima	0,089	0,105	0,116	0,202
Desvio-padrão	0,008	0,013	0,008	0,019

TABELA 7
Estatística descritiva HHI – relatório 4040

	HHI_ativo	HHI_empréstimos	HHI_depósitos	HHI_depósitos à vista	HHI_patrimônio líquido
Mínimo	0,068	0,065	0,097	0,124	0,056
Média	0,263	0,091	0,112	0,164	0,072
Mediana	0,080	0,085	0,108	0,164	0,066
Máxima	1,903	0,134	0,146	0,201	0,120
Desvio-padrão	0,513	0,019	0,014	0,019	0,019

Fonte: Elaboração própria.

HHI é a medida de concentração de Hirfindhal para ativos, empréstimos, depósitos totais, depósitos à vista, e patrimônio líquido.

Período: Dez./1999 a jun./2010 com dados trimestrais.

As medidas de HHI para estas duas últimas tabelas, quando apresentamos a estatística descritiva para comparação das bases 4010 e 4040, revelam consistentemente o resultado esperado, ou seja, os valores de concentração medidos na base de dado conglomerado financeiro são mais elevados do que na base de dados de CNPJ.

Finalmente, a tabela 8 apresenta as estatísticas descritivas das variáveis utilizadas no modelo de Panzar e Rosse (1987). Pode-se notar que há uma maior diversificação das variáveis construídas, com base no relatório 4010, e que os valores da média, mediana e máximo também são maiores para todas as variáveis, exceto para despesas de captação.

11. O índice HHI é uma medida de tamanho das empresas em relação à indústria e um indicador de concentração. Formalmente este índice pode ser expresso por: $HHI = \sum_{i=1}^n \beta_i^2$, na qual β_i é a participação de mercado do banco i .

TABELA 8
Estatística descritiva – variáveis do modelo P&R – relatórios 4010 e 4040

	4010				4040			
	R_tot	D_work	D_adm	D_cap	R_tot	D_work	D_adm	D_cap
Mínimo	-3,613	-9,163	-0,151	-7,342	1,218	-6,174	0,961	1,443
Média	19,214	13,230	7,938	9,866	11,457	3,803	6,842	10,530
Mediana	19,295	13,168	8,048	10,531	11,441	3,872	6,856	10,604
Máxima	25,797	21,055	14,509	20,484	17,968	8,892	10,456	13,992
Desvio-padrão	2,588	2,360	1,492	2,805	2,300	1,313	1,205	0,920

Fonte: Elaboração própria.

Em que: R_tot é LN das receitas totais, D_work é LN das despesas com funcionários, D_adm é o LN das despesas administrativas, e D_cap é o LN das despesas de captação.

Período: Dez./1999 a jun./2010 com dados trimestrais.

4 RESULTADOS

As próximas subseções destinam-se à apresentação das especificações adotadas, apresentação das bases de dados utilizadas e discussão dos resultados. As descrições sempre seguirão a apresentação dos testes para base de dados do relatório 4010, com a denominação neste trabalho de base CNPJ, seguida dos testes para a base de dados do relatório 4040, denominada neste trabalho de base conglomerado.

4.1 Panzar & Rosse (1987)

Na discussão sobre utilização de medidas sobre a competição bancária, o modelo Panzar e Rosse (1987) foi o primeiro a motivar a pesquisa e o debate desenvolvidos neste trabalho sobre o caso brasileiro.

É importante notar que há diversas estimações do nível de concorrência na indústria bancária brasileira por meio do modelo de Panzar e Rosse, como nos trabalhos Araújo e Jorge Neto (2007), Lucinda (2010), e Nakane e Rocha (2010).¹² Entretanto, todos esses trabalhos utilizam a base de dados de bancos individuais (relatório 4010 do Bacen) como unidade de análise. Assim, esses trabalhos podem estar sujeitos a viés por desconsiderarem os efeitos de coordenação em empresas pertencentes a um mesmo conglomerado financeiro.

As grandes referências internacionais comuns para o uso do modelo de Panzar e Rosse são Bikker e Haaf (2002) e Belaisch (2003), na definição de três fatores que sumariam os efeitos de elasticidade custo da receita, dados por: despesas administrativas, despesas operacionais e despesas de captação.

12. Em termos de período de análise, estes três trabalhos seguem, respectivamente, dezembro/1995 a junho/2004, 2000 a 2005, e 2001 a 2009. Nakane e Rocha (2010), em particular, realizam estimações com dados em *cross-section*. Diferentemente, neste artigo o modelo de Panzar e Rosse é estimado com dados em painel. Na seção 4.4 será realizado um teste com o modelo de Panzar-Rosse, com estimações de médias móveis para a estatística H.

Para a composição das variáveis que compõem a estatística H, foram utilizadas as despesas correspondentes ao custo do trabalho, despesas administrativas e despesas de captação, com a seguinte construção:

QUADRO 2
Descrição das variáveis¹³ de H-Panzar e Rosse

Variável da regressão	Construção da variável
D_work	$Ln\left(\frac{\text{Despesas Pessoal}}{n^{\circ} \text{ funcionários}}\right)$
D_adm	$Ln\left(\frac{\text{Despesas Administrativas}}{\text{Ativos}}\right)$
D_cap	$Ln\left(\frac{\text{Despesas Captação} + \text{Operacionais}}{\text{Passivo} + \text{Permanente}}\right)$

Fonte: Elaboração própria.

Nota: Variáveis deflacionadas pelo IPCA.

Os dados utilizados foram providos pelos relatórios 4010 (base CNPJ) ou 4040 (base conglomerado) do Banco Central do Brasil. Foi utilizado IPCA da base de dados do IBGE para as variáveis, como receita total, que precisaram sofrer deflação.

Panzar & Rosse (1987) – CNPJ. A equação estimada em dados de painel, com efeitos fixos, é dada para este caso por:

$$Ln(RT_{it}) = \alpha + \beta_1 \cdot Ln(d_work) + \beta_2 \cdot Ln(d_adm_{it}) + \beta_3 \cdot Ln(d_cap_{it}) + Ln(Z_{it})' \cdot \gamma + \mu_i + \varepsilon_{it} \quad (6)$$

na qual z_{it} são as variáveis de controle, o termo μ_i é o efeito fixo do banco e ε_{it} é o termo errático. Para esses testes, foram utilizadas as seguintes variáveis: *i*) (vc_11) como variabilidade do lucro do banco *i*; *ii*) (HHI_{DEP_VISTA})¹⁴ como índice de concentração de depósitos à vista; *iii*) ($Desp_NOper$) como somatório de despesas não operacionais; e *iv*) (z_n) como lucratividade padronizada pela estatística normal de cada banco *i*. Os resultados para o período de dezembro/1999 a junho/2006 seguem na tabela 9.

13. Não existe uma forma consensual na construção dessas variáveis. Por exemplo, ao olhar cuidadosamente os referidos trabalhos que se utilizam desta metodologia, não é possível extrair uma forma única. O importante, independentemente da forma de elaborar a construção dessas variáveis, é conseguir distinguir a separação dos fatores de custo que entram como elasticidade custo da receita.

14. As medidas de HHI para depósitos e ativos são utilizadas nos testes desta subseção e demais seções deste artigo e do próximo. Esta variável é o meio mais consagrado de mensurar a concentração de mercado, o que, por sua vez, é uma variável relevante para explicar a conduta das firmas, como se propõe o modelo de Panzar e Rosse (1987). É importante reportar que há críticas a sua utilização como variável explicativa, dado ser a estrutura de mercado endógena, gerando um problema de identificação (Martin, 2002: 173).

TABELA 9

Testes com o modelo Panzar-Rosse sobre a base de dados do relatório 4010 (CNPJ)
(Variável dependente: LN da receita total)

	Modelo (1)	Modelo (2)
D_work	0.221 (0.001)***	0.218 (0.001)***
D_adm	-0.141 (0.054)***	-0.036 (0.055)
D_cap	0.270 (0.041)***	0.202 (0.041)***
VC_11	1.595 (0.032)***	
(HHI) _{dep_vista}	1.553 (0.222)***	1.334 (0.225)***
Desp _{NOper}		7.791 (1.705)***
Z_N		3.712 (0.077)***
Constante	7.208 (0.079)***	7.393 (0.082)***
Nº observações	6.559	6.559
Nº bancos	223	223
R2 (within)	0.86	0.85
H-panzar	0.350	0.384
DP(H)	0.066	0.067
T-stat	5.288	5.727
H=1 (p-value)	0.000	0.000
H=0 (p-value)	0.000	0.000

(Em parênteses estão os desvios-padrão robustos)

* Significativo a 10%; ** Significativo a 5%; *** Significativo a 1%

Fonte: Elaboração própria.

Para as especificações testadas, há coeficientes positivos para (d_work) e (d_cap), e negativo para (d_adm), o que corresponde aos sinais esperados, segundo a construção das variáveis. Assim, ao aumentar gastos com funcionários ou com captação, o banco deve esperar um incremento na receita. E ao contrário, quando os custos administrativos aumentam, isso deve acarretar um decréscimo da receita do banco, dando sentido de ineficiência.

Vale destacar que foram adotadas as sugestões de Bikker, Shaffer e Spierdijk (2012) na inferência do parâmetro H de Panzar e Rosse, o qual foi obtido por meio de estimação da equação (6). Isto é, foi utilizada a receita não normalizada do banco como variável dependente na estimação de (6) e o valor ativo do banco (e nenhuma outra variável como *proxy* para a escala do banco) foi incluída como

variável de controle naquela regressão. Assim, a nossa estimação do parâmetro H não padece dos problemas de viés de estimação destacados por Bikker, Shaffer e Spierdijk.¹⁵

Os valores estimados para o parâmetro H ($H = 0,35$ e $H=0,384$), reportados na tabela 9, revelam que o mercado bancário brasileiro de dezembro/1999 a junho/2010 operou, em média, em equilíbrio de concorrência monopolística. Esse resultado é interessante, pois revela que o tipo de concorrência estimado nesse artigo pela estatística H, e com base de dados do relatório 4010, não difere qualitativamente do tipo de concorrência estimado por Araújo e Jorge Neto (2007), para o período de junho/1999 a junho/2004, (estatística H média de 0,737)¹⁶ e por Nakane e Rocha (2010), para o período de junho/2001 a junho/2010, que encontram uma estatística H de 0,797.

Contudo, quantitativamente obtemos resultados diferentes dos obtidos por Araújo e Jorge Neto (2007) e Nakane e Rocha (2010), uma vez que as nossas estimativas do parâmetro H são significativamente menores que os valores estimativos por tais autores. Tal resultado é particularmente importante quando se leva em conta o trabalho de Vesala (1995), no qual tal autor mostra que o parâmetro H cresce com o grau de concorrência de um mercado. Assim, os resultados das estimações do parâmetro H revelam que o grau de concorrência estimado neste artigo para o setor bancário brasileiro é menor que o grau de concorrência estimado para o mesmo setor por aqueles autores.

Lucinda (2010), em contrapartida, estimou a estatística H igual a 0,85 e 0,497 para o período do primeiro trimestre de 2000 ao quarto trimestre de 2005, com dados trimestrais, utilizando duas diferentes especificações de variáveis de controle. Vale notar que, em pelo menos uma das estimações de Lucinda (2010), a estimativa pontual do parâmetro H é semelhante a nossa estimativa de H.¹⁷ Possivelmente, o fato de Lucinda (2010) ter estimado a equação de Panzar e Rosse com dados em

15. No modelo 2 da tabela 2, reportamos as estimações da equação (6), quando a variável despesas não operacionais (Desp_NOper) é utilizada como variável de controle. As despesas não operacionais são despesas que ocorrem ocasionalmente e não fazem parte das atividades operacionais do banco. Portanto, não é uma *proxy* para a escala de um banco. Note que não há diferença estatística entre o parâmetro H estimado por meio do modelo (1), sem inclusão daquelas despesas como variável de controle, e o modelo (2), com a inclusão daquelas despesas como variável de controle. Esse resultado demonstra que a inclusão da variável Desp_NOper, como variável dependente na equação (6), não afeta o valor e o desvio-padrão estimados para a estatística H de Panzar e Rosse na tabela 2.

16. Araújo e Jorge Neto (2007) apresentam a estatística H para cada semestre no período de estudo, pois foram realizadas diversas regressões para cada período em vez de dados em painel.

17. Lucinda (2010) traz uma discussão interessante ao montar as estimações da estatística H para toda a amostra e diferentes subamostras definidas por: grandes, médios e pequenos bancos, bancos comerciais, e bancos de investimento. A divisão de porte do banco para Lucinda (2010) segue: pequeno, médio, grande porte referente à média dos ativos respectivamente nos seguintes intervalos de valores até 250 milhões, de 250 a 5.000 milhões, e acima de 5.000 milhões de reais.

painel e efeito fixo, tal como fizemos neste artigo, tenha permitido que as nossas estimativas sejam similares.¹⁸

Assim, uma primeira conclusão que se pode derivar dessa seção é que, independentemente da forma de estimação e da base de dados utilizada, tanto as nossas estimações como as estimações dos trabalhos existentes na literatura do parâmetro de conduta H de Panzar e Rosse indicam que a indústria bancária brasileira opera em concorrência monopolística, uma vez que, conforme o quadro 1, qualquer valor da estatística H entre 0 e 1 é indicativo de padrão de concorrência. Esse mesmo resultado repete-se neste modelo por diferentes especificações, e nos outros modelos também, como será mostrado nas demais subseções.

Por outro lado, as estimativas deste artigo para o parâmetro H são menores que os estimados pelos demais artigos da literatura. Assim, as estimativas de H revelam que o grau de concorrência estimado para o setor bancário brasileiro neste artigo é menor que o grau de concorrência estimado para o mesmo setor por aqueles autores, tendo em vista a existência de uma relação negativa entre o parâmetro H e o poder de mercado, no caso de concorrência monopolística com livre entrada (conforme Vesala, 1995).

Panzar & Rosse (1987) – conglomerado. Para a base dos conglomerados bancários, foi aplicada a mesma especificação da subseção anterior com o uso de variáveis de controle diferentes.

A equação estimada em dados de painel, com efeitos fixos, é dada para este caso por:

$$\begin{aligned} \ln(RT_{it}) = & \alpha + \beta_1 \cdot \ln(d_work_{it}) + \beta_2 \cdot \ln(d_adm_{it}) + \beta_3 \cdot \ln(d_cap_{it}) + \\ & \ln(Z_{it})' \cdot \gamma + \mu_i + \varepsilon_{it} \end{aligned} \quad (7)$$

na qual z_i são as variáveis de controle, o termo μ_i é o efeito fixo do banco e ε_{it} é o termo errático. Para esses testes, utilizamos as seguintes variáveis: (vc_II_{it}) como variabilidade do lucro do banco i no período t , (d_01) como *dummy* para diferença entre conglomerado bancário I ou II, sendo que a definição técnica de conglomerado bancário I corresponde ao grupo que tem ao menos um banco com carteira comercial e tipo II que não possui carteira comercial. As demais variáveis de controle: (d_cons) como *dummy* para distinguir banco correspondente à conglomerado financeiro de banco independente; (d_03), (d_06) e (d_09) *dummy* para marcar o trimestre correspondente do ano; ($HHI_{DEP_VISTA_t}$) como índice de concentração de depósitos

18. Uma possível explicação para essa diferença nos valores estimativos do parâmetro H é que Araújo e Neto (2007) e Nakano e Rocha (2010) utilizam períodos diferentes dos utilizados neste artigo. Outra possível explicação é que aqueles autores utilizam dados em *cross-section* na estimação de H, ao passo que este artigo utiliza dados em painel. A omissão do efeito fixo do banco, fruto da utilização de dados em *cross-section*, pode trazer prejuízo à estimação do H e mensuração do grau de concorrência (viés de variável omitida) se as características do banco que são invariantes no tempo forem correlacionadas com os preços dos insumos (variáveis dependentes na equação de Panzar e Rosse).

à vista, $(HHI_{ATIVO,t})$ como índice de concentração de ativos. Os resultados para o período de dezembro/1999 a junho/2006 seguem na tabela 10.

TABELA 10

Testes com o Modelo Panzar-Rosse sobre a base de dados do relatório 4040 (conglomerado bancário)
(Variável dependente: LN da receita total)

	Modelo (1)	Modelo (2)	Modelo (3)
D_work	0.125 (0.016)***	0.114 (0.016)***	0.139 (0.016)***
D_adm	-0.203 (0.017)***	-0.195 (0.017)***	-0.190 (0.017)***
D_cap	0.320 (0.015)***	0.343 (0.015)***	0.349 (0.015)***
VC_11	0.596 (0.113)***	0.599 (0.112)***	0.694 (0.113)***
D_01	0.396 (0.074)***	0.354 (0.074)***	
D_cons	0.651 (0.054)***	0.628 (0.053)***	
D_03	-0.577 (0.030)***	-0.560 (0.030)***	-0.539 (0.031)***
D_06	0.051 (0.022)**	0.057 (0.022)**	0.059 (0.023)***
D_09	-0.503 (0.030)***	-0.479 (0.030)***	-0.445 (0.030)***
$(HHI)_{ATIVO}$		4.890 (0.534)***	
$(HHI)_{DEP_VISTA}$			4.960 (0.481)***
Constante	8.661 (0.208)***	8.037 (0.217)***	8.018 (0.226)***
Nº observações	6035	6035	6035
Nº bancos	260	260	260
R ² (within)	0.34	0.35	0.34
H-Panzar	0.243	0.262	0.298
DP(H)	0.026	0.026	0.264
T-stat	9.221	9.990	1.129
H=1 (p-value)	0.000	0.000	0.000
H=0 (p-value)	0.000	0.000	0.000

Fonte: Elaboração própria.

Na avaliação do valor da estatística H, na base conglomerado, obtém-se o valor 0,24 para a primeira especificação da tabela 10, o qual indica um mercado em concorrência monopolística. Note que os demais valores de H estimados e

reportados na tabela 10, e os reportados na tabela 9, revelam que, utilizando a base de dados 4010 ou 4040, o mercado bancário opera em concorrência monopolística.

Por outro lado, vale destacar que as estimativas pontuais do parâmetro H com a base de dados do relatório 4040 (tabela 10) são menores que as estimativas pontuais do parâmetro H estimados com a base de dados do relatório 4010 (tabela 9). Assim, conforme Vesala (1995), as estimativas de H sugerem que o grau de concorrência do setor bancário brasileiro calculado com a base de dados de conglomerado financeiro é menor que o grau de concorrência calculado com a base de dados de CNPJ. Essa é a primeira indicação de que a desconsideração de efeitos de coordenação entre empresas do mesmo grupo econômico pode levar ao viés de superestimação da intensidade de concorrência bancária.

Uma possibilidade de teste que este trabalho traz sobre o uso das duas bases de dados é a comparação estatística sobre os valores estimados da estatística H em um teste de diferença de médias. Na seção 4.4, é apresentado o teste de diferença de média dos valores estimados, comparando se há ou não diferença estatística no uso das bases de dados CNPJ e conglomerado. Esta comparação é feita para todo o período analisado (de dezembro/1999 a junho/2010), bem como para o período subsequente a um processo de concentração na indústria bancária, em meados dos anos 2000.

Por fim, vale apontar duas leituras das duas últimas tabelas: *i*) os sinais positivos para (d_work) e (d_cap), e negativo para (d_adm) mantiveram-se; *ii*) os valores da estatística H em testes de Wald para $H = 0$ ou $H = 1$ nas duas bases apontam valores estatisticamente diferentes de 0 ou 1. Logo, segundo essas medidas o setor bancário realmente encontra-se em concorrência monopolística – o que reafirmam os resultados da maioria dos testes para o caso brasileiro.¹⁹

4.2 Boone (2008)

A mensuração da elasticidade preço-lucro,²⁰ em relação aos fatores de custo, tem neste modelo uma forma funcional direta de mensurar competitividade, segundo a equação:

$$Ln(\pi_{it}) = \alpha - \beta \cdot Ln(C_{it}) + \mu_i + \varepsilon_{it} \quad (8)$$

19. Vale destacar que, tal como na estimação do modelo de Panzar e Rosse com a base de dados do relatório 4010 (CNPJ), na estimação do mesmo modelo com a base de dados do relatório 4040 (conglomerado bancário) também seguimos a sugestão de Bikker, Shaffer e Spierdijk (2012). Assim, utilizamos a receita não normalizada do banco variável dependente na equação estimada (8) e não utilizamos *proxies* para o tamanho do banco (por exemplo, ativo) na estimação de (8) para não incorrerem no problema de viés de estimação da estatística H destacado por aqueles autores.

20. No Brasil, este deve ser o primeiro trabalho acadêmico a utilizar este modelo na mensuração da concorrência do setor bancário – esta é mais uma contribuição.

As variáveis μ_{it} e ε_{it} representam, respectivamente, o lucro e o custo marginal do banco i no período t . O termo μ_i é o efeito fixo do banco e ε_{it} é o termo errático. A estimação foi na forma de um painel em dois estágios, em que no segundo estágio foi estimada a seguinte equação:

$$\ln(\pi_{it}) = \alpha - \beta_1 \cdot \ln(CMg_{it}) + \beta_2 \cdot d_{1st} + \beta_3 \cdot HHI_{ATIVOt} + \mu_i + \varepsilon_{it} \quad (9)$$

O parâmetro β_1 estimado representa o beta do modelo Boone (2008), e os parâmetros $(\alpha_1, \alpha_2, \alpha_8, \alpha_9)$ na equação (10) são parâmetros estimados na equação de custo total do primeiro estágio que geram a necessidade de correção da variância dos parâmetros estimados no segundo estágio.²¹

Nessa estrutura, para estimar a equação (9), foi estimado no primeiro estágio a equação de custo total, na forma funcional *translog*, como em Shaffer (1993),²² dada por:

$$\begin{aligned} \ln C_{it} = & \alpha_0 + \alpha_1 \ln Q_{it} + \alpha_2 (\ln Q_{it})^2 + \alpha_3 \ln W_{1t} + \alpha_4 \ln W_{2t} \\ & + \alpha_5 \frac{\ln(W_{1t})^2}{2} + \alpha_6 \frac{\ln(W_{2t})^2}{2} + \alpha_7 \ln W_{1t} \ln W_{2t} \\ & + \alpha_8 \ln Q_{it} \ln W_{1t} + \alpha_9 \ln Q_{it} \ln W_{2t} + \mu_i + \varepsilon_{it} . \end{aligned} \quad (10)$$

A partir da equação (10), os parâmetros avaliados foram estimados para calcular o Custo Marginal, dado por:

$$CMg_{it} = [C/Q][\alpha_1 + 2 \cdot \alpha_2 \ln Q_{it} + \alpha_8 \ln W_{1t} + \alpha_9 \ln W_{2t}], \quad (11)$$

com o CMg_{it} estimado no primeiro estágio, estimou-se o BETA do modelo Boone na equação (9) do segundo estágio.

As variáveis são de simples descrição: C_{it} corresponde aos custos totais do mercado de empréstimo; Q_{it} corresponde à quantidade total de empréstimos; W_{1t} corresponde aos custos administrativos e de pessoal; e W_{2t} corresponde aos custos de captação.

Para garantir uma mensuração correta das significâncias individuais do segundo estágio, foi aplicado o método descrito nos apêndices técnicos, disponível no material suplementar. Esse mesmo procedimento foi adotado para a base CNPJ

21. A metodologia que utilizamos para realizar tal correção está baseada em Wooldridge (2002) e segue descrita nos apêndices técnicos disponíveis no material suplementar.

22. Outras referências, antecessoras à Shaffer (1993), para a especificação translog são os trabalhos de Christensen e Green (1976), Mester (1987) e Berger, Hanweck e Humphrey (1987).

e conglomerado, em que os cálculos matriciais sugeridos foram realizados com o uso do software Matlab.²³

Para resumir os resultados encontrados nas duas bases, a seguir, são apresentados os valores do segundo estágio das estimações para o modelo Boone (2008), descritas acima, com os valores das variâncias corrigidos.²⁴

Boone (2008) – CNPJ. A equação estimada em dados de painel, com efeitos fixos, é dada para este caso por:

$$\text{Ln}(\pi_{it}) = \alpha - \beta_1 \cdot \text{Ln}(\text{CMg}_{it}) + \beta_2 \cdot d_{1st} + \beta_3 \cdot \text{HHI}_{\text{ATIVO}_t} + \mu_i + \varepsilon_{it}. \quad (12)$$

E os resultados obtidos seguem na próxima tabela.

TABELA 11

Apresentação dos valores estimados no 2º estágio do modelo Boone [equação (7)] com a utilização dos dados do relatório 4010 (CNPJ).
(Variável dependente: LN do lucro)

Ln(CM _g)	-1.787 (0.02)***
D-1s	0.001 (0.00)***
(HHI) _{ativo}	-5.018 (0.10)***
Constante	-117.142 (4.008)***
Nº observações	5124
Nº bancos	193
R ² (within)	0.42

(Em parênteses estão os desvios-padrão robustos)

* Significativo a 10%; ** Significativo a 5%; *** Significativo a 1%

Fonte: Elaboração própria.

O parâmetro de beta = 1,78, medido pelo coeficiente associado ao Ln(CM_g), revela uma elasticidade do lucro não muito elevada quando comparada com resultados de outros trabalhos. Leuvensteijn *et al.* (2007), por exemplo, encontram betas iguais a 3,38 e 4,15 para Alemanha e Espanha, respectivamente.

23. Note que isso é necessário porque os parâmetros estimados no primeiro estágio interagem de forma não linear com os parâmetros a serem estimados no segundo estágio. Se esta interação fosse linear, entre os parâmetros do primeiro e segundo estágio, o procedimento mais usual seria projeção das variáveis dos dois estágios no espaço das médias e desvios das médias, para então a utilização da função *reg3* no Stata.

24. As estimações de primeiro e segundo estágio, bem como a descrição das variáveis de controle adotadas seguem no material suplementar.

Com esse resultado, os bancos no Brasil poderiam ser considerados com níveis menores de concorrência, mas para tal afirmação seriam necessárias comparações do período de análise e demais variáveis da base de dados, além de um maior aprofundamento do modelo Boone (2008) para o caso brasileiro.

Boone (2008) – conglomerado. A equação estimada em dados de painel, com efeitos fixos, é dada para este caso pela equação 13, sendo os resultados apresentados na tabela 12.

$$\text{Ln}(\pi_{it}) = \alpha - \beta_1 \cdot \text{Ln}(\text{CMg}_{it}) + \beta_2 \cdot d_{1st} + \beta_3 \cdot \text{HHI}_{\text{ATIVO}_t} + \mu_i + \varepsilon_{it}. \quad (13)$$

TABELA 12

Valores estimados no 2º estágio do modelo Boone equação 13, com a utilização dos dados do relatório 4040 (conglomerado)

(Variável dependente: LN do lucro)

Ln(CM _g)	-0.860 (0.350)**
D-1s	0.073 (0.277)
(HHI) _{ativo}	0.401 (0.003)***
Constante	3.882 (0.110)**
Nº observações	3766
Nº bancos	176
R ² (within)	0.39

(Em parênteses estão os desvios-padrão robustos)

* Significativo a 10%; ** Significativo a 5%; *** Significativo a 1%

Fonte: Elaboração própria.

O parâmetro de beta = 0,86, medido pelo coeficiente associado ao Ln(CMg) revela uma elasticidade do lucro menor na base conglomerado em relação à base CNPJ, o que valida a hipótese inicial de diferenças das bases. De acordo com Boone (2008), uma menor elasticidade lucro indica um menor grau de concorrência na base de dados do relatório 4040 (conglomerado), que na base de dados do relatório 4010 (CNPJ). Note que esse resultado é semelhante ao resultado obtido nas estimações da estatística H de Panzar e Rosse.

Entretanto, somente o teste de médias realizado na seção 4.4 é que pode inferir quanto à diferença estatística desses resultados.

4.3 Bresnahan-Lau (1982)

As formas funcionais adotadas para demanda e oferta seguem a referência pioneira do trabalho de Shaffer (1993). Esta, talvez seja, a mais citada referência no uso de modelos Bresnahan-Lau.

Para tanto, utilizamos as formas de estimação em dois estágios, com dados em painel, sendo que no segundo estágio há um termo não linear.

Para demanda, a forma funcional é dada por:

$$Q_{it} = \alpha_0 + \alpha_1 \cdot P_{it} + \alpha_2 \cdot Y_{it} + \alpha_3 \cdot P_{it} \cdot Z_{it} + \alpha_5 \cdot P_{it} \cdot Y_{it} + \alpha_6 \cdot Z_{it} \cdot Y_{it} + \sum_{j=1}^{k-1} \delta_j \cdot D_{ij} + \mu_i + \varepsilon_{it} \quad (14)$$

Para Oferta, a forma funcional é dada por:²⁵

$$P_{it} = \frac{C_{it}}{q_{it}} \cdot [s_0 + s_1 \cdot \ln(q_{it}) + s_2 \cdot \ln(w_{1t}) + s_3 \cdot \ln(w_{2t})] - \lambda \cdot \frac{Q_{it}}{\alpha_1 + \alpha_3 \cdot Z_{it} + \alpha_5 \cdot Y_{it}} + \sum_{j=1}^{m-1} \delta_j \cdot D_{ij} \cdot \frac{Q_{it}}{\alpha_1 + \alpha_3 \cdot Z_{it} + \alpha_5 \cdot Y_{it}} + \mu_i + \varepsilon_{it}, \quad (15)$$

na qual:²⁶

- P : razão entre renda dos empréstimos e total de empréstimos
- Q : total dos empréstimos
- Y : PIB deflacionado – Ipeadata
- Z : Selic mensal – Ipeadata
- C_j : Custo total
- W_{1j} : Despesas (proventos + encargos + benefícios pagos a funcionários)
- W_{2j} : Custo de captação
- λ : parâmetro de conjectura (igual a 0 implica concorrência perfeita; e igual a 1 monopólio).
- D_{ij} : variáveis *dummies* utilizadas como controle

25. As estimações de primeiro e segundo estágio, bem como descrição das variáveis de controle adotadas seguem no material suplementar.

26. As variáveis Q_{it} , C_{it} , W_{1it} e W_{2it} foram deflacionadas pelo IPCA.

Para esse modelo, temos a mesma ressalva referida anteriormente sobre a não linearidade dos parâmetros estimados no primeiro estágio e interagindo com especificação não linear no segundo estágio, como explicado na nota de rodapé.

Bresnahan-Lau (1982) – CNPJ. No material suplementar, encontram-se as tabelas com os valores estimados para o primeiro e segundo estágio do modelo. A seguir apresentam-se apenas os valores do coeficiente associado aos termos (λ) e variáveis *dummies* utilizadas. Tais como:

$$-\lambda \cdot \frac{Q_{it}}{\alpha_1 + \alpha_3 \cdot Z_{it} + \alpha_5 \cdot Y_{it}} \text{ e } \delta_i \cdot D_{ij} \cdot \frac{Q_{it}}{\alpha_1 + \alpha_3 \cdot Z_{it} + \alpha_5 \cdot Y_{it}}. \quad (16)$$

Os parâmetros α_1 , α_3 e α_5 são parâmetros estimados na equação de demanda do primeiro estágio – conforme estimações no material suplementar. Dessa forma, é preciso corrigir a variância dos parâmetros estimados no segundo estágio para garantir uma mensuração correta das significâncias individuais. A tabela 13 contém os parâmetros após a correção da variância do segundo estágio, descrito nos apêndices técnicos no material suplementar.

TABELA 13

Apresentação dos valores estimados no 2º estágio do modelo Bresnahan-Lau com a utilização dos dados do relatório 4010 (CNPJ)
(Variável dependente: oferta de empréstimos (P_j))

Lamb	-0.0026 (40.312)
Dum1	0.000 (1.4729)
Dum5	0.036 (572,73)
Constante	-1.269 (0.226)***
Nº observações	5448
Nº bancos	207
R ² (within)	0.56

(Em parênteses estão os desvios-padrão robustos)

* Significativo a 10%; ** Significativo a 5%; *** Significativo a 1%

Fonte: Elaboração própria.

Com as variáveis de controle utilizadas sendo: *i*) (dum1) como variabilidade do lucro do banco, e *ii*) (dum5) como índice de concentração do volume de crédito. Note que essas variáveis foram construídas como indicado nas equações (15) e (16), com a presença do fator não linear.

Carvalho (2006) estima este mesmo modelo simplesmente removendo a não linearidade e escrevendo no segundo estágio o parâmetro λ como uma constante, e as demais variáveis *dummies*. Para fins de comparação, o modelo também foi estimado dessa forma, tendo-se encontrado valores muito próximos ao referido trabalho; por exemplo, λ assume valores como 0,23 ou, 0,18 ou, 0,21 a depender da especificação. Entretanto, opta-se aqui pela utilização da especificação descrita nas equações 15 e 16, que reproduzem mais fielmente a literatura original.

Os valores encontrados para a base CNPJ não foram significantes para diferentes testes de especificações, mas aponta um mercado mais próximo de concorrência do que monopólio.

Bresnahan-Lau (1982) – conglomerado. Para esta subseção, valem as mesmas ressalvas dadas na subseção anterior. No material suplementar, encontram-se as tabelas com os valores estimados para o primeiro e segundo estágios do modelo.

Conforme observado na tabela 14, os valores encontrados para a base conglomerado foram significantes, inclusive para diferentes testes de especificações, e apontam um mercado mais próximo de concorrência do que monopólio, apesar de o valor $\lambda = 0,022$ não ser tão próximo de Nakane (2002), que encontra $\lambda = 0,0017$. Note-se, contudo, que, se fosse feita a regressão do segundo estágio com o parâmetro λ sem o termo não linear, seriam obtidos valores dez vezes maiores, o que poderia levar à conclusão de uma medida de competição mais acirrada.

TABELA 14

Valores estimados no 2º estágio do modelo Bresnahan-Lau com a utilização dos dados do relatório 4010 (CNPJ)

(Variável dependente: oferta de empréstimos (P_j))

Lamb	0.02228 (0.0023)***
Dum1	-0.0004 (0.000)***
Dum5	-0.259 (0.0145)***
Constante	0.106 (0.002)***
Nº observações	3614
Nº bancos	175
R ² (within)	0.17

(Em parênteses estão os desvios-padrão robustos)

* Significativo a 10%; ** Significativo a 5%; *** Significativo a 1%

Fonte: Elaboração própria.

Em resumo, o modelo Bresnahan-Lau (1982) também nos leva a concluir que o setor bancário, segundo essas medidas de competição, apresenta-se em concorrência monopolística. Por outro lado, vale destacar também que as estimações do λ indicam um menor grau de concorrência na base de dados do relatório 4040 (conglomerado) que na base de dados do relatório 4010 (CNPJ). Note que esse resultado é semelhante aos resultados obtidos nas estimações da estatística H de Panzar e Rosse e do Beta de Boone.

Contudo, tal análise comparativa será realizada com mais rigor na próxima seção, quando será verificado, através de um teste de diferença de médias, se há diferença estatística entre os parâmetros estimados com as duas bases de dados (4010 e 4040) para os três diferentes dos modelos de conduta (Panzar e Rosse, Boone e Bresnahan-Lau). A próxima seção, portanto, identifica se o uso de bases de dados diferentes pode gerar distorções na avaliação da intensidade de competição.

4.4 Diferenças de médias

Esta subseção é destinada a concluir se há ou não diferença no uso da base CNPJ (relatório 4010) em relação ao uso da base conglomerado (relatório 4040).

Como discutido ao longo da apresentação dos resultados dos modelos, os valores para a estatística H , para o λ , do Bresnahan-Lau, e para o beta, do modelo Boone, são diferentes quando estimados com uso de bases diferentes. Contudo, isso não implica que sejam diferentes em média do ponto de vista estatístico.

Para responder esta pergunta, são apresentados nesta seção testes de médias para os três modelos. No caso do modelo de Panzar e Rosse, a análise foi estendida para diferentes períodos, procurando captar se os resultados se mantêm robustos com o processo de reestruturação da indústria bancária ao longo da década de 2000. Esse exercício se restringiu ao modelo de Panzar e Rosse, pois apenas para ele foi possível fazê-lo para diferentes períodos.

Testes gerais. A proposta inicial foi realizar o teste de diferença de média sobre a amostra como um todo, ou seja, para os valores obtidos nas três subseções anteriores, estimados para o período de dez./1999 a jun./2010. Para a estatística H , foram realizados três testes, que são combinações dos valores encontrados para especificações de variáveis de controle das estimações apresentadas nas subseções de anteriores. Os resultados são apresentados na tabela 15.

Para o período de dez./1999 a jun./2010, não há diferenças significativas para o uso das bases de dados diferentes. Somente o modelo de Boone a 5% que rejeita a hipótese nula, ou seja, de que os valores estimados pelas bases de dados diferentes têm diferença estatística.

TABELA 15

Apresentação dos testes sobre diferenças das médias realizados sobre os parâmetros estimados em relação à conduta bancária, nos três diferentes modelos Panzar-Rosse (1987), Bresnahan-Lau (1982) e Boone (2008)

	CNPJ	Conglomerado	T-stat	P-valor
H-Panzar e Rosse				
Teste 1	0,35049	0,24251	1,51435	0,129
Teste 2	0,35049	0,26172	1,24553	0,212
Teste 3	0,35049	0,29794	0,73664	0,461
Lambda				
Bresnahan-Lau	-0,0026411	0,022279	-0,00061818	0,99
Beta-Boone	-1,768801	-0,85997	-2,25036262	0,024

Fonte: Elaboração própria.

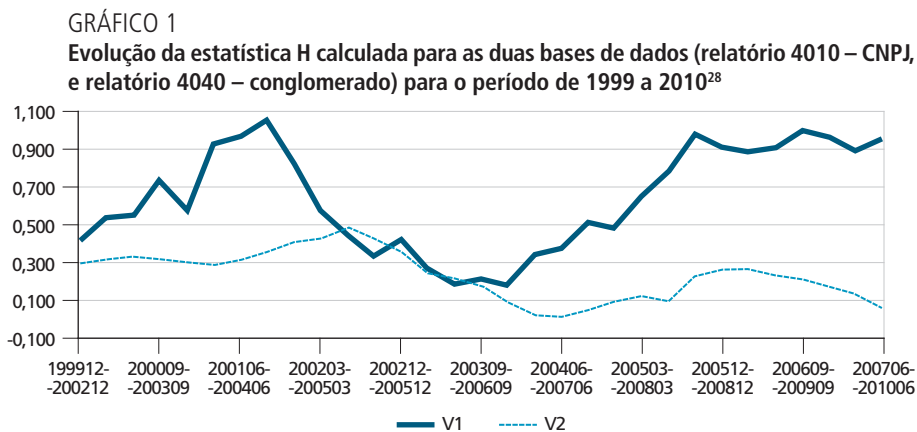
Assim, tomando-se um longo período de tempo, não há diferenças relevantes no uso das diferentes bases de dados para os modelos de Panzar e Rosse, e Bresnahan-Lau. Dessa forma, tomando-se a média do período analisado, não há prejuízos em se avaliar a intensidade de concorrência por meio da base de dados de instituições individuais, ainda que tal procedimento, conforme já argumentado, seja conceitualmente incorreto para aqueles modelos. Porém, as estimações do modelo de Boone (2008) no nosso artigo também sugerem uma superestimação do grau de concorrência no mercado brasileiro quando a análise é realizada com a base de dados de bancos individuais.

Contudo, essa conclusão não se mantém para qualquer período do tempo. A diferença nos dois procedimentos de mensuração do grau de concorrência deve ser tanto maior quanto maior for a reunião de empresas em um mesmo grupo econômico. Como parte das fusões observadas ao longo da década de 2000 preservou o CNPJ original das empresas fusionadas, é no final dessa década que a mensuração que tem por base a pessoa jurídica está sujeita a maiores distorções.

Utilizando novamente o modelo de Panzar e Rosse para janelas de tempo, com períodos de 03 anos, ao longo da amostra de dados de dez./1999 a jun./2010, há evidências de que a desconsideração do grupo econômico gera distorções relevantes à análise. Esses resultados seguem na próxima subseção.

Testes específicos com Panzar-Rosse (1987). Para o modelo de Panzar e Rosse foi possível estimar a evolução da estatística H ao longo do tempo, como em Nakane e Rocha (2010) e em Araújo e Jorge Neto (2007)²⁷. Para tanto, foi estimada a estatística H em janelas de tempo de três anos, para ambas as bases, sendo os resultados apresentados no Gráfico 1.

27. No caso dos modelos de Bresnahan-Lau (1982) e Boone (2008), não foi possível obter especificação que se mantivesse robusta ao longo de diferentes períodos para as duas bases de dados, CNPJ e conglomerado, o que decorre do maior volume de parâmetros a serem estimados.



Nota-se no gráfico 1 que a série de estatísticas H avaliada na base 4010 (CNPJ) se aproxima bastante dos valores apontados por Nakane e Rocha (2010), principalmente entre 2002 e 2004. Outra leitura de resultados que o gráfico sugere, e nesse caso dissonante de Nakane e Rocha (2010), é que nos anos mais recentes, após os processos das grandes fusões bancárias de 2008, a concorrência parece ter uma tendência de queda.

Essa conclusão, extraída do gráfico 1, sugere a realização do teste de diferença de médias para dois períodos distintos, de modo a captar as diferenças observadas na evolução da estatística H. Esses resultados são apresentados na tabela 16.

TABELA 16²⁹

Apresentação de dois testes sobre diferenças das médias realizados sobre os parâmetros estimados em relação à conduta bancária com o modelo Panzar-Rosse (1987) – até junho/2004 e após junho/2004

	CNPJ	Conglomerado	T-stat	P-valor
	H-Panzar e Rosse			
Até junho/2004	0,480	0,298	1,141	0,254
Pós junho/2004	0,447	0,125	2,870	0,004

Fonte: Elaboração própria.

Nos resultados constantes na tabela 16, em que se separam as estimativas da estatística H para os períodos até junho/2004 e após junho/2004, pode-se notar que há diferenças significantes no uso das duas bases de dados. Em outras palavras,

28. As séries de valores da estatística H apresentadas no gráfico 1, bem com as variáveis (d_work), (d_adm) e (d_cap), com os respectivos testes de significância, e testes Wald sobre $H = 0$ ou $H = 1$, seguem no material suplementar.

29. As estimações que geram os valores da tabela 16 acima estão no material suplementar.

a conclusão da subseção anterior, de que não haveria prejuízo da desconsideração do grupo econômico no modelo de Panzar e Rosse (isto é, uso da base de CNPJ) não se mantém para o período após junho/2004. E note-se que esta conclusão pode ser estendida como ressalva sobre o uso de qualquer modelo que estime nível de concorrência.

Ao revisitar a literatura aplicada ao caso brasileiro, pode-se afirmar que Nakane (2002), mesmo usando a base de dados de CNPJ, não teria incorrido em medidas viesadas, apesar de desconsiderar efeitos de coordenação dentro do mesmo grupo econômico. Também Carvalho (2006), que faz uso do modelo Bresnahan-Lau, por utilizar período semelhante, também não teria incorrido em viés significativo em decorrência do uso da base de dados imprópria.³⁰ No caso dos trabalhos que se utilizam do modelo de Panzar e Rosse, tanto Araujo e Jorge Neto (2007), quanto Lucinda (2010), não teriam suas estimações enviesadas pela utilização da base de dados de CNPJ, porque o primeiro utiliza dados até junho/2004 e o segundo até dezembro/2005.

Entretanto, Nakane e Rocha (2010)³¹ podem estar superestimando a intensidade de concorrência por desconsiderarem a constituição de grupos econômicos. Conforme observado pelo teste de médias, há diferença significativa com o uso das bases dos relatórios 4010 e 4040, após junho/2004.

Em síntese, a análise ao longo do período 1999 a 2010 indica que nos anos mais recentes, após a elevação nos níveis de concentração do setor bancário, deve haver diferença na estimação de parâmetros de concorrência quando a unidade de análise for restrita à pessoa jurídica (base CNPJ), em comparação com a base de dados que agrega as pessoas jurídicas pertencentes a um mesmo conglomerado financeiro (base conglomerado).

5 CONSIDERAÇÕES FINAIS

A década de 2000 marcou uma importante mudança na estrutura da indústria bancária no Brasil, com algumas aquisições de elevado impacto, como as fusões entre Itaú e Unibanco e entre Banco do Brasil e Nossa Caixa. Essas fusões têm o potencial de afetar não apenas a concorrência bancária, mas também a acurácia de medidas de competição que desconsiderem a existência de coordenação dentro de um mesmo grupo econômico. As estimativas do grau de competição bancária no Brasil, ao adotarem os relatórios 4010 do Banco Central como fonte de dados, estão sujeitas a esse problema, uma vez que as informações são ali agrupadas pelo CNPJ dos bancos individuais e não pelo grupo econômico.

30. Entretanto, sua opção pela linearização do parâmetro de interesse pode resultar em distorções.

31. Uma ressalva deve ser posta: o referido trabalho cita o uso da base "Top 50" que seria correspondente ao relatório 4040, entretanto ao indicar contas realizadas com os lançamentos contábeis dos bancos indica o uso de informações do relatório 4010.

Este artigo apresenta duas contribuições a este tema. O primeiro foi estimar empiricamente se a desconsideração do grupo econômico produz viés nas estimativas de medidas de competição no mercado bancário brasileiro. Para tanto, foram estimadas medidas de competição no setor bancário brasileiro, utilizando dados financeiro-contábeis de bancos individuais (relatório 4010) e de conglomerados bancários (relatório 4040). A segunda contribuição, menos central ao artigo, foi a utilização do modelo de Boone (2008), ainda não aplicado à indústria bancária brasileira, além dos modelos de Bresnahan e Lau (1982) e de Panzar e Rosse (1987), frequentemente utilizados pela literatura aplicada.

As estimativas mostram que, a partir de 2004, há evidências de diferenças significantes entre as duas bases de dados, o que revela que estudos que desconsideram efeitos de coordenação entre empresas do mesmo grupo econômico estão sujeitos a viés de superestimação da intensidade de concorrência bancária. Esse resultado foi observado, em particular, para o caso da estatística H, de Panzar e Rosse (1987), para o qual foi possível realizar estimações para cada ano, bem como testar as diferenças para os períodos antes e após 2004. As estimações do modelo de Boone (2008) também apontam a existência de diferença estatística entre os parâmetros de conduta estimados com as duas bases de dados para o período de 1999 a 2010, indicando uma superestimação do grau de concorrência no mercado bancário brasileiro, quando a análise é realizada com a base de dados de bancos individuais.

Vale destacar, por outro lado, que as estimações para o modelo de Bresnahan e Lau (1982), os quais foram realizadas também para o período de 1999 a 2010, não têm revelado diferenças significantes entre as duas bases de dados, tendo também confirmado resultados anteriores sobre o grau de concorrência na indústria.

Os resultados sugerem que as mudanças ocorridas na estrutura da indústria ao longo da década de 2000 tornaram as medidas de competição mais sensíveis ao pressuposto implícito da literatura aplicada, de competição entre bancos individuais. Desse modo, as pesquisas conduzidas com dados referentes à primeira metade da década de 2000 não estão sujeitas a viés de desconsideração do grupo econômico. Por outro lado, os resultados também indicam que estimativas do grau de competição nos períodos mais recentes devem utilizar o conceito de conglomerado financeiro como unidade de concorrência.

O aprofundamento do processo de concentração da indústria bancária na década de 2010, com eventos de porte como a saída do grupo HSBC, deve motivar novos estudos sobre as medidas de competição nessa indústria. É desejável que esses novos estudos passem a utilizar dados referentes aos conglomerados financeiros, visto que a utilização de dados de bancos individuais pode produzir distorções significantes nas medidas de competição.

REFERÊNCIAS

- ARAÚJO, L.A.; JORGE NETO, P. M.; PONCE, D.S. **Competição e concentração entre os bancos brasileiros**. In: XXXIII Encontro Nacional de Economia – Anpec, 2005.
- ARAÚJO, L.A.; JORGE NETO P. M. Risco e Competição Bancária no Brasil. **Revista Brasileira de Economia**, n. 2, v. 61, 2007.
- BARAJAS, A.; STEINER, R.; SALAZAR. Interest spreads in banking Colombia, 1974-96. **IMF Staff Papers**, v.46, 196-224 n. 1999.
- BARBOSA, K.; DE PAULA ROCHA, B.; SALAZAR, F. Assessing competition in the banking industry: a multi-product approach. **Journal of Banking & Finance**, 50, 340-362, 2015.
- BELAISCH Agnès. Do Brazilian banks compete? **IMF WP**, 3/113, 2003.
- BERGER, Allen. *et al.* Bank Concentration and Competition: An Evolution in the Making. **Journal of Money, Credit, and Banking**, 36, 433-451, 2004.
- _____. Competitive Viability in Banking: Scale, Scope and Product Mix Economies. **Journal of Monetary Economics**, 501-20, 1987.
- _____. Efficiency of financial institutions: international survey and directions for future research. **European Journal of Operational Research** v. 98(2), 175-212, 1997.
- Bikker, Jacob A.; Haff, Katharina. Measures of Competition and Concentration in the Banking Industry: a Review of the Literature. **Economic & Financial Modelling**, 1-46, 2002.
- _____. Competition, concentration and their relationship: An empirical analysis of the banking industry. **Journal of Banking and Finance**, v.26, 2191-2214, 2002.
- BIKKER, J. A.; SHAFFER, S.; SPIERDIJK, L. Assessing Competition with the Panzar-Rosse Model: The Role of Scale, Costs, and Equilibrium. **Review of Economics and Statistics**, v. 94, n. 4, p. 1025-1044, 2012.
- BOONE, J. Competition. **Discussion Paper Series Tilburg University**, Center for Economic Research, v. 104, 2000.
- _____. A New Way to Measure Competition. **CEPR Discussion Paper Series**, n. 4330, 2004.
- _____. Optimal competition: a benchmark for competition policy. **The Economic Journal**, v. 118, n. 531, p. 1245-1261, 08, 2008.
- BOONE, J.; OURS, J. C. V.; WIEL, Henry. How (not) to Measure Competition. **CPB Discussion Paper Series**, n. 2007-32, 2007.

BRESNAHAN, Timothy F. The Oligopoly Solution is Identified. **Economic Letters** n. 10, p. 87-92, 1982.

CARVALHO, Vitor Aguiar de. Análise dos Atos de Concentração no Setor Bancário: um estudo motivado pelas recentes fusões no mercado brasileiro. **Prêmio SEAE 2006 de Monografia sobre Defesa da Concorrência**, 2006.

CHRISTENSEN, Laurits R.; GREENE, William H. Economies of Scale in U.S. Electric Power Generation. **The Journal of Political Economy** v. 84, n. 4 part 1 655-676, 1976.

DEGRYSE, H.; KIM M.; ONGENA S. Microeconometrics of Banking: Methods, Applications, and Results. **Oxford University Press**, 2009.

IWATA, G. Measurement of Conjectural Variations in Oligopoly. **Econometrica** n. 4, 947-966, 1974.

LAU, L. On Identifying the Degree of Competitiveness from Industry Price and Output Data. **Economic Letters** 10, 93-99, 1982.

LEUVENSTEIJN, M. *et al.* A New Approach to Measuring Competition in the Loan Markets of the Euro Area. **European Central Bank Working Papers Series**, n. 84, 2007.

LUCINDA, C. Competition in the Brazilian loan market: an empirical analysis. **Estudos Econômicos**, v. 40, p. 831-858, 2010.

MARTIN, S. Advanced Industrial Economics. **Blackwell**, 2nd ed., 2002.

MESTER, Loretta J. A Multiproduct Cost Study of Savings and Loans. **The Journal of Finance**, n. 42, 423-45, 1987.

NAKANE, Márcio I. A test of competition in Brazilian banking. **Estudos Econômicos**, v. 32, 203-224, 2002.

NAKANE, Márcio I.; ROCHA, Bruno. Concentração, Concorrência e Rentabilidade no setor bancário brasileiro: uma visão atualizada. **Tendências Consultoria Integrada** – Febraban, 2010.

NEVEN, D.; RÖLLER, L.-H. An aggregate structural model of competition in the European banking industry. **International Journal of Industrial Organization**, v.17, 1059-1074, 1999.

O'BRIEN, D. P.; SALOP, S. C. Competitive effects of partial ownership: Financial interest and corporate control. **Antitrust Law Journal**, 559-614, 2000.

PANZAR, J.; ROSS, E.J. Testing for 'Monopoly' Equilibrium. **Journal of Industrial Economics**, 35, 443-456, 1987.

SALVO, Alberto. O Método de Bresnahan-Lau: uma nota cautelosa sobre a sua aplicabilidade. **Revista de Direito da Concorrência**, n. 18, 9-60, 2007.

SHAFFER, Sherrill. Competition, conduct and demand elasticity. **Economics Letters**, v.10, n. 1-2, p. 167-171, 1982.

SHAFFER, Sherrill. Non-structural measures of competition: Toward a synthesis of alternatives. **Economics Letters**, v. 12, n. 3-4, p. 349-353, 1983.

SHAFFER, S. Competition in the U.S. banking industry. **Economic Letters**, 29, 321- 323, 1989.

_____. A test of competition in Canadian banking. **Journal of Money, Credit and Banking**, 25, 49-61, 1993.

SPILLER, Pablo T.; FAVARO, Edgardo. The Effects of Entry Regulation on Oligopolistic Interaction: The Uruguayan Banking Sector. **RAND Journal of Economics**, v. 15, n. 2, p. 244-254, 1984.

VESALA, J. Testing for Competition in Banking: behavioral evidence from Finland. **Bank of Finland Studies – Bank of Finland**, 1995.

WILLIAMSON, Dean V. Organization, Control, and the Single Entity Defense in Antitrust. **Journal of Competition Law and Economics**, v. 5, n. 4, p. 723-745, 2009.

(Originais submetidos em fevereiro de 2014. Última versão recebida em agosto de 2015. Aprovada em dezembro de 2015.)

INFLUÊNCIA DO AMBIENTE ECONÔMICO SOBRE A CONFIANÇA DO CONSUMIDOR E AS DECISÕES DE CONSUMO NO BRASIL: UMA ANÁLISE EMPÍRICA CONSIDERANDO OS EFEITOS DAS CREDIBILIDADES MONETÁRIA E FISCAL¹

Gabriel Caldas Monte²

Bruno Campos Visconti³

A credibilidade é um aspecto importante para a formação de expectativas no regime de metas para a inflação. Este trabalho analisa se a credibilidade da autoridade monetária e a credibilidade fiscal afetam a confiança dos consumidores. Após observar a influência do ambiente econômico sobre a confiança dos consumidores, o estudo verifica a influência da confiança do consumidor, e de outras variáveis, sobre as decisões de consumo. Os resultados sugerem que as credibilidades monetária e fiscal influenciem a confiança dos consumidores, e que a confiança dos consumidores seja relevante para as decisões de consumo de diferentes tipos de bens.

Palavras-chave: credibilidade; confiança; consumo.

INFLUENCE OF THE ECONOMIC ENVIRONMENT ON CONSUMER CONFIDENCE AND CONSUMPTION DECISIONS IN BRAZIL: AN EMPIRICAL ANALYSIS CONSIDERING THE EFFECTS OF MONETARY CREDIBILITY AND FISCAL CREDIBILITY

Credibility is important for the expectation formation in the regime of inflation targeting. This paper analyzes whether the credibility of the monetary authority and the fiscal credibility affect consumer confidence. After observing the influence of the economic environment on consumer confidence, the study verifies the influence of consumer confidence, together with other variables, on the consumption decisions. The results suggest that monetary credibility and fiscal credibility influence the confidence of Brazilian consumers and, consumer confidence is relevant to consumption decisions of different types of goods.

Keywords: credibility; confidence; consumption.

JEL: E21; E52; E62.

1. Os autores agradecem pelos comentários e sugestões do parecerista anônimo da revista. Todos os eventuais erros presentes no texto são de responsabilidade dos autores.

2. Professor do Departamento de Economia da Universidade Federal Fluminense (UFF) e bolsista de produtividade do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). *E-mail:* <gabrielmontesuff@yahoo.com.br>.

3. *E-mail:* <b_visconti@hotmail.com>.

1 INTRODUÇÃO

A confiança do consumidor é abordada na literatura por meio de formas e metodologias distintas. Enquanto alguns trabalhos investigam quais fatores contribuem com a interpretação do contexto econômico e a formação da confiança dos consumidores (Fuhrer, 1993; Vuchelen, 2004; de Mendonça, 2009), outros analisam se a confiança do consumidor atua como mecanismo de transmissão de choques para a economia (Eppright, Arguea e Huth, 1998; Li, 2011; Starr, 2012).

Em parte, as decisões dos consumidores são influenciadas por suas expectativas e confiança, as quais são afetadas pelas incertezas que envolvem o ambiente econômico. Nesse sentido, a manutenção de um ambiente macroeconômico estável é fundamental de modo a evitar oscilações abruptas no processo de geração de renda, nas expectativas e, assim, no consumo. Entretanto, estudos apontam que a manutenção de um ambiente macroeconômico estável resulta da condução de políticas econômicas pautadas em objetivos claros e críveis (Kydland e Prescott, 1977; Barro e Gordon, 1983a e 1983b; Persson e Tabellini, 1990).

Atualmente, o regime de metas para inflação é adotado por diversos países de modo a aumentar o comprometimento dos *policymakers* com seus objetivos. A estabilidade dos preços é o principal objetivo a ser alcançado nesse regime. Para isso, é necessário não apenas uma política monetária comprometida com esse objetivo, mas também uma política fiscal. Ou seja, as políticas monetária e fiscal devem atuar de maneiras coordenadas (Mishkin, 2007). A ideia é que a utilização da política monetária para controlar a inflação e guiar as expectativas do público, sem que as finanças públicas estejam equilibradas, e, portanto, que a dívida pública esteja controlada, pode causar instabilidades no ambiente macroeconômico e deteriorar as expectativas e a confiança dos agentes.

Devido ao reconhecimento de que a efetividade da condução da política econômica depende, em parte, da capacidade do público de antecipar as ações dos *policymakers*, então, é possível levantar a hipótese de que as credibilidades monetária e fiscal representam importantes aspectos no processo de formação das expectativas e na construção da confiança dos consumidores acerca do ambiente econômico. Nesse sentido, o artigo tem os seguintes objetivos e traz as seguintes contribuições. Em primeiro lugar, na linha dos trabalhos que observam os determinantes da confiança do consumidor, o estudo destaca os efeitos da credibilidade monetária e da credibilidade fiscal sobre a confiança do consumidor no Brasil. Para o caso brasileiro, não existem estudos que observem o efeito da credibilidade fiscal sobre a confiança do consumidor; há somente o estudo de de Mendonça (2009) que estima o efeito da credibilidade monetária sobre a confiança do consumidor. Em segundo lugar, na linha dos trabalhos que analisam a confiança do consumidor como mecanismo de transmissão, o estudo destaca os efeitos da confiança do

consumidor sobre as decisões de consumo. No caso brasileiro, a literatura é escassa, e o presente estudo contribui ao analisar, pela primeira vez, a influência da percepção dos consumidores sobre diferentes bens de consumo. Portanto, o estudo desdobra-se em dois estágios: *i*) verificar o efeito das credibilidades (monetária e fiscal) e do ambiente macroeconômico sobre a confiança dos consumidores; e *ii*) analisar se a confiança do consumidor afeta as decisões dos consumidores considerando diferentes bens de consumo.

O trabalho divide-se da seguinte maneira: *i*) a seção 2 apresenta os argumentos teóricos e as evidências empíricas existentes na literatura para a importância da credibilidade no regime de metas para a inflação, como também para os determinantes do *índice de confiança do consumidor* e como este afeta as decisões de consumo; *ii*) a seção 3 apresenta evidências empíricas por meio de estimações por mínimos quadrados ordinários (MQO), método generalizado dos momentos (GMM) e sistema GMM para a economia brasileira para o período de janeiro de 2002 a dezembro de 2012; e *iii*) a seção 4 apresenta a conclusão do trabalho.

2 ARGUMENTAÇÃO TEÓRICA E EVIDÊNCIAS EMPÍRICAS

A partir do trabalho de Kydland e Prescott (1977) e do reconhecimento da importância da credibilidade para a condução da política monetária, tecnologias de compromisso – tais como, bancos centrais independentes (Cukierman, 1992), conservadorismo (Rogoff, 1985) e contratos ótimos (Walsh, 1995) – têm sido propostas com intuito de restringir as ações discricionárias dos *policymakers*. Isso porque ações discricionárias que renegam compromissos assumidos resultam em inflação elevada e, por conseguinte, em baixa credibilidade.

O regime de metas para inflação representa uma tecnologia de compromisso que busca servir como um guia para o processo de formação de expectativas dos agentes. Assim, confere importância à credibilidade e impõe comprometimento à autoridade monetária com o seu principal objetivo de longo prazo.

A credibilidade é um elemento central para o sucesso do regime de metas para a inflação. Com base na definição de credibilidade de Cukierman e Meltzer (1986) e no argumento de Agénor e Taylor (1992) de que séries de expectativas de inflação poderiam ser utilizadas para derivar esses índices, alguns economistas empenharam-se em construir índices para medir a credibilidade – por exemplo, Cecchetti e Krause (2002), de Mendonça (2007), de Mendonça e de Guimarães e Souza (2009).

O trabalho elaborado por de Mendonça (2007) contribui com a literatura sobre a mensuração da credibilidade da política monetária ao sugerir outra abordagem em relação à mensuração da credibilidade. Neste, um índice de credibilidade é construído com base na definição de Cukierman e Meltzer (1986) para credibilidade e na ideia de Agénor e Taylor (1992 e 1993). Dando sequência, o estudo elaborado

por de Mendonça e de Guimarães e Souza (2009) analisa, entre uma série de índices de credibilidade e de reputação, quais deles são mais úteis na previsão de variações na taxa de juros para o caso brasileiro. O trabalho encontra evidências de que uma credibilidade mais elevada exige menores variações nas taxas de juros para o controle da inflação com fundamento na adoção do regime de metas para inflação, como também que a credibilidade mais elevada conduz a taxa de juros a níveis menores.

O trabalho de Montes e Bastos (2011) investiga a influência da credibilidade do regime de metas para inflação sobre o comportamento da taxa nominal de juros de longo prazo (TNJLP) e, por conseguinte, sobre o *spread* da TNJLP. As evidências encontradas sugerem que os ganhos de credibilidade promovam a redução do *spread* e a suavização da curva de juros.

Montes e Bastos (2013) analisam a influência das variáveis macroeconômicas e das políticas monetária e fiscal sobre as expectativas e a confiança dos empresários no Brasil após a adoção do regime de metas para inflação. Os resultados indicam que um aumento na credibilidade da política monetária cria um ambiente mais estável, contribuindo positivamente para aumentar a confiança dos empresários na economia e nos seus negócios e, por conseguinte, sendo capaz de afetar a atividade econômica. Para o mesmo período da economia brasileira, o estudo de Montes (2013) faz uma análise para os efeitos da credibilidade sobre a condução da política monetária e seus canais de transmissão, com ênfase no canal das expectativas empresariais. Os achados apontam que os ganhos de credibilidade auxiliem na obtenção de um ambiente macroeconômico mais estável, o que afeta positivamente as expectativas dos empresários, servindo como um estímulo ao aumento de seus investimentos.

O artigo de Montes e Bastos (2014) desenvolve um modelo teórico e uma análise empírica para demonstrar que, quando a autoridade monetária age comprometida com o objetivo de estabilidade de preços e possui credibilidade, o esforço de política monetária é menor. Os achados para a economia brasileira após a adoção do regime de metas para inflação apontam que quando aumenta a reputação do Banco Central, a credibilidade do regime de metas para inflação também aumenta; e que os ganhos de credibilidade reduzem os esforços do Banco Central na condução da política monetária, pois a capacidade de ancorar as expectativas dos agentes torna-se maior e o viés inflacionário reduz.

Montes e Machado (2014) analisam empiricamente a importância da credibilidade das metas para inflação e das reputações das autoridades fiscal e monetária para a condução da política monetária e a transmissão da política monetária pelo canal do preço dos ativos. A análise para o período de metas para inflação no Brasil encontra evidências de que a estabilidade dos preços, a ancoragem

das expectativas de inflação e a sustentabilidade da dívida pública proporcionaram graus de liberdade à autoridade monetária para conduzir a política monetária, de tal modo que fosse possível programar uma trajetória de redução da taxa básica de juros. Ademais, o trabalho oferece evidências acerca do mecanismo de transmissão da política monetária pelo canal dos preços dos ativos, observando o efeito da política monetária sobre o investimento agregado.

Os estudos até então elaborados sugerem que o regime de metas para a inflação contribui para construção de um ambiente macroeconômico mais previsível e estável desde que os *policymakers* atuem de maneira comprometida com seus objetivos. Ou seja, a credibilidade é fundamental para a eficácia das políticas econômicas, pois, influencia as expectativas dos agentes de forma a ancorá-las sem maiores esforços e custos, possibilitando, assim, que sejam tomadas decisões em um ambiente menos incerto.

Em ambientes de incerteza, as decisões dos agentes são pautadas por suas expectativas e pela confiança nessas expectativas. Certas decisões de consumo são mais sensíveis ao ambiente econômico, e as incertezas existentes. Por exemplo, enquanto gastos com bens não duráveis costumam ocorrer por necessidades e sem maiores ponderações, os gastos com bens duráveis requerem uma parcela maior da renda, acesso ao crédito e planejamento por parte das famílias. Sendo assim, tais gastos são mais sensíveis ao contexto econômico e, portanto, mais influenciados pelas expectativas e pela confiança dos consumidores.

A confiança dos consumidores baseia-se na percepção destes acerca do estado atual de suas finanças e da economia, assim como nas expectativas formadas sobre o rumo futuro que esta pode tomar. Nesse sentido, os consumidores mostram-se atentos a fatores da conjuntura econômica (tais como o desemprego, a taxa de juros, produto e inflação) e também a fatores relacionados à condução das políticas econômicas (tais como as ações tomadas pelas autoridades monetária e fiscal e, portanto, a credibilidade construída).

No momento de fazerem suas escolhas, os agentes levam em consideração todas as informações disponíveis. Caso os *policymakers* tenham comportamentos inconsistentes, haverá perda de credibilidade, o que deteriorará as expectativas dos agentes e tornará o ambiente mais incerto. A incerteza terá impacto na confiança dos agentes e em suas decisões. Portanto, o argumento é que os consumidores formem suas expectativas levando em conta que as políticas econômicas alteram o ambiente econômico e que essas políticas também respondam a choques nesse ambiente. Posto isto, a interpretação dos consumidores acerca da conjuntura e da condução das políticas econômicas é transmitida para suas decisões por meio da confiança e das expectativas formadas.

Existem estudos que utilizam o índice de confiança do consumidor (*icc*) para analisar a transmissão de choques para a economia por meio da percepção dos consumidores relacionada a mudanças no ambiente econômico. Por sua vez, há, também, os estudos que analisam a influência da confiança dos consumidores sobre as decisões dos agentes.

Um dos pioneiros no uso do *icc* como medida das expectativas dos consumidores foi Katona (1951, 1960, 1968, 1974, 1975). O autor introduziu a distinção entre a “capacidade” de comprar (*ability to buy*) e a “disposição” em comprar (*willingness to buy*), dando início ao que foi classificado como “teoria adaptativa do comportamento do consumidor” (*adaptive theory of consumer behaviour*). Segundo a teoria desenvolvida por Katona, os gastos discricionários⁴ com bens de consumo são uma função da “capacidade” de comprar dos agentes e da “disposição” em comprar. A “capacidade” de comprar refere-se aos fatores objetivos que determinam os gastos, como renda, riqueza e acesso a crédito, enquanto a “disposição” em comprar captura fatores subjetivos, como preferências, atitudes e expectativas dos consumidores. A “capacidade” e a “disposição” de comprar interagem entre si, porém de forma independente uma da outra. O *icc* seria uma medida representativa do sentimento e das expectativas dos consumidores, ao refletir a leitura destes sobre os rumos de suas finanças pessoais e da economia como um todo.

Ao assumir que os agentes respondem a estímulos provenientes do ambiente em que se encontram, passa a ser importante identificar os fatores que contribuem com a formação da confiança do consumidor. Nesse sentido, Fuhrer (1993) argumenta que o *icc* não pode ser visto apenas como um número isolado, pois reflete a cada divulgação a análise de um contexto econômico mutante. Ao utilizar dados trimestrais entre os anos de 1960-1990, o autor estimou, por uma equação de cointegração, que quase 70% da variação do *icc* da Universidade de Michigan é explicada por oscilações no desemprego, na inflação, na renda e na taxa de juros real. Todavia, parte dos movimentos da confiança do consumidor não refletem variáveis da conjuntura econômica.

O trabalho de Vuchelen (2004) tenta elucidar parte desses movimentos na confiança do consumidor que não são explicados por variáveis do contexto econômico. O estudo utilizou dados trimestrais de 1985 até 2000 para investigar a variação da confiança. O autor utilizou duas variáveis para a incerteza e as expectativas dos agentes quanto às condições econômicas. Com isso, concluiu que a variação do *icc* não explicada pela conjuntura reduz-se pela metade com a inclusão dessas variáveis, o que indica a importância de fatores subjetivos na ponderação dos agentes.

4. Costuma-se fazer uma distinção entre gastos não discricionários e discricionários. Os gastos não discricionários incluem obrigações contratuais dos agentes (aluguel, gastos com despesas da casa, plano de saúde e seguro etc) e gastos com necessidades básicas (comida, vestuário e transporte). Esses gastos são difíceis de reduzir ou adiar, uma vez assumidos, variando muito pouco no curto prazo. Por sua vez, os gastos discricionários são mais maleáveis, podendo ser adiados ou cancelados. Esses gastos consistem na compra ou reposição de bens duráveis (eletrodomésticos, automóveis etc) e em demais gastos, como lazer. Segundo Katona, a confiança do consumidor contribui mais na previsão dos gastos discricionários.

Um estudo acerca dos fatores macroeconômicos que determinam a confiança do consumidor no Brasil foi feito por de Mendonça (2009). Utilizando dados compreendendo janeiro de 2000 até outubro de 2008, encontrou evidências de que mudanças no índice de confiança do consumidor podem ser explicadas por variações no produto interno bruto (PIB), na taxa de juros real e na credibilidade da política monetária. O impacto da credibilidade sobre o *icc* indica que a atuação da autoridade monetária influencia a interpretação relacionada ao contexto econômico e a formação de expectativas por parte dos consumidores.

Estes trabalhos chamam atenção para o fato de que a confiança do consumidor corresponde, em parte, à interpretação da conjuntura econômica pelos agentes. Todavia, a variação do índice não é completamente esclarecida por fatores objetivos. Dessa forma, cabe espaço para que parte da explicação recaia sobre outros fatores, ligados à formação de expectativas dos agentes. Os trabalhos a seguir abordam a confiança do consumidor como um mecanismo transmissor de choques para a economia.

O estudo de Eppright, Arguea e Huth (1998) encontrou evidências, compreendendo o período de janeiro de 1978 até agosto de 1992, para a hipótese de que a confiança do consumidor, medida por diferentes índices agregados, possui informações relevantes (não contidas em outros indicadores econômicos) para antecipar movimentos nos gastos com consumo nos Estados Unidos. Os achados reforçam a hipótese de que fatores subjetivos relativos à confiança e às expectativas impactam as decisões de consumo. O autor concluiu também que a confiança e as expectativas do consumidor causam variações nos gastos com automóveis e com o consumo de bens duráveis.

Starr (2012) investiga se o noticiário econômico americano provoca alterações na confiança do consumidor e se choques provocados por notícias possuem reflexos nas decisões de consumo. Utilizando dados mensais para o período de 1978 até 1986, o estudo sugere que choques provocados por notícias descoladas da realidade econômica são importantes para explicar flutuações de curto prazo no *icc* americano. Além disso, após controlar para choques provenientes da mídia, constatou que choques na confiança do consumidor possuem um impacto pequeno e positivo sobre os gastos com consumo.

A transmissão de choques para a economia via *icc* também foi abordada por Li (2011). O estudo mostrou que o índice de expectativas do consumidor chinês, um subíndice que compõe o índice de confiança, contém informações úteis sobre choques puramente expectationais (não relacionado aos fundamentos econômicos) e que são importantes causadores de variações na produção industrial.

As evidências apontam que a confiança do consumidor responde a estímulos provenientes do contexto em que está inserido e funciona como mecanismo transmissor de choques. Ademais, os trabalhos também buscam analisar se a inclusão da confiança do consumidor, como variável explicativa das decisões de consumo, melhoram as previsões de gastos das famílias. Entretanto, parece haver pouco consenso

na literatura quanto ao *icc* conter informações adicionais relevantes para a previsão dos gastos das famílias que não sejam capturadas pelos fundamentos econômicos. Os resultados costumam variar de acordo com as variáveis de controle utilizadas com a variável dependente a ser prevista e com o período de tempo. Trabalhos em que a janela temporal engloba crises econômicas costumam apresentar resultados mais relevantes da confiança prevendo gastos de consumo, o que aponta o impacto da deterioração da confiança sobre as decisões de consumo.

Fuhrer (1993) e Carroll, Fuhrer e Wilcox (1993), por exemplo, concluíram que a confiança do consumidor, por si só, possui um poder preditivo significativo e relevante sobre o consumo das famílias americanas trimestres à frente. Todavia, o poder preditivo do *icc* reduz-se conforme outras variáveis explicativas relacionadas à conjuntura econômica (tais como renda real disponível passada e variação do consumo passado) são inseridas na estimação, permanecendo estatisticamente significante, mas não mais economicamente.

O estudo de Bram e Ludvigson (1998) utilizou dados trimestrais de 1968 até 1996 e encontrou evidências de que a confiança do consumidor pode ajudar a prever variações nos gastos futuros com consumo. Porém, o resultado depende do índice utilizado: o índice de confiança da Universidade de Michigan ou o índice de confiança do consumidor do Conference Board. Os achados sugerem que o índice de confiança do Conference Board apresente melhores resultados no momento de prever os gastos com diferentes categorias de bens de consumo do que o índice de confiança da Universidade de Michigan, mesmo controlando para variáveis da conjuntura. Os acréscimos de previsão são mais relevantes para as categorias de gastos com veículos e com bens duráveis. Essa superioridade se deve à diferença na metodologia de construção dos índices.

Posteriormente, as evidências encontradas por Ludvigson (2004) apontam para o fato de o *icc* conter informações relevantes para prever os gastos com consumo das famílias americanas quando considerado como única variável preditiva. Porém, quando incluído em uma equação com variáveis que representam a conjuntura econômica, o acréscimo na previsão dos gastos totais é apenas marginal. Os dados utilizados foram trimestrais entre os anos de 1968 e 2002, e a metodologia consistiu na mesma utilizada nos trabalhos anteriores: a comparação do grau de ajuste e da significância conjunta dos parâmetros de uma equação básica (representando os fundamentos econômicos) contra uma que inclui uma medida da confiança do consumidor.

Outro trabalho que identificou a confiança e as expectativas dos consumidores como relevantes para a previsão dos gastos de consumo nos EUA foi o de Qiao, McAleer e Wong (2009). A base de dados utilizada foi mensal entre os anos de 1985 e 2005. Os resultados indicaram a existência de uma relação causal não linear do índice de confiança do consumidor para as decisões de consumo. O estudo de Easaw e Heravi (2004), para o caso britânico, utilizou dados trimestrais de 1974

a 2000. O estudo encontrou evidências de que o sentimento do consumidor é útil para prever variações no consumo das famílias (principalmente de bens duráveis).

O estudo elaborado por Al-Eyd, Barrel e Davis (2009) envolve dados trimestrais de 1973 a 2005, para Alemanha, Itália, França, Reino Unido e EUA. As evidências apontam que o *icc* possui pouca informação relevante para a previsão dos gastos com diferentes bens de consumo quando controlado para fundamentos. Outra conclusão foi que a relação entre a confiança e o consumo reduziu nos EUA e no Reino Unido após o processo de liberalização financeira ocorrido no último quarto do século XX.

3 ANÁLISE EMPÍRICA

O regime de metas para a inflação foi adotado formalmente no Brasil em junho de 1999, após uma série de ataques especulativos contra o Real que desvalorizou fortemente a moeda brasileira. O objetivo da adoção do regime foi consolidar a estabilidade de preços alcançada com o Plano Real em 1994, de forma a propiciar uma nova âncora nominal para guiar as expectativas dos agentes. O Brasil faz parte de um importante grupo de países em desenvolvimento, os BRIC (Brasil, Rússia, Índia e China), e seu destaque o qualifica como um interessante estudo de caso para avaliar se o regime de metas para a inflação tem sido eficiente em guiar as expectativas e proporcionar um ambiente mais estável para a tomada de decisão por parte dos agentes privados.

A seguir, o trabalho analisa se a credibilidade das autoridades monetária e fiscal impactam a interpretação do contexto econômico e a formação de expectativas por parte dos consumidores, e se as expectativas e confiança dos consumidores exercem influência sobre as decisões de consumo.

3.1 Dados

O período analisado compreende o período entre janeiro de 2002 e dezembro de 2012⁵. As séries (mensais) utilizadas neste estudo são:

- **índice de confiança do consumidor** (*icc*)⁶ – este índice capta as expectativas e confiança dos consumidores relativas às condições de suas finanças, às condições econômicas do país e de suas perspectivas futuras. O índice (dessazonalizado pelo método *census X12*) varia de 0 a 200;
- **índice de bens de consumo** (*ibc*)⁷ – utilizado como variável *proxy* para a demanda por bens de consumo;

5. A justificativa para utilização deste período baseia-se na disponibilidade das séries no site do Banco Central do Brasil (BCB).

6. Série do índice de confiança do consumidor, calculada pela Federação do Comércio do Estado de São Paulo (Fecomércio), disponível no portal do Ipeadata.

7. Série de consumo aparente – bens de consumo – índice encadeado – dessazonalizado (média 2002 = 100), disponível no portal do Ipeadata.

- **índice de bens de consumo duráveis** (*ibcd*)⁸ – utilizado como variável *proxy* para a demanda de bens de consumo duráveis;
- **índice de bens de consumo semi e não duráveis** (*ibcsd*)⁹ – utilizado como variável *proxy* para a demanda de bens de consumo semi e não duráveis;
- **taxa de desemprego aberto** (*desemprego*)¹⁰ – relação entre o número de pessoas desocupadas (procurando trabalho) e o número de pessoas economicamente ativas num determinado período de referência (semana);
- **taxa de juros real *ex ante*** (*juro_real*) – série calculada pela diferença entre a taxa básica de juros (Over/Selic)¹¹ e a expectativa formada para a inflação (IPCA)¹². Pelo fato do *icc* refletir as expectativas e a confiança dos consumidores em relação à economia, a taxa de juros real *ex ante* é indicada;
- **renda esperada** (*Y_e*) – série referente à expectativa de renda pessoal (índice – média 2001 = 100 – Confederação Nacional da Indústria);
- **crédito às famílias** (*crédito*)¹³ – refere-se ao saldo das operações de empréstimo, financiamento, adiantamento e arrendamento mercantil concedidos pelas instituições do Sistema Financeiro Nacional às pessoas físicas como proporção do PIB;
- **índice de credibilidade da autoridade monetária** (*icm*) – para elaborar a série do índice de credibilidade, foi seguida a metodologia de cálculo proposta em de Mendonça (2007).¹⁴ Para o cálculo do índice foi utilizada a série de expectativas de inflação acumulada para os próximos doze meses, divulgada pelo Banco Central do Brasil. O índice assume valor igual a 1 quando a inflação anual esperada ($E(\pi)$) for igual a meta de

8. Série de consumo aparente – bens de consumo duráveis – índice encadeado – dessaz (média 2002 = 100), disponível no portal do Ipeadata.

9. Série de consumo aparente – bens de consumo semi e não duráveis – índice encadeado – dessaz (média 2002 = 100), disponível no portal do Ipeadata.

10. Série 10777 – taxa de desemprego – região metropolitana – Brasil (na semana), disponível no portal do BCB.

11. Representa o principal instrumento de política monetária no regime de metas para inflação no Brasil. Alterações nesse instrumento afetam as expectativas dos agentes e da economia. Série 4189 – taxa de juros – Selic acumulada no mês, anualizada, disponível no portal do BCB.

12. Série de expectativas de inflação (IPCA) em doze meses, disponível no portal do BCB.

13. Série 17471 – saldo das operações de crédito a pessoas físicas em relação ao PIB, disponível no portal do BCB.

14. Apesar de diferentes índices de credibilidade já terem sido propostos – como resumido nos trabalhos elaborados por de Mendonça e de Guimarães e Souza (2009) e Nahon e Meurer (2009) – e, portanto, atualmente existem outros de índices de credibilidade capazes de serem usados em análises empíricas, o presente trabalho não pretende analisar a influência e o poder de cada índice sobre o principal instrumento de política monetária no Brasil, embora essa pesquisa seja importante. Assim, a opção de usar o índice proposto por de Mendonça (2007) justifica-se nos seguintes argumentos: *i*) o índice é reconhecido pela literatura internacional, sendo este índice utilizado em vários estudos aplicados; *ii*) simplicidade de compreensão e de preparação; *iii*) o índice capta as mudanças e as flutuações na credibilidade de uma forma compatível com o regime de metas para inflação adotado no Brasil, ou seja, o índice utiliza bandas pré-determinadas de tolerância, e não bandas de tolerância ad-hoc, como proposto por outros índices; e *iv*) o índice é suficientemente rigoroso e pune adequadamente os desvios das expectativas de inflação em relação à meta de inflação.

inflação e decresce à medida que as expectativas para a inflação desviam-se da meta anunciada. O índice assume valores entre zero e um quando a expectativa de inflação situa-se dentro dos limites mínimo e máximo (π_t^*) estabelecidos para cada ano e, caso exceda um desses limites, assume valor igual a zero. O índice é calculado da seguinte maneira:

$$icm = \left\{ \begin{array}{ll} 1 & \text{se } E(\pi) = \pi_t^c \\ 1 - \frac{1}{\pi_t^* - \pi_t^c} [E(\pi) - \pi_t^c] & \text{se } \pi_{tMin}^* < E(\pi) < \text{se } \pi_{tMax}^* \\ 0 & \text{se } E(\pi) \geq \text{se } \pi_{tMax}^* \text{ ou } E(\pi) \leq \text{se } \pi_{tMin}^* \end{array} \right\};$$

- **índice de credibilidade da autoridade fiscal (*icf*)** – o estudo utiliza o índice desenvolvido por de Mendonça e Machado (2013). Uma vez que a credibilidade é essencialmente uma variável *forward-looking*, o índice de credibilidade fiscal leva em conta a informação disponibilizada por meio do relatório Focus do Banco Central do Brasil sobre as expectativas do mercado para a relação dívida/PIB. O índice considera como limites inferior e superior uma relação dívida/PIB¹⁵ de 40% e 60% respectivamente. Tais limites foram estabelecidos no Tratado de Maastricht e, de acordo com o IMF (2002), fornecem pontos de referência úteis, tendo em visto que uma razão dívida/PIB, maior que esta referência, aumenta a probabilidade de uma crise da dívida. Assim, a credibilidade fiscal é calculada da seguinte maneira:

$$icf = \left\{ \begin{array}{ll} 1 & \text{se } E_t(dlsp_{t+12}) \leq dlsp^{min} \\ 1 - \frac{1}{dlsp^{max} - dlsp^{min}} [E_t(dlsp_{t+12}) - dlsp^{min}] & \text{se } dlsp^{min} < E_t(dlsp_{t+12}) < dlsp^{max} \\ 0 & \text{se } E_t(dlsp_{t+12}) \geq dlsp^{max} \end{array} \right\}.$$

Logo, quando as expectativas de mercado acerca da relação dívida/PIB para os doze meses seguintes são inferiores a 40%, o risco de desequilíbrio fiscal é nulo e, portanto, o índice de credibilidade é igual a um. Caso as expectativas superem o valor de 60%, existe uma elevada probabilidade de desequilíbrio fiscal e, portanto, o índice de credibilidade é igual a zero. Caso as expectativas se situem entre 40% e 60%, o índice de credibilidade varia entre zero e um. Dessa maneira, espera-se a existência de uma relação positiva entre a credibilidade fiscal e a confiança do consumidor;

- ***dummy subprime* (*subprime*)** – variável *dummy* que representa os meses de maior incerteza devido à crise financeira internacional iniciada nos Estados Unidos em 2007. A variável assume o valor um, para o período de

15. A relação dívida/PIB refere-se à relação entre a dívida líquida do setor público e o PIB.

setembro de 2008 até janeiro de 2009, e o valor zero, para os demais. O estouro da crise e sua disseminação pelo mundo tiveram impacto negativo sobre o sistema financeiro e o lado real da economia de diversos países;

- **hiato do produto** (*gap*) – esta variável expressa a influência dos ciclos na economia, sendo a diferença entre o produto efetivo e sua tendência de longo prazo. O produto efetivo é o *log* do produto interno bruto dessazonalizado (método *census* X12) e deflacionado pelo IPCA¹⁶. A tendência de longo prazo do produto foi calculada utilizando-se o filtro Hodrick-Prescott;
- **índice do mercado acionário** (*ibov*) – a série do índice de preços do mercado acionário, Ibovespa, foi obtida junto ao Banco Central do Brasil (série 7845 do BCB).

A tabela 1 abaixo apresenta as correlações entre as variáveis testadas nas estimações. Em particular, são observadas correlações positivas entre a credibilidade monetária e o *icc*, a credibilidade fiscal e o *icc*, e, também, entre o *icc* e os diferentes índices de bens de consumo.

TABELA 1
Correlações

	ICC		IBC	IBCD	IBCS
Desemprego	-0,86	ICC	0,86	0,86	0,82
Juro_real	-0,79	Desemprego	-0,94	-0,93	-0,93
ICF	0,80	Juro_real	-0,89	-0,89	-0,86
ICM	0,41	GAP	0,16	0,18	0,13
Y_e	0,82	Y_e	0,82	0,84	0,77
Ibovespa	0,86	Crédito	0,97	0,97	0,94

Fonte: Tabela elaborada pelos autores.

3.2 METODOLOGIA E ESTIMAÇÕES

De modo a evitar resultados espúrios e sem significado estatístico, a estacionariedade das séries foi testada.¹⁷ Foram realizados os testes Augmented Dickey-Fuller (ADF), Phillips-Perron (PP) e Kwiatkowski-Phillips-Schmidt-Shin (KPSS). Ademais, os testes modificados de Dickey-Fuller (MADF^{GLS}) e de Phillips-Perron (MPP^{GLS}), propostos por Elliot, Rotemberg e Stock (1996), e Ng e Perron (2001) são também aplicados. Os resultados encontram-se na tabela 2 a seguir.¹⁸ Pelo fato dos testes apontarem que as séries do *crédito*, *ibov* e *icf* não são estacionárias, nas estimativas foram utilizadas as séries em primeira diferença (ou seja, $d(\text{crédito})$, $d(\text{ibov})$ e $d(\text{icf})$).

16. Série do Produto Interno Bruto (PIB), deflacionada pelo IPCA, disponível no portal do Ipeadata.

17. Todas as estimações foram feitas pelo software Eviews 7.0.

18. Nos casos divergentes encontrados (como os do *crédito* e da *credibilidade fiscal*), recorremos à observação da natureza das séries por meio da metodologia de Box-Jenkins, analisando as funções de autocorrelação.

TABELA 2
Testes de raiz unitária

Variáveis	ADF			PP			KPSS			MADFGLS			MZaGLS								
	Lags	Estatística	10% I/T, I, N	Bandwidth	Estatística	10% I/T, I, N	Bandwidth	Estatística	1% I/T, I	Lags	Estatística	10% I/T, I, N	Lags	Estatística	10% I/T, I, N						
ICC	0	-3,38	-3,15	I/T	2	-3,45	-3,15	I/T	8	0,11	0,22	I/T	12	-1,12	-2,72	I/T	0	-3,04	-14,20	I/T	
d(ICC)													0	-11,30	-2,71	I/T	0	-65,00	-14,20	I/T	
IBC	1	-3,81	-3,15	I/T	4	-4,23	-3,15	I/T	8	0,12	0,22	I/T	2	-1,76	-2,71	I/T	1	-7,01	-14,20	I/T	
d(IBC)													16	-2,46	-2,73	I/T	12	-1,12	-14,20	I/T	
IBCD	0	-3,37	-3,15	I/T	2	-3,25	-3,15	I/T	8	0,12	0,22	I/T	1	-2,45	-2,71	I/T	1	-11,86	-14,20	I/T	
d(IBCD)													0	-13,21	-2,71	I/T	0	-63,52	-14,20	I/T	
IBCSd	1	-4,22	-3,15	I/T	6	-5,22	-3,15	I/T	8	0,15	0,22	I/T	14	-0,93	-2,72	I/T	11	2,31	-5,70	I/T	
d(IBCSd)													12	-1,78	-2,72	I/T	10	0,44	-5,70	I/T	
Desemprego	12	-4,76	-3,15	I/T	7	-4,55	-3,15	I/T	7	0,06	0,22	I/T	16	-2,01	-2,73	I/T	16	-11,30	-14,20	I/T	
d(desemprego)													16	-0,80	-2,73	I/T	16	-0,37	-14,20	I/T	
Juro_real	1	-4,12	-3,15	I/T	6	-3,08	-3,15	I/T	8	0,06	0,22	I/T	1	-4,10	-2,71	I/T	2	-35,40	-14,20	I/T	
d(Juro_real)					1	-5,62	-1,62	N													
Crédito	2	2,78	-1,62	N	8	0,06	-2,58	I	9	0,17	0,22	I/T	2	-1,23	-2,71	I/T	2	-3,22	-14,20	I/T	
d(credito)					0	-6,30	-2,58	I	4	-6,34	-2,58	I	11	-1,51	-2,72	I/T	11	-2,59	-14,20	I/T	
Gap	0	-5,07	-1,62	N	5	-5,17	-1,62	N	8	0,04	0,74	I	2	-3,05	-2,71	I/T	2	-16,70	-14,20	I/T	
d(Gap)																					
Credib_Monetária	0	-2,85	-2,58	I	2	-2,94	-2,58	I	9	0,23	0,22	I/T	8	-1,67	-1,61	I	8	-5,91	-5,70	I	
d(Credib_Monetária)													9	0,09	0,74	I					
Credib_fiscal	3	0,89	-1,62	N	9	1,22	-1,62	N	9	0,19	0,22	I/T	3	-1,86	-2,71	I/T	3	-4,64	-14,20	I/T	
d(Credib_fiscal)					2	-2,97	-1,62	N	5	-2,83	-1,62	N	0	-1,53	-2,71	I/T	0	-4,55	-14,20	I/T	
y_e	0	-3,96	-3,15	I/T	9	-3,79	-3,15	I/T	7	0,12	0,22	I/T	11	-1,53	-2,72	I/T	11	-3,82	-14,20	I/T	
lbovespa	1	1,45	-1,62	N	5	1,48	-1,62	N	9	0,29	0,22	I/T	1	-1,47	-2,71	I/T	1	-5,17	-14,20	I/T	
d(lbovespa)					0	-9,70	-1,62	N	5	-9,92	-1,62	N	4	-4,72	-2,71	I/T	2	-30,72	-14,20	I/T	

Fonte: Tabela elaborada pelos autores.

Nota: ADF – a escolha final de defasagem (lag) foi feita com base no critério de Schwarz. PP e KPSS – a defasagem é o lag truncation escolhido para o Bartlett kernel. O número ótimo de defasagens nos testes MAD^{GLS} e MPP^{GLS} segue o critério de Akaike modificado (MAIC). "I" denota constante; "I/T" denota constante e tendência; e, "N" denota nada.

Para estimar as equações, foi utilizado o método de mínimos quadrados ordinários (MQO)¹⁹ e o método generalizado dos momentos (*generalized method of moments* – GMM). Uma das razões para utilizar o GMM é o fato de que este método fornece estimadores consistentes para a regressão quando verificadas endogeneidades e não linearidade (Hansen, 1982). Como apontado por Wooldridge (2001, p. 95), “*to obtain a more efficient estimator than two-stage least squares (or ordinary least squares), one must have overriding restrictions*”. A matriz de ponderação na equação foi escolhida para permitir que as estimativas por GMM sejam robustas, considerando a possível presença de heterocedasticidade e autocorrelação de forma desconhecida.²⁰

De modo a buscar evidências acerca da influência que exercem as credibilidades das autoridades fiscal e monetária sobre a percepção dos consumidores, as seguintes equações foram estimadas utilizando o *icc* como variável dependente:

$$icc_t = \beta_1 + \beta_2 desemprego_{t-6} + \beta_3 juro_real_{t-3} + \beta_4 d(ibov)_{t-5} + \beta_5 Y_e_t + \beta_6 subprime, \quad (1a)$$

$$icc_t = \beta_7 + \beta_8 desemprego_{t-6} + \beta_9 juro_real_{t-3} + \beta_{10} d(ibov)_{t-5} + \beta_{11} Y_e_t + \beta_{12} d(icf)_{t-2} + \beta_{13} icm_{t-4} + \beta_{14} subprime. \quad (1b)$$

No tocante às especificações, a escolha das defasagens foi determinada por meio da metodologia “geral para o específico”, utilizando os critérios de informação tradicionalmente aplicados, além do princípio da parcimônia e com base na teoria econômica. Ademais, foi considerado não somente a significância estatística dos parâmetros, mas também os testes de diagnósticos, com intuito de assegurar que o modelo escolhido apresenta poder explicativo (Hendry, 2001). Ademais, reconhecendo a racionalidade dos agentes e, portanto, o comportamento *forward-looking* desses agentes no processo de formação de expectativas, ou seja, que suas expectativas e percepções acerca do contexto atual e futuro de variáveis econômicas relevantes sejam capazes de afetar sua confiança, a seguinte especificação foi estimada por GMM utilizando as variáveis no instante *t*:

$$icc_t = \beta_{15} + \beta_{16} desemprego_t + \beta_{17} juro_real_t + \beta_{18} d(ibov)_t + \beta_{19} Y_e_t + \beta_{20} d(icf)_t + \beta_{21} icm_t + \beta_{22} subprime. \quad (1c)$$

A tabela 3 apresenta as estimativas. Nas estimativas por MQO, a estatística *F* indica que as regressões são significativas e os resultados dos testes *reset* de Ramsey

19. As estatísticas *t* reportadas na estimação por OLS são baseadas no estimador de Newey e West (1987), o qual é consistente na presença de heterocedasticidade e autocorrelação de forma desconhecida.

20. Ainda em relação à técnica empregada na estimação GMM, Cragg (1983) indicou que a análise de sobreidentificação possui um importante papel na seleção das variáveis instrumentais para melhorar a eficiência dos estimadores. Sendo assim, o teste *J* padrão foi realizado com o objetivo de testar esta propriedade para a validade das restrições de sobreidentificação (Hansen, 1982).

indicam que as especificações não apresentam problemas. Em relação às estimações por GMM²¹, os resultados do teste-J indicam que não se pode rejeitar a hipótese de que o modelo está corretamente especificado.

TABELA 3
Estimações por MQO e GMM para o *icc*

Váriavel dependente: <i>icc</i>					Váriavel dependente: <i>icc</i>	
Variáveis	MQO	MQO	GMM	GMM	Variáveis	GMM
	Equação 1a	Equação 1b	Equação 1a	Equação 1b		Equação 1c
	53.748	61.593	70.650 **	29.805	Constante	79.472
Constante	(50.910)	(45.366)	(33.854)	(53.990)		(55.166)
	[1.055]	[1.357]	[2.086]	[0.552]		[1.440]
	-3.199 **	-2.936 **	-4.488 ***	-3.934 ***	Desemprego _t	-5.318 ***
Desemprego _{t-6}	(1.245)	(1.161)	(0.644)	(0.827)		(0.971)
	[-2.569]	[-2.529]	[-6.963]	[-4.756]		[-5.472]
	-1.700 **	-1.778 ***	-1.161 ***	-1.191 ***	Juro_real _t	-0.675
Juro_real _{t-3}	(0.687)	(0.598)	(0.262)	(0.332)		(0.728)
	[-2.471]	[-2.971]	[-4.420]	[-3.584]		[-0.926]
	29.210 **	20.973 **	74.863 ***	35.771 **	d(ibov) _t	39.031 *
d(ibov) _{t-5}	(12.424)	(9.312)	(16.069)	(16.563)		(22.150)
	[2.351]	[2.252]	[4.658]	[2.159]		[1.762]
	1.162 ***	1.013 ***	1.071 ***	1.356 ***	Y_e _t	0.919 **
Y_e _t	(0.441)	(0.375)	(0.275)	(0.432)		(0.457)
	[2.632]	[2.703]	[3.887]	[3.139]		[2.007]
		189.542 *		201.399 *	d(icf) _t	280.766 *
d(icf) _{t-2}		(113.062)		(116.824)		(145.577)
		[1.676]		[1.724]		[1.928]
		9.740 **		7.579 ***	icm _t	11.618 ***
icm _{t-4}		(4.321)		(2.851)		(3.973)
		[2.253]		[2.658]		[2.923]
	-12.039 ***	-12.278 ***	-20.169 **	-32.359 **	Subprime	-17.188 **
Subprime	(2.100)	(2.614)	(9.508)	(14.188)		(7.718)
	[-5.731]	[-4.697]	[-2.121]	[-2.281]		[-2.227]

(Continua)

21. Os instrumentos utilizados nas estimações por GMM são:

equação 1a: *icc*(-1), desemprego (-7 até -10), *juro_real* (-4 até -8), *d(ibov)* (-6 até -8), *Y_e* (-1) *d(icf)* (-3 até -4), *icm*(-5 até -8), *selic* (-1 até -4);

equação 1b: *icc*(-1), desemprego (-7 até -10), *juro_real*(-4 até -8), *d(ibov)* (-6 até -8), *Y_e*(-1) *d(icf)* (-3 até -4), *icm*(-5 até -8), *selic*(-1 até -4);

equação 1c: *icc*(-1 até -2), desemprego (-1 até -2), *juro_real* (-1 até -5), *d(ibov)* (-1 até -4), *Y_e*(-1 até -3) *icf*(-1), *icm*(-1 até -5), *selic*, *renda_real*(-1).

(Continuação)

Váriavel dependente: <i>icc</i>					Váriavel dependente: <i>icc</i>	
Variáveis	MQO	MQO	GMM	GMM	Variáveis	GMM
	Equação 1a	Equação 1b	Equação 1a	Equação 1b		Equação 1c
R ²	0,79	0,82	0,72	0,74	R ²	0,77
R ² ajustado	0,78	0,81	0,71	0,72	R ² ajustado	0,75
Estatística J			14,28	15,20	Estatística J	17,36
P-valor (est. J)			0,76	0,58	P-valor (est. J)	0,43
Instrument rank			25	25	Instrument rank	25
Estatística F	90,03	79,42				
P-valor (est. F)	0,00	0,00				
Ramsey	0,09	0,58				
P-valor Ramsey	0,76	0,45				
Serial Correl (1)	189,75	170,93				
P-valor	0,00	0,00				
Serial Correl (2)	94,98	84,81				
P-valor	0,00	0,00				
ARCH (1)	87,04	49,12				
P-valor	0,00	0,00				
ARCH (2)	45,02	24,61				
P-valor	0,00	0,00				
Jarque-Bera	2,33	0,17				
P-valor	0,31	0,91				

Fonte: Estimções elaboradas pelos autores. Níveis de significância marginais: *** denota 0.01, ** denota 0.05 e * denota 0.1; erro-padrão entre parênteses e estatística *t* entre colchetes.

As evidências sugerem que a taxa de desemprego afeta negativamente – e com significância estatística – a confiança do consumidor em todas as especificações. Como apontado pela literatura, o desemprego é uma variável importante na avaliação dos agentes e, portanto, reflete-se na percepção e na confiança dos consumidores em relação ao cenário econômico.

Os coeficientes estimados para o efeito da taxa de juros real *ex ante* sobre o índice de confiança do consumidor são negativos em todas as especificações. Entretanto, significância estatística foi encontrada somente nos modelos que consideram o efeito defasado sobre a confiança dos agentes, sugerindo que, de fato, leva algum tempo (no caso das estimções, três meses) até que a taxa de juros afete a confiança dos consumidores. Esse resultado é condizente com o preconizado pela literatura e corrobora a teoria econômica.

Os coeficientes estimados para a variável *d(ibov)* apresentaram sinais positivos e significância para todas as estimções. Isto é, a influência do desempenho do mercado

de ações sobre a confiança do consumidor é direta. Quanto maior (menor) o índice Bovespa, maior (menor) a confiança dos consumidores em relação à economia.

Em relação aos coeficientes estimados para a renda esperada (Y_e), os resultados apontam, com significância estatística em todas as estimações, que, quando os agentes esperam rendas maiores no futuro (sendo essas expectativas formadas no instante corrente), a confiança e o otimismo dos consumidores em relação à economia aumenta.

No tocante às estimativas para o índice de credibilidade da autoridade fiscal, os resultados sugerem que o gerenciamento responsável da dívida pública afeta positivamente e com significância estatística a confiança do consumidor. Ou seja, quando são verificadas variações positivas no índice de credibilidade fiscal, isso sinaliza maior preocupação do governo com a sustentabilidade da dívida pública e, portanto com os fundamentos macroeconômicos. Quando um ambiente macroeconômico mais estável é criado, os agentes formam expectativas mais otimistas em relação ao estado da economia, o que, por conseguinte, traz um impacto positivo sobre a confiança dos consumidores.

Quanto à credibilidade da autoridade monetária, os resultados indicam uma relação positiva e significativa entre o *icm* e o *icc*. Logo, quando o Banco Central do Brasil atua no sentido de manter a inflação controlada, e consegue êxito nesta tarefa, as expectativas formadas para a inflação são afetadas e a credibilidade aumenta, ampliando, nesse sentido, a confiança dos consumidores.

Em relação à influência da crise financeira sobre a confiança do consumidor, os resultados revelam que em um contexto de crise, ocorre a deterioração das expectativas e da confiança do consumidor. Os resultados encontrados para a variável *subprime* apontam que a crise afetou negativamente e com significância estatística a confiança dos consumidores brasileiros.

A seguir é analisada a influência da confiança do consumidor (*icc*) sobre as decisões de consumo. Assim, três equações foram estimadas, cada uma possuindo um índice de bens de consumo (*ibc*, *ibcd* e *ibcsd*) como variável dependente. As variáveis independentes são as mesmas nas três equações, todavia as defasagens com que cada uma afeta as decisões de consumo variam, uma vez que cada tipo de bem possui uma característica de demanda. As equações a serem estimadas são:

$$ibc_t = \alpha_1 + \alpha_2 \text{desemprego}_{t-7} + \alpha_3 \text{juro_real}_{t-1} + \alpha_4 \text{gap}_{t-1} + \alpha_5 Y_e t + \alpha_6 d(\text{credito})_{t-5} + \alpha_7 \text{icc}_t + \alpha_8 \text{subprime}, \quad (2)$$

$$ibcd_t = \gamma_1 + \gamma_2 \text{desemprego}_{t-8} + \gamma_3 \text{juro_real}_t + \gamma_4 \text{gap}_{t-1} + \gamma_5 Y_e t + \gamma_6 d(\text{credito})_{t-5} + \gamma_7 \text{icc}_t + \gamma_8 \text{subprime}, \quad (3)$$

$$ibcsd_t = \theta_1 + \theta_2 \text{desemprego}_{t-8} + \theta_3 \text{juro_real}_t + \theta_4 \text{gap}_{t-4} + \theta_5 Y_e t + \theta_6 d(\text{credito})_{t-7} + \theta_7 \text{icc}_t + \theta_8 \text{subprime}. \quad (4)$$

Novamente, a escolha das defasagens foi determinada por meio da metodologia “geral para o específico”, a qual leva em consideração não somente a significância estatística dos parâmetros, mas também os testes de diagnósticos, com intuito de assegurar que o modelo escolhido apresenta poder explicativo (Hendry, 2001). Mais uma vez, reconhecendo a racionalidade dos agentes e, portanto, o comportamento *forward-looking* desses agentes, as seguintes especificações foram estimadas por GMM utilizando as variáveis explicativas no instante t :

$$ibc_t = \alpha_9 + \alpha_{10}desemprego_t + \alpha_{11}juro_real_t + \alpha_{12}gap_t + \alpha_{13}Y_e_t + \alpha_{14}d(credito)_t + \alpha_{15}icc_t + \alpha_{16}subprime, \quad (5)$$

$$ibcd_t = \gamma_9 + \gamma_{10}desemprego_t + \gamma_{11}juro_real_t + \gamma_{12}gap_t + \gamma_{13}Y_e_t + \gamma_{14}d(credito)_t + \gamma_{15}icc_t + \gamma_{16}subprime, \quad (6)$$

$$ibcsd_t = \theta_9 + \theta_{10}desemprego_t + \theta_{11}juro_real_t + \theta_{12}gap_t + \theta_{13}Y_e_t + \theta_{14}d(credito)_t + \theta_{15}icc_t + \theta_{16}subprime. \quad (7)$$

A tabela 4 apresenta os resultados das estimações por MQO e GMM²² para as equações 2, 3 e 4. Por sua vez, a tabela 5 apresenta os resultados das estimações por GMM²³ para as equações 5, 6 e 7.

TABELA 4
Estimações por MQO e GMM (equações 2, 3 e 4)

Variável dependente: IBC (equação 2)			Variável dependente: IBCD (equação 3)			Variável dependente: IBCSD (equação 4)		
Variáveis	MQO	GMM	Variáveis	MQO	GMM	Variáveis	MQO	GMM
	100.079 ***	84.071 ***		7.121	16.058		100.563 ***	91.855 ***
Constante	(17.232)	(16.275)	Constante	(65.963)	(68.007)	Constante	(12.341)	(11.773)
	[5.807]	[5.165]		[0.107]	[0.236]		[8.148]	[7.801]
	-3.593 ***	-3.749 ***		-10.441 ***	-11.593 ***		-2.071 ***	-2.249 ***
Desemprego _{t-7}	(0.378)	(0.331)	Desemprego _{t-8}	(1.251)	(0.747)	Desemprego _{t-8}	(0.263)	(0.214)
	[-9.482]	[-11.298]		[-8.341]	[-15.518]		[-7.871]	[-10.474]

(Continua)

22. Os instrumentos utilizados nas estimações por GMM são:

equação 2: $ibc(-1 \text{ até } -2)$, $desemprego(-8 \text{ até } -12)$, $juro_real(-2 \text{ até } -4)$, $gap(-2 \text{ até } -8)$, $Y_e(-1)$, $d(credito)(-6)$, $icc(-1)$; equação 3: $ibcd(-1 \text{ até } -2)$, $desemprego(-9 \text{ até } -10)$, $juro_real(-1 \text{ até } -8)$, $gap(-2 \text{ até } -6)$, $Y_e(-1)$, $d(credito)(-6 \text{ até } -7)$, $icc(-1 \text{ até } -6)$;

equação 4: $ibcsd(-1 \text{ até } -3)$, $desemprego(-9 \text{ até } -10)$, $juro_real(-1 \text{ até } -6)$, $gap(-5 \text{ até } -8)$, $Y_e(-1)$, $d(credito)(-8 \text{ até } -11)$, $icc(-1)$, $selic(-1 \text{ até } -3)$.

23. Os instrumentos utilizados nas estimações por GMM são:

equação 5: $ibc(-1)$, $desemprego(-1 \text{ até } -3)$, $juro_real(-1 \text{ até } -4)$, $gap(-1 \text{ até } -3)$, $Y_e(-1)$, $d(credito)(-1 \text{ até } -5)$, $icc(-1 \text{ até } -10)$, icm , icf ;

equação 6: $ibcd(-1)$, $desemprego(-1 \text{ até } -3)$, $juro_real(-1 \text{ até } -6)$, $gap(-1 \text{ até } -4)$, $Y_e(-1)$, $d(credito)(-1 \text{ até } -6)$, $icc(-1 \text{ até } -10)$, icm , $icf(-1 \text{ até } -2)$;

equação 7: $ibcsd(-1)$, $desemprego(-1)$, $juro_real(-1 \text{ até } -6)$, $Y_e(-1)$, $d(credito)(-1 \text{ até } -4)$, $icc(-1 \text{ até } -11)$, icm , icf , $selic(0 \text{ até } -1)$, $taxa_de_cambio$.

(Continuação)

Variável dependente: IBC (equação 2)			Variável dependente: IBCD (equação 3)			Variável dependente: IBCSD (equação 4)		
Variáveis	MQO	GMM	Variáveis	MQO	GMM	Variáveis	MQO	GMM
Juro_real _{t-1}	-1.380 *** (0.168) [-8.184]	-1.214 *** (0.152) [-7.985]	Juro_real _t	-3.928 *** (0.656) [-5.984]	-3.934 *** (0.537) [-7.320]	Juro_real _t	-0.762 *** (0.119) [-6.373]	-0.623 *** (0.088) [-7.032]
Gap _{t-1}	49.425 *** (16.359) [3.021]	70.528 *** (22.384) [3.151]	Gap _{t-1}	226.886 *** (71.263) [3.183]	290.256 *** (59.094) [4.911]	Gap _{t-4}	45.122 *** (15.745) [2.865]	64.872 *** (15.989) [4.057]
Y_e _t	0.464 *** (0.145) [3.193]	0.601 *** (0.128) [4.671]	Y_e _t	2.233 *** (0.607) [3.679]	2.391 *** (0.591) [4.038]	Y_e _t	0.232 ** (0.113) [2.044]	0.291 *** (0.092) [3.146]
d(crédito) _{t-5}	7.388 * (3.845) [1.921]	11.729 *** (4.976) [2.356]	d(crédito) _{t-5}	24.039 * (13.281) [1.810]	33.949 ** (15.176) [2.236]	d(crédito) _{t-7}	4.978 * (2.848) [1.747]	7.241 * (4.033) [1.795]
icc _t	0.081 ** (0.038) [2.118]	0.084 * (0.044) [1.894]	icc _t	0.288 ** (0.127) [2.257]	0.175 * (0.104) [1.686]	icc _t	0.055 * (0.031) [1.750]	0.072 ** (0.036) [1.998]
Subprime	-5.696 *** (1.577) [-3.611]	-6.385 *** (1.823) [-3.502]	Subprime	-18.595 *** (5.802) [-3.204]	-23.537 *** (7.751) [-3.036]	Subprime	-4.625 *** (1.043) [-4.431]	-7.300 *** (2.352) [-3.103]
R ²	0,95	0,95	R ²	0,94	0,94	R ²	0,92	0,92
R ² ajustado	0,95	0,95	R ² ajustado	0,93	0,93	R ² ajustado	0,91	0,91
Estatística J		13,44	Estatística J		14,68	Estatística J		16,03
P-valor (est. J)		0,41	P-valor (est. J)		0,74	P-valor (est. J)		0,52
Instrument rank		21	Instrument rank		27	Instrument rank		25
Estatística F	335,83		Estatística F	261,46		Estatística F	197,98	
P-valor (est. F)	0,00		P-valor (est. F)	0,00		P-valor (est. F)	0,00	
Ramsey	0,32		Ramsey	1,19		Ramsey	6,62	
P-valor Ramsey	0,57		P-valor Ramsey	0,27		P-valor Ramsey	0,02	
Serial Correl (1)	20,74		Serial Correl (1)	37,34		Serial Correl (1)	14,83	
P-valor	0,00		P-valor	0,00		P-valor	0,00	
Serial Correl (2)	11,02		Serial Correl (2)	21,29		Serial Correl (2)	8,27	
P-valor	0,00		P-valor	0,00		P-valor	0,00	
Arch (1)	2,50		Arch (1)	7,36		Arch (1)	3,90	
P-valor	0,11		P-valor	0,00		P-valor	0,05	
Arch (2)	1,52		Arch (2)	4,14		Arch (2)	1,43	
P-valor	0,22		P-valor	0,02		P-valor	0,24	
Jarque-Bera	1,87		Jarque-Bera	2,51		Jarque-Bera	0,52	
P-valor	0,39		P-valor	0,28		P-valor	0,77	

Fonte: Estimções elaboradas pelos autores. Níveis de significância marginais: *** denota 0.01, ** denota 0.05 e * denota 0.1; erro-padrão entre parênteses e estatística *t* entre colchetes.

TABELA 5
Estimações por GMM (equações 5, 6 e 7)

Variável dependente: IBC		Variável dependente: IBCD		Variável dependente: IBCSD	
Variáveis	GMM	Variáveis	GMM	Variáveis	GMM
Constante	50.675 *** (19.018) [2.664]	Constante	-209.323 *** (41.455) [-5.049]	Constante	90.278 *** (12.321) [7.327]
Desemprego _t	-3.731 *** (0.480) [-7.769]	Desemprego _t	-10.053 *** (1.339) [-7.505]	Desemprego _t	-2.421 *** (0.270) [-8.967]
Juro_real _t	-0.739 *** (0.174) [-4.247]	Juro_real _t	-2.198 *** (0.473) [-4.639]	Juro_real _t	-0.456 *** (0.122) [-3.730]
Gap _t	1.142 (12.568) [0.090]	Gap _t	99.712 * (57.017) [1.748]	Gap _t	-11.563 (15.684) [-0.737]
Y_e _t	0.830 *** (0.180) [4.598]	Y_e _t	4.019 *** (0.415) [9.664]	Y_e _t	0.319 *** (0.109) [2.931]
d(crédito) _t	8.989 ** (4.207) [2.136]	d(crédito) _t	22.745 ** (10.917) [2.083]	d(crédito) _t	2.415 (2.918) [0.827]
icc _t	0.101 * (0.054) [1.857]	icc _t	0.257 * (0.154) [1.669]	icc _t	0.055 * (0.029) [1.892]
Subprime	-4.312 * (2.209) [-1.951]	Subprime	-15.365 ** (6.151) [-2.497]	Subprime	-4.243 *** (1.155) [-3.671]
R ²	0,93	R ²	0,92	R ²	0,91
R ² ajustado	0,93	R ² ajustado	0,91	R ² ajustado	0,90
Estatística J	18,70	Estatística J	18,90	Estatística J	14,94
P-valor (est. J)	0,66	P-valor (est. J)	0,87	P-valor (est. J)	0,86
Instrument rank	30	Instrument rank	35	Instrument rank	30

Fonte: estimações elaboradas pelos autores. Níveis de significância marginais: *** denota 0.01, ** denota 0.05 e * denota 0.1; erro-padrão entre parênteses e estatística t entre colchetes.

As estimações apontam o impacto negativo que o *desemprego* desempenha sobre o consumo. Os coeficientes da variável *desemprego* foram negativos e estatisticamente significativos em todas as estimações. Os achados sugerem que um aumento (queda) no desemprego exerce impacto negativo (positivo) sobre todos os índices de bens de consumo.

A taxa de juros é um fator determinante para que os agentes aloquem de forma ótima o seu consumo ao longo do tempo, de maneira a postergar ou antecipar gastos. Os resultados encontrados em todas as estimações sugerem este comportamento ao apresentarem sinais negativos e significativos para a relação entre o *juro_real* e o consumo de bens. Um aumento (queda) dos juros leva os agentes a postergar (antecipar) suas decisões de consumo no tempo.

Para medir a sensibilidade do consumo à atividade econômica, o hiato do produto foi incluído na análise. As estimativas apresentadas na tabela 4 apontam – com significância estatística para todas as equações – a relação positiva entre o *gap* e os gastos de consumo dos agentes. Por sua vez, as estimações apresentadas na tabela 5 apontam para uma relação positiva entre atividade econômica e consumo, entretanto, significância estatística foi encontrada somente na estimação referente aos bens de consumo duráveis. De um modo geral, as evidências sugerem que o consumo seja maior (menor) quanto mais aquecida (desaquecida) a economia estiver.

As estimativas para todas as especificações nas tabelas 4 e 5 apontam que a renda esperada (*Y_e*) afeta positivamente e com significância estatística todos os índices de bens de consumo. Assim, os achados sugerem que o consumo seja maior (menor) quanto maior (menor) a renda esperada for.

Os coeficientes estimados para a variável *d(crédito)* apresentaram sinais positivos e significância estatística em quase todas as estimações (a exceção foi a especificação para o IBCSD na tabela 5, o qual não foi significativo). De um modo geral, os resultados apontam para importância do crédito nas decisões de consumo, pois amplia o poder de compra dos consumidores.

No tocante à influência da confiança do consumidor (*icc*) sobre o consumo dos agentes, os resultados de todas as estimações apontam para existência de relações positivas e estatisticamente significativas, revelando que a confiança do consumidor brasileiro influencia os gastos com consumo. As expectativas formadas pelos consumidores, fundamentadas em percepções de sua situação financeira e dos rumos da economia, impactam as decisões de consumo, representando, nesse sentido, um canal de transmissão de choques para a economia que não deve ser negligenciado.

Por sua vez, os resultados encontrados para *dummy subprime* apontam que a crise financeira internacional de 2008 atingiu de forma negativa o consumidor brasileiro. Os coeficientes apresentaram sinais negativos e significância estatística para todas as estimações.

Outra análise pode ser depreendida dos resultados obtidos. É possível observar que as mesmas variáveis impactam cada categoria de bens de consumo com defasagens (tabela 4) e magnitudes diferentes (tabelas 4 e 5). Os bens de consumo duráveis e os semi e não duráveis possuem como principal diferença o tempo de durabilidade: enquanto os bens de consumo duráveis têm como característica

principal sua durabilidade, podendo ser usados diversas vezes ao longo do tempo, os bens de consumo semi e não duráveis são consumidos imediatamente ou se desgastam em um curto espaço de tempo. Podemos citar como exemplos de bens duráveis veículos, eletrodomésticos e máquinas, enquanto bens semi e não duráveis são alimentos, calçados e roupas, entre outros. Dos exemplos citados, percebe-se que as características de gastos com esses bens também são distintas. O consumo de bens semi e não duráveis ocorre de acordo com a necessidade diária, enquanto o consumo de bens duráveis exige um planejamento e um horizonte de gastos mais longo. Ou seja, de acordo com o argumento apresentado por Katona (1968), o consumo de bens duráveis depende tanto da vontade de consumir quanto da capacidade de consumir, enquanto o consumo de bens semi e não duráveis depende primordialmente da necessidade e da capacidade de consumir dos agentes.

A análise dos coeficientes estimados permite observar que o impacto do desemprego, do juro real, da renda esperada, do crédito, da atividade econômica (gap) e da confiança do consumidor (icc), bem como os efeitos da crise do *subprime*, são maiores sobre as decisões de consumo de bens duráveis. Ou seja, o consumo de bens duráveis é mais sensível às flutuações no ambiente macroeconômico. Vale destacar, por meio dos resultados encontrados, que o impacto das expectativas do consumidor e da confiança sobre o consumo não pode ser desprezado, principalmente, sobre o consumo de bens duráveis, pois estes se mostraram mais sensíveis às expectativas e à confiança dos consumidores que os bens de consumo semi e não duráveis.

Até então os resultados das estimações apontam que as credibilidades fiscal e monetária influenciam a confiança dos consumidores e que a confiança dos consumidores é relevante para o processo de decisão de consumo dos agentes. Além disso, as estimações apontam a maior sensibilidade dos bens de consumo duráveis em relação ao ambiente econômico.

Para dar maior robustez às estimações individuais realizadas por MQO e GMM e observar a influência do ambiente econômico sobre as decisões de consumo, são realizadas estimações por meio de sistemas de equações simultâneas. Cada sistema tem um índice de bem de consumo como variável dependente. Para tratar possíveis problemas de endogeneidade e simultaneidade as estimações do sistema foram feitas por GMM²⁴. Os sistemas de equações estimados são:

24. Os sistemas utilizam as seguintes variáveis instrumentais:

S1 – equação icc: icc(-1 até -3) desemprego(-1 até -4) juro_real(-1 até -4) d(ibov)(-1 até -4) Y_e(-1) selic(-1) icm(-1 até -5) icf(-1); equação ibc: ibc(-1) desemprego(-1 até -3) juro_real(-1 até -3) gap(-1 até -7) icc(-1 até -8) d(crédito)(-1) Y_e(-1) selic(-1) taxa_de_cambio(0 até -2) icm icf;

S2 – equação icc: icc(-1 até -3) desemprego(-1 até -4) juro_real(-1 até -4) d(ibov)(-1 até -4) selic(-1) icm(-1 até -5) icf(-1) Y_e(-1); equação ibcd: ibcd(-1 até -3) desemprego(-1) juro_real(-1 até -2) gap(-1 até -2) icc(-1 até -8) d_crédito(-1) taxa_de_cambio(0 até -2) selic(-1 até -2) icm icf Y_e(-1);

S3 – equação icc: icc(-1 até -2) desemprego(-1 até -4) juro_real(-1 até -4) d(ibov)(-1 até -4) Y_e(-1) icm(-1 até -5) icf(-1) selic(-1) crédito(-1 até -3); equação ibcsd: ibcsd(-1 até -2) desemprego(-1 até -2) juro_real(-1 até -5) icc(-1 até -9) icm(0 até -3) icf selic(0 até -1) Y_e(-1) gap(-1 até -3) taxa_de_cambio.

$$\begin{aligned}
 (S1) \quad & \left\{ \begin{aligned}
 & icc_t = \rho_1 + \rho_2 desemprego_t + \rho_3 juro_real_t + \rho_4 d(ibov)_t \\
 & \quad + \rho_5 Y_e_t + \rho_6 d(icf)_t + \rho_7 icm_t + \rho_8 subprime \\
 & ibc_t = \tau_1 + \tau_2 desemprego_t + \tau_3 juro_real_t + \tau_4 gap_t \\
 & \quad + \tau_5 d(credito)_t + \tau_6 Y_e_t + \tau_7 icc_t + \tau_8 subprime,
 \end{aligned} \right. \\
 (S2) \quad & \left\{ \begin{aligned}
 & icc_t = w_1 + w_2 desemprego_t + w_3 juro_real_t + w_4 d(ibov)_t \\
 & \quad + w_5 Y_e_t + w_6 d(icf)_t + w_7 icm_t + w_8 subprime \\
 & ibcd_t = \phi_1 + \phi_2 desemprego_t + \phi_3 juro_real_t + \phi_4 gap_t \\
 & \quad + \phi_5 d(credito)_t + \phi_6 Y_e_t + \phi_7 icc_t + \phi_8 subprime,
 \end{aligned} \right. \\
 (S3) \quad & \left\{ \begin{aligned}
 & icc_t = \eta_1 + \eta_2 desemprego_t + \eta_3 juro_real_t + \eta_4 d(ibov)_t \\
 & \quad + \eta_5 Y_e_t + \eta_6 d(icf)_t + \eta_7 icm_t + \eta_8 subprime \\
 & ibcsd_t = \pi_1 + \pi_2 desemprego_t + \pi_3 juro_real_t + \pi_4 gap_t \\
 & \quad + \pi_5 d(credito)_t + \pi_6 Y_e_t + \pi_7 icc_t + \pi_8 subprime.
 \end{aligned} \right.
 \end{aligned}$$

As tabelas 6, 7 e 8 mostram as estimativas dos sistemas S1, S2 e S3, respectivamente.

Os resultados dos testes J indicam que não se pode rejeitar a hipótese de que os sistemas estão corretamente especificados. Além disso, todos os coeficientes estimados estão com seus sinais de acordo com o esperado. A significância estatística não foi encontrada somente para os coeficientes relativos ao efeito da taxa de juros real sobre a confiança do consumidor (icc); os demais coeficientes estimados nos três sistemas apresentaram todos eles significância estatística. De modo geral, os achados apresentados pelos sistemas confirmam os resultados observados nas equações individuais.

As evidências encontradas corroboram a ideia de que a atuação das autoridades econômicas no sentido de manter a estabilidade do ambiente econômico é fundamental para a formação de expectativas dos agentes. Isto é, a credibilidade fiscal e a credibilidade monetária atuam sobre confiança dos consumidores, o que, por sua vez, é levado em consideração pelos agentes na hora do consumo, principalmente de bens duráveis. Esse encadeamento de credibilidade-confiança-consumo caracteriza um possível canal de transmissão via expectativas dos agentes.

Portanto, os resultados das estimações apontam que as credibilidades fiscal e monetária influenciam a formação de expectativas e a confiança dos consumidores, e que a confiança dos consumidores é relevante para o processo de decisão de consumo dos agentes. Além disso, as estimações apontam a maior sensibilidade dos bens de consumo duráveis em relação ao ambiente econômico.

TABELA 6
Sistema GMM (S1 – ICC e IBC)

Variáveis dependentes	ICC		IBC
Variáveis independentes		Variáveis independentes	
	46.178		72.658 ***
Constante	(34.911)	Constante	(12.697)
	[1.322]		[5.722]
	-5.348 ***		-4.341 ***
Desemprego _t	(0.499)	Desemprego _t	(0.296)
	[-10.709]		[-14.629]
	-0.234		-0.616 ***
Juro_real _t	(0.511)	Juro_real _t	(0.131)
	[-0.458]		[-4.672]
	38.541 **		17.795 *
d(ibov) _t	(15.008)	Gap _t	(10.587)
	[2.567]		[1.681]
	1.192***		9.174 ***
Y_e _t	(0.282)	d(crédito) _t	(2.500)
	[4.219]		[3.669]
	12.261 ***		0.726 ***
icm _t	(2.469)	Y_e _t	(0.103)
	[4.965]		[7.006]
	209.063 *		0.057 **
d(icf) _t	(112.182)	icc _t	(0.028)
	[1.863]		[2.034]
	-17.750 ***		-6.503 ***
Subprime	(4.863)	Subprime	(1.192)
	[-3.649]		[-5.453]
R ²	0,77		0,93
R ² ajustado	0,75		0,93
Estatística J	0,17		
P-valor (est. J)	0,98		

Fonte: Estimções elaboradas pelos autores. Níveis de significância marginais: *** denota 0.01, ** denota 0.05 e * denota 0.1; erro-padrão entre parênteses e estatística t entre colchetes.

TABELA 7
Sistema GMM (S2 – ICC e IBCD)

Variáveis dependentes	ICC		IBCD
Variáveis independentes		Variáveis independentes	
	50.996		77.124 ***
Constante	(42.942)	Constante	(14.211)
	[1.187]		[5.426]
	-5.387 ***		-4.413 ***
Desemprego _t	(0.515)	Desemprego _t	(0.339)
	[-10.449]		[-12.989]
	-0.245		-0.650 ***
Juro_real _t	(0.517)	Juro_real _t	(0.134)
	[-0.474]		[-4.818]
	35.529 **		21.831 *
d(ibov) _t	(15.522)	Gap _t	(12.127)
	[2.288]		[1.800]
	1.156 ***		6.509 *
Y_e _t	(0.350)	d(crédito) _t	(3.389)
	[3.301]		[1.920]
	11.528 ***		0.702 ***
icm _t	(2.387)	Y_e _t	(0.110)
	[4.828]		[6.375]
	246.643 **		0.052 *
d(icf) _t	(114.330)	icc _t	(0.031)
	[2.157]		[1.660]
	-18.579 ***		-6.912 ***
Subprime	(4.989)	Subprime	(1.257)
	[-3.724]		[-5.496]
R ²	0,77		0,94
R ² ajustado	0,76		0,93
Estatística J		0,18	
P-valor (est. J)		0,87	

Fonte: Estimações elaboradas pelos autores. Níveis de significância marginais: *** denota 0.01, ** denota 0.05 e * denota 0.1; erro-padrão entre parênteses e estatística t entre colchetes.

TABELA 8
Sistema GMM (S3 – ICC e IBCSD)

Variáveis dependentes	ICC		IBCSD
Variáveis independentes		Variáveis independentes	
Constante	34.795 (32.097) [1.084]	Constante	71.285 *** (13.034) [5.469]
Desemprego _t	-5.373 *** (0.426) [-12.597]	Desemprego _t	-4.339 *** (0.256) [-16.900]
Juro_real _t	-0.044 (0.440) [-0.101]	Juro_real _t	-0.599 *** (0.131) [-4.567]
d(ibov) _t	25.099 * (13.769) [1.822]	gap _t	22.051 ** (10.277) [2.145]
Y_e _t	1.283 *** (0.258) [4.974]	d(crédito) _t	8.942 *** (2.332) [3.833]
icm _t	13.211 *** (2.642) [5.000]	Y_e _t	0.723 *** (0.104) [6.954]
d(icf) _t	152.261 * (79.324) [1.919]	icc _t	0.068 *** (0.025) [2.740]
Subprime	-19.457 *** (3.419) [-5.689]	Subprime	-5.103 *** (1.248) [-4.089]
R ²	0,78		0,93
R ² ajustado	0,77		0,93
Estatística J	0,18		
P-valor (est. J)	0,96		

Fonte: Estimções elaboradas pelos autores. Níveis de significância marginais: *** denota 0.01, ** denota 0.05 e * denota 0.1; erro-padrão entre parênteses e estatística t entre colchetes.

4 CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo oferecer evidências de que o sucesso dos *policymakers* em atingir seus objetivos e construir um ambiente econômico mais estável auxilia no processo de tomada de decisões dos agentes privados (mais precisamente,

consumidores), pois afeta positivamente as expectativas e aumenta a confiança dos consumidores na economia. Desde a adoção do regime de metas para a inflação, em 1999, o banco central e a autoridade fiscal brasileira têm buscado construir suas reputações e ampliar a credibilidade do regime, de forma a reduzir seus esforços e propiciar um ambiente menos incerto para a formação de expectativas.

O presente estudo utiliza o *índice de confiança do consumidor* como variável central da análise, e busca identificar os fatores determinantes dessa confiança (com destaque para as credibilidades monetária e fiscal), como também, a influência que essa confiança exerce sobre as decisões de consumo. Assim, o trabalho analisa os fatores da conjuntura econômica capazes de impactar a confiança dos consumidores, e se a credibilidade exerce alguma influência sobre a percepção dos agentes acerca do ambiente econômico e na formação de suas expectativas. Em seguida, analisa o impacto da confiança dos consumidores sobre diferentes categorias de bens de consumo.

Os resultados encontrados sugerem que a credibilidade das políticas monetária e fiscal exerce influência sobre a confiança dos consumidores brasileiros. Além disso, a confiança do consumidor é afetada por variáveis da conjuntura econômica, como o desemprego, a taxa de juros e o crédito, além de ter sido abalada pela crise do *subprime*. Ou seja, a confiança do consumidor reflete tanto fatores do cenário macroeconômico quanto a percepção dos agentes acerca da condução das políticas econômicas.

A análise empírica sugere ainda que a confiança do consumidor brasileiro influencia os gastos com consumo. As expectativas formadas pelos agentes, fundamentadas em percepções de sua situação financeira e dos rumos da economia, impactam as decisões de consumo, representando, nesse sentido, um importante canal de transmissão das políticas econômicas.

REFERÊNCIAS

AGÉNOR, P.; TAYLOR, M. P. Testing for credibility effects. **IMF Staff Papers**, 39, September, p. 545-71, 1992.

_____. Analysing credibility in high-inflation countries: a new approach. **Economic Journal**, 103, p. 329-36, 1993.

AL-EYD, A.; BARREL, R.; DAVIS, E. P. Consumer confidence indices and short-term forecasting of consumption. **The Manchester School**, v. 77, n. 1, p. 96-111, 2009.

BARRO, R. J.; GORDON, D. B. A Positive Theory of Monetary Policy in a Natural Rate Model. **Journal of Political Economy**. v. 91, n. 4, p. 589-610, 1983a.

_____. Rules, Discretion and Reputation in a Model of Monetary Policy. **Journal of Monetary Economics**. 12 (1), p. 101-121, 1983b.

- BRAM, J.; LUDVIGSON, S. Does consumer confidence forecast household expenditure? A sentiment index horse race. **Federal Reserve Bank of New York: Economic Policy Review**. v. 4, n. 2, p. 59-78, 1998.
- CARROLL, C. D.; FUHRER, J. C.; WILCOX, D. W. Does Consumer Sentiment Forecast Household Spending? If So Why? **American Economic Review**. v. 84, n. 5, p. 1397-1408, 1994.
- CECCHETTI, S. G.; KRAUSE, S. Central Bank Structure, Policy Efficiency and Macroeconomic Performance: Exploring Empirical Relationships. **Federal Reserve Bank of St. Louis, Review**, v. 84, n.4, p. 45-60, 2002.
- CUKIERMAN, A. Central Bank Strategy, Credibility, and Independence: Theory and Evidence. **Cambridge, Mass.: The MIT Press**, 1992.
- CUKIERMAN, A.; MELTZER, A. H. A theory of ambiguity, credibility and inflation under discretion and asymmetric information. **Econometrica**, 54, p. 1099-128, 1986.
- DE MENDONÇA, H. F. Towards credibility from inflation targeting: the Brazilian experience. **Applied Economics** v. 39, n. 20, p. 2599-2615, 2007.
- _____. Brazil: how macroeconomic variables affect consumer confidence? **Cepal Review** v. 99, p. 81-94, 2009.
- DE MENDONÇA, H. F.; DE GUIMARÃES E SOUZA, G. J. Inflation targeting credibility and reputation: the consequences for the interest rate. **Economic Modelling**, v. 26, n. 6, p. 1228-1238, 2009.
- DE MENDONÇA, H. F.; MACHADO, M. R. Public debt management and credibility: evidence from an emerging economy. **Economic Modelling**, v. 30, n. 1, p. 10-21, 2013.
- EASAW, J. Z.; HERAVI, S. M. Evaluating consumer sentiments as predictors for UK household consumption behavior: Are they accurate and useful? **International Journal of Forecasting**, v. 20, n. 4, p. 671-681, 2004.
- ELLIOT, G.; ROTHENBERG, T. J.; STOCK, J. H. Efficient tests for an autoregressive unit root. **Econometrica**, v. 64, n. 4, p. 813-836, 1996.
- EPPRIGHT, D. R.; ARGUEA, N. M.; HUTH, W. L. Aggregate consumer expectation indexes as indicators of future consumer expenditures. **Journal of Economic Psychology**, v. 19, n. 2, p. 215-235, 1998.
- FUHRER, J. What role does consumer sentiment play in the US macroeconomy? **New England Economic Review**, Boston, Federal Reserve Bank of Boston, January-February, p. 32-44, 1993.

HENDRY, D.F. Achievements and challenges in econometric methodology. **Journal of Econometrics**, v. 100, n. 1, p. 7-10, 2001.

IMF. **Assessing sustainability**. Washington D.C, 2002.

Katona, G. **Psychological Analysis of Economic Behavior**. McGraw-Hill, 1951.

_____. **The powerful consumer**: Psychological studies of the American economy. New York: McGraw-Hill, 1960.

_____. Consumer behavior: Theory and findings on expectations and aspirations. **The American Economic Review**, v. 58, n. 2, p. 19-30, 1968.

_____. Psychology and consumer economics. **Journal of Consumer Research**, v. 1, n. 1, p. 1-8, 1974.

_____. **Psychological Economics**. New York: Elsevier Scientific Publishing Company, 1975.

KYDLAND, F.E.; PRESCOTT, E. C. Rules Rather than Discretion: the Inconsistency of Optimal Plans. **Journal of Political Economic**. v. 85, n. 3, p. 473-492, 1977.

LI, C. Consumer expectation and output growth: The case of China. **Economics Letters**, v. 113, n. 3, p. 298-300, 2011.

LUDVIGSON, S. Consumer confidence and consumer spending. **Journal of Economic Perspectives**, v. 18, n. 2, p. 20-50, 2004.

MISHKIN, F. S. **Monetary Policy Strategy**. The MIT Press, 2007.

MONTES, G. C. Credibility and monetary transmission channels under inflation targeting: an econometric analysis from a developing country. **Economic Modelling**, v. 30, p. 670-684, 2013.

MONTES, G. C.; BASTOS, J. C. A. Metas de inflação e estrutura a termo das taxas de juros no Brasil. **Economia Aplicada**, v. 15, n. 3, p. 391-415, 2011.

_____. Economic policies, macroeconomic environment and entrepreneurs' expectations: Evidence from Brazil. **Journal of Economic Studies**, v. 40, n. 3, p. 334-354, 2013.

_____. Effects of reputation and credibility on monetary policy: Theory and evidence for Brazil. **Journal of Economic Studies**, v. 41, n. 3, 2014.

MONTES, G. C.; MACHADO, C. C. Efeitos da credibilidade e reputação sobre a taxa Selic e a transmissão da política monetária para o investimento agregado pelo canal dos preços dos ativos. **Pesquisa e Planejamento Econômico (PPE)**, v. 44, n. 2, p. 241-287, 2014.

NAHON, B. F.; MEURER, R. Measuring Brazilian Central Bank Credibility Under Inflation Targeting. **International Research Journal of Finance and Economics**, 27, p. 72-81, 2009.

NEWKEY, W. K.; WEST, K. D. A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. **Econometrica**, v. 55, n. 3, p. 703-708, 1987.

NG, S.; PERRON, P. Lag length selection and the construction of unit root tests with good size and power. **Econometrica**, v. 69, n. 6, p. 1519-1554, 2001.

PERSSON, T.; TABELLINI, G. **Macroeconomics Policy, Credibility and Politics**. London: Harwood Academic, 1990.

QIAO, Z.; MCALEER, M.; WONG, W. Linear and nonlinear causality between changes in consumption and consumer attitudes. **Economics Letters**, v. 102, n. 3, p. 161-164, 2009.

ROGOFF, K. The optimal degree of commitment to an intermediate monetary target. **The Quarterly Journal of Economics**, v. 100, n. 4, p. 1169-1189, 1985.

STARR, M. A. Consumption, sentiment, and economic news. **Economic Inquiry**, v. 50, n. 4, p. 1097-1111, 2012.

VUCHELEN, J. Consumer sentiment and macroeconomic forecast. **Journal of Economic Psychology**, v. 25, n. 4, p. 493-506, 2004.

WALSH, C. E. Optimal contracts for central bankers. **The American Economic Review**, v. 85, n. 1, p. 150-67, 1995.

(Originals submetidos em outubro de 2013. Última versão recebida em dezembro de 2014. Aprovada em março de 2015.)

ELEIÇÃO NAS ESCOLAS: UMA ANÁLISE DO IMPACTO DO DIRETOR ELEITO SOBRE O DESEMPENHO EDUCACIONAL NO ESTADO DA BAHIA

Fernanda Julyanna Silva dos Santos¹

Raquel Menezes Bezerra Sampaio²

Luciano Menezes Bezerra Sampaio³

Analisam-se os efeitos da eleição do diretor escolar sobre o desempenho dos alunos por meio dos modelos de dados em painel e diferença em diferenças. Comparam-se os desempenhos das escolas municipais e estaduais da Bahia, tendo, estas últimas, adotado a eleição desde 2008. São utilizados os dados da Prova Brasil de 2007 (anterior à política), 2009 e 2011. Os resultados mostram que em geral não houve aumento no desempenho dos alunos devido à adoção da política e, quando significativa, o efeito foi negativo – caso apenas da nota de Matemática no modelo de diferença em diferenças com os dados de 2007 e 2011. Constaram-se ainda reduções sobre as taxas de aprovação e abandono escolar no período.

Palavras-chave: seleção de diretores; desempenho escolar; diferença em diferenças.

ELECTION IN SCHOOLS: AN ANALYSIS OF PRINCIPAL'S ELECTION ON THE STUDENTS' PROFICIENCY LEVELS IN THE STATE OF BAHIA

The main goal of this paper is to analyze the school's principal's election on the students' proficiency levels. We compare the school's proficiency between state and city schools in Bahia, where a selection by election was implemented only in city schools (the control group), after 2008. The databases used are Prova Brasil 2007 (before the change), 2009 and 2011. Our results shows that in general there were no gains on mean proficiency levels both in Mathematics and in Portuguese exams due to principals' elections, and, when this effect was significant it was negative – only for Mathematics in the difference-in-difference model for the 2007-2011 period. We also documented a reduction in attainment and dropout rates.

Keywords: principals' selection; school performance; difference-in-differences.

JEL: C21; C54; I21.

1. Professora na Escola de Saúde da Universidade Federal do Rio Grande do Norte (UFRN) e Doutoranda em Administração do PPGA da UFRN. *E-mail:* <fernandajusa@gmail.com>.

2. Professora da Escola de Ciência e Tecnologia da UFRN, Doutorado em Economia pela Toulouse School of Economics (TSE), Mestrado em Economia na EPGE-FGV. *E-mail:* <raquelmsampaio@gmail.com>.

3. Professor do Departamento e Pós-graduação em Administração da UFRN, Pós-doutorado em Economia na University of Illinois e Doutorado em Economia no Pimes/UFPE, com estágio doutoral na Université Paris 1-Sorbonne. *E-mail:* <lucianombsampaio@gmail.com>.

1 INTRODUÇÃO

Apesar da ampliação, a partir da década de 1990, do acesso à educação no Brasil, os dados da última avaliação do Programa Internacional de Avaliação de Alunos (Pisa, na sigla em inglês) mostram que os alunos brasileiros na faixa dos 15 anos têm obtido resultados entre os piores do mundo em leitura, matemática e ciência. O Pisa é realizado a cada três anos por países membros ou parceiros da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) – foram 65 países em 2009, com o Brasil ocupando a 53ª posição, atrás de Chile, Colômbia, México, entre outros. Mesmo com um desempenho absoluto um pouco melhor em relação ao exame de 2006, a comparação mostra que a educação brasileira evoluiu bem menos do que a de outros países em desenvolvimento. Esse fato é preocupante dada a importância indiscutível da educação para o desenvolvimento econômico e social de um país.

A literatura procura identificar os fatores que influenciam a qualidade do aprendizado com o propósito de fornecer subsídios à formulação de políticas educacionais. Entre eles, citam-se os fatores extraescolares, como as características familiares e a aptidão do aluno, entre outros, e os fatores intraescolares, os quais estão mais sujeitos às ações de políticas educacionais. Neste último grupo, a qualidade da educação muitas vezes é atribuída ao trabalho do professor, desenvolvido dentro da sala de aula, e ao esforço do aluno. Contudo, uma visão mais ampla da instituição escolar demonstra a importância da sua organização, protagonizada pelo seu diretor e sua capacidade de liderança (Luck, 2011; Tekeste, 2000; Normore, 2004; Meyer; Fleistritz, 2003).

Alguns estudos demonstram efeitos pequenos, contudo significantes, de variáveis ligadas ao diretor e sua liderança sobre o desempenho escolar dos alunos. Sebastian e Allensworth (2012) e Bryk *et al.* (1999), usando equações estruturais multiníveis e um modelo hierárquico linear para escolas públicas de Chicago, mostraram que a liderança do diretor, representada pelo desenvolvimento de metas, pelo relacionamento com os professores e pelo fortalecimento da relação entre escola, pais e comunidade, afeta positivamente o desempenho dos alunos. Essas variáveis representativas da liderança do diretor explicam de três a cinco por cento da variação na aprendizagem escolar. Apesar de responderem por um percentual pequeno, essa variação representa aproximadamente um quarto da variação total que é explicada por todas as variáveis intraescolares (Hallinger; Heck, 1996a, 1996b, 1998; Creemers; Reezigt, 1996; Hill, 1998).

Diante do papel que o diretor exerce dentro da unidade escolar “como agente de gestão de processos socioeducacionais dinâmicos e participativos” (Luck, 2011, p.31), e do impacto de suas características, sobretudo aquelas representativas da liderança, no aprendizado dos alunos, emerge a discussão sobre a importância da forma de seleção de tal profissional.

No Brasil, a gestão democrática no ensino foi promulgada por meio da Lei de Diretrizes e Bases da Educação Nacional (LDB) – Lei nº 9.393/1996, que em seu Artigo 3º, VIII, enuncia que a gestão democrática do ensino público será disposta “na forma desta lei e da legislação dos sistemas de ensino”, e, em seu Artigo 14, quando evidencia que “os sistemas de ensino definirão as normas da gestão democrática do ensino público na educação básica” (Brasil, 1996). Fica assim regulamentado o princípio da “gestão democrática de ensino, na forma da lei” no Artigo 206, VI, da Constituição Federal de 1988 (Brasil, 1988).

Apesar de tal regulamentação institucionalizar a gestão democrática, como não foi estabelecido um prazo para sua adoção, a aplicação dessa regulamentação deu-se de forma diferenciada entre os sistemas educacionais de estados e municípios do país. Atualmente, em vários desses sistemas ocorrem eleições de diretores de escolas enquanto em outros, esses diretores continuam sendo indicados pelos governos.

Entre os processos de seleção de diretores nos diversos sistemas estaduais e municipais de ensino, encontram-se: o provimento do cargo de diretor por meio de concurso público, seleção e eleição, apenas eleição e, em algumas situações, indicações políticas e/ou técnicas.

Discussões sociais apontam que a eleição de diretor apresenta-se como uma solução ótima e equivalente a uma sociedade dita democrática (Paro, 1996), eliminando o clientelismo político dentro das instituições de ensino, o qual se baseia na confiança pessoal e política dos que indicam, e não na capacidade própria dos diretores, “ficando distante da ordenação impessoal que caracteriza a administração burocrática” (Mendonça, 2001, p. 89).

Souza (2008) ressalta que em redes de ensino em que os diretores foram eleitos existe uma participação maior da comunidade dentro da escola, com um melhor funcionamento de organismos coletivos, em contraponto às redes de escolas nas quais os diretores foram indicados, em que a ação coletiva nos processos de gestão escolar é menos presente.

Por outro lado, Paro (1996) e Myers (2008) colocam que a escolha de diretores via processo eleitoral pode levar ao corporativismo já que, em grande parte das vezes, o voto dos professores tem um maior peso que o de pais e funcionários nos resultados da eleição. Outro ponto de discussão no que se refere a efeitos perversos da eleição é a presença de poucos ou de nenhum candidato à eleição na escola (Luck, 2011).

Alguns estudos nacionais já analisaram o impacto da gestão democrática sobre o aprendizado escolar, mesmo não sendo este o seu interesse principal. Por exemplo, Barros e Mendonça (1997) abordaram o impacto de inovações na gestão escolar – autonomia financeira, eleição de diretor e colegiado ou conselho escolar – sobre o desempenho educacional; Menezes-Filho (2007), por sua vez, relacionou o desempenho

nos ensinos fundamental e médio a diversas variáveis extraescolares e interescolares, entre elas, a seleção de diretor; Biondi e Felício (2007) consideraram variáveis de *background* familiar, características da escola e do diretor da escola e o procedimento pelo qual o diretor assumiu a escola sobre o desempenho dos alunos brasileiros do ensino fundamental; Alves (2008) utilizou variáveis socioeconômicas dos alunos, professores, volume de recursos recebidos e a forma de seleção do diretor escolar para investigar a associação entre as políticas públicas e o desempenho das redes de ensino das capitais brasileiras.

Esses estudos, de modo geral, apontaram um efeito significativo e positivo sobre o desempenho dos alunos de diretores escolhidos por seleção e eleição. Entre os métodos utilizados, citam-se o de regressão múltipla, regressão com dados em painel e modelo multinível. Como o objetivo deles não era, especificamente, a avaliação da eleição do diretor sobre o rendimento dos alunos, os métodos usados em seus artigos podem não captar apropriadamente o efeito desta mudança política, pois não foram delimitadas estratégias claras de identificação do efeito causal da eleição dos diretores no desempenho dos alunos.

Saes e Alves (2004) e mais recentemente Luck (2011) apontam uma carência de estudos que demonstram a efetividade das políticas de seleção de diretores, independentemente da forma de provimento para o cargo adotado. Parte desta lacuna deve ser atribuída à ausência de dados que permitam a comparação de desempenho de escolas com e sem eleição de diretor.

Em 2008, porém, foi iniciada a gestão democrática no estado da Bahia, por via da seleção de diretores escolares com processo de eleição entre os candidatos aprovados anteriormente no Curso de Gestão Escolar. A gestão democrática foi promessa de campanha do governador eleito em 2006. Em 2007, o governador manteve a indicação política dos diretores e vice-diretores, enquanto um projeto de lei não era aprovado. As primeiras eleições aconteceram em 2008.

Para se candidatar ao cargo, são necessários os seguintes pré-requisitos: *i*) possuir formação superior em grau de licenciatura plena, tendo desenvolvido atividade docente durante no mínimo dois anos; *ii*) ter disponibilidade para atendimento à demanda de carga horária de quarenta horas semanais; e *iii*) estar em exercício, há no mínimo seis meses, na unidade escolar para a qual concorre ao cargo. Na inexistência de candidato que atenda aos requisitos previstos ou de não haver candidaturas, o secretário estadual da Educação indica um representante do segmento do magistério da unidade escolar para assumir a direção (Bahia, 2008).

A eleição ocorre por voto direto nas escolas da rede estadual do ensino que ofertem séries a partir do 5º ano, participando alunos, pais de alunos, professores e servidores das escolas. As categorias de eleitores têm os seguintes pesos: *i*) 25% para pais e responsáveis; *ii*) 25% para estudantes; *iii*) 45% para os membros do magistério;

e *iv*) 5% para os servidores. Após eleito, o diretor da escola tem um mandato de três anos, podendo ser reconduzido mais uma vez ao cargo (Bahia, 2008).

São atribuições do diretor: *i*) elaborar o planejamento pedagógico; *ii*) manter um fluxo de informação atualizado com a Secretaria da Educação; *iii*) coletar, analisar e divulgar os resultados de desempenho dos alunos; *iv*) assegurar a participação do colegiado escolar; *v*) gerenciar o funcionamento e a administração da escola; *vi*) coordenar atividades financeiras; *vii*) elaborar e responder pela prestação de contas dos recursos; *viii*) promover ações que estimulem a utilização de espaços físicos; *ix*) controlar a frequência dos servidores; entre outras. (BAHIA, 2002).

Não existe nenhum mecanismo efetivo de fiscalização ou punição aos diretores que não cumpram suas atribuições adequadamente. Apesar da Lei nº 11.043, de 9 de maio de 2008, que dispõe sobre a existência do colegiado escolar – formado por representantes dos diversos segmentos da comunidade escolar, entre eles, a direção, professores ou coordenadores pedagógicos, servidores técnico-administrativos, estudantes, pais ou responsáveis e comunidade local – atribuir a este colegiado o controle, o acompanhamento e a avaliação dos processos administrativos e pedagógicos, a legislação não detalha como estes são feitos e tampouco se há punição aos responsáveis.

Como esse processo eleitoral é restrito às escolas estaduais do estado, é possível utilizar as escolas municipais como grupo de controle e avaliar as diferenças de desempenho nas escolas das duas redes.

Assim, o objetivo deste estudo é analisar os efeitos de uma política de seleção de diretor sobre o rendimento escolar, comparando as escolas estaduais – nas quais houve eleição – às municipais, do estado da Bahia, por meio das metodologias de dados em painel e de diferença em diferenças. São utilizadas bases de dados anteriores (2007) e posteriores (2009 e 2011) à data de implementação da política a ser testada.

Além desta introdução, este artigo contém outras seis seções: *i*) na seção seguinte, apresenta-se uma discussão da literatura sobre desempenho educacional no Brasil e sobre gestão escolar democrática no Brasil e no estado da Bahia; *ii*) na terceira seção, as metodologias de diferença em diferenças e dados em painel são dispostas; *iii*) na quarta seção, as bases de dados são descritas; *iv*) na quinta seção, são mostrados os resultados; e, *v*) na sexta seção, as conclusões.

2 FATORES DETERMINANTES DO DESEMPENHO EDUCACIONAL

Em 1966, foi desenvolvida uma grande pesquisa sobre a educação, solicitada pelo congresso americano, com o objetivo de estudar em larga escala o desempenho de alunos de diferentes raças nas escolas dos Estados Unidos. Até então, os americanos acreditavam que as intervenções políticas – por exemplo, o financiamento do

governo federal – conseguiram corrigir as desigualdades sociais entre as escolas com predominância de alunos negros ou brancos. O relatório concluiu que as diferenças de recursos destinados às escolas não apresentavam relevância frente à desigualdade no aprendizado dos alunos de diferentes raças e que os principais fatores influenciadores do desempenho dos alunos eram o *background* familiar e o *status* socioeconômico (Lee, 2001). A publicação desses resultados gerou uma ideia predominante de que apenas as famílias e os colegas de classe podiam afetar o desempenho escolar, gerando forte pessimismo pedagógico frente à conclusão de que a escola teria pouco impacto no desempenho dos alunos e que diferentes maneiras de organizá-las seriam igualmente eficientes (Soares, 2004).

Resguardando a constatação de que uma parte importante do desempenho discente é explicada pelos fatores extraescolares, ainda assim existe uma variação entre resultados de escolas de um mesmo sistema que atendem a alunos com perfis socioeconômicos similares, implicando que características da unidade escolar que o aluno frequenta podem explicar essa diferença de desempenho (Brooke; Soares, 2008 *apud* Andrade; Soares, 2008). Como demonstrado no estudo de Rivkin, Hanushek e Kain (2005), ao controlar para fatores extraescolares, ainda foi identificado um importante efeito das características das escolas sobre o aprendizado dos alunos.

Nessa linha da literatura internacional e a partir dos primeiros resultados de exames de proficiência, de 1995, vários estudos brasileiros procuraram identificar a contribuição de características observáveis externas e internas à escola, como da família, dos próprios alunos, dos professores, de diretores e das escolas no desempenho dos alunos.

Para avaliar o papel desenvolvido por fatores externos à escola, sobretudo o ambiente familiar e comunitário do docente, e ainda considerando características individuais e familiares, destacam-se os trabalhos desenvolvidos por Barros e Mendonça (1997), Barros *et al.* (2001) e Bonelli (2002).

Outro grupo de estudos, alguns também usando fatores externos como controle, focou a identificação dos fatores intraescolares que exercem impacto sobre o aprendizado, entre eles, Barros e Mendonça (1997), Lee, Bryk e Smith (1993), Franco *et al.* (2003), Satyro e Soares (2007), Klein (2006), Araújo e Luzio (2005), Menezes Filho (2007), Soares e Teixeira (2006), Tavares (2012), Curi e Menezes Filho (2013). Os resultados encontrados em alguns desses estudos são pouco robustos e o efeito das variáveis estudadas sobre o desempenho do discente varia em significância e sentido, não havendo assim um consenso.

Algumas dessas referências incluem como variáveis explicativas a forma do provimento do cargo de direção escolar, computando seus efeitos sobre o desempenho médio dos alunos. Barros e Mendonça (1997) abordaram o impacto de inovações na gestão escolar – autonomia financeira, eleição de diretor e presença

de colegiado ou conselho escolar – sobre o desempenho educacional, comparando o desempenho dos sistemas estaduais mais e menos inovadores. Para tanto, fizeram uso do modelo de regressão múltipla e do modelo de efeitos fixos. Em todos os modelos, utilizaram como variáveis dependentes a taxa de reprovação dos alunos (obtida com base no censo escolar), a proporção de crianças fora da escola, duas medidas de atraso escolar (obtidas com base na Pesquisa Nacional por Amostra de Domicílios – PNAD) e, ainda, indicadores de rendimento escolar, obtidos a partir do Sistema Nacional de Acompanhamento da Educação Básica (Saeb). Como variáveis de controle, consideraram características da família (renda e escolaridade da família) e dos professores (salários e escolaridade). A base de dados cobre o período de 1981 a 1993 e todas as unidades da federação, exceto o Distrito Federal, os estados da região Norte e o estado de Alagoas. Seus resultados mostram que a eleição aumenta o desempenho educacional em um ponto percentual.

Utilizando uma estratégia de estimação semelhante ao trabalho anterior, Biondi e Felício (2007), com um painel de dados do Saeb e do censo escolar para os anos de 1999, 2001 e 2003, aplicaram os métodos de mínimos quadrados ordinários e de efeitos fixos para identificar os atributos escolares que possam ser alvos de políticas educacionais voltadas à melhoria do desempenho escolar. Eles utilizaram variáveis de *background* familiar, das escolas e dos alunos, além de variáveis representantes de características do diretor da escola – experiência do diretor e procedimento pelo qual o diretor assumiu a escola. Entre as variáveis com efeitos significativos, encontraram a ausência de rotatividade dos professores ao longo do ano letivo, a experiência dos professores (mais de dois anos em sala de aula) e a existência na escola de conexão com a internet, todas afetando positivamente o desempenho médio. O procedimento de escolha do diretor por concurso público, eleição ou indicação (de técnicos, políticos ou outras formas) tem efeito negativo sobre o desempenho médio dos alunos relativamente à categoria usada como base de comparação que foi a seleção, seguida de eleição de diretores.

Menezes-Filho (2007), por sua vez, utilizando dados do Sistema de Avaliação da Educação Básica (Saeb) e um modelo de regressão múltipla relacionou o desempenho dos alunos das 4ª e 8ª séries do ensino fundamental e da 3ª série do ensino médio nos testes de proficiência em matemática, em escolas públicas e particulares, às variáveis relacionadas ao aluno, à escola, ao professor, ao diretor e ao modo de escolha do diretor – dos quais inclui a indicação técnica ou política e ou a seleção. O estudo não identifica relação significativa entre o desempenho e a forma de escolha do diretor nas escolas públicas. Com relação às escolas privadas, existe um impacto significativo e positivo, mas apenas para os alunos da 8ª série.

Alves (2008), por meio de modelos multiníveis de classificação cruzada, buscou associar o efeito de políticas públicas ao desempenho das redes de ensino

das capitais brasileiras, mais especificamente, dos alunos de 4ª série do ensino fundamental nos anos de 1999, 2001 e 2003, usando como medida de desempenho a média em matemática obtida no sistema de avaliação da educação básica. A autora utilizou como variáveis explicativas: o nível socioeconômico dos alunos, a dependência administrativa, o ensino fundamental com duração de nove anos, o sistema de avaliação da rede de ensino, o número de professores com ensino superior, a participação no Programa Dinheiro Direto na Escola (PDDE), o atendimento em educação infantil, e a forma de provimento do cargo de direção escolar. Os resultados mostraram que as variáveis educacionais de maior impacto no desempenho dos estudantes foram aquelas relacionadas aos processos de escolha por mérito de diretores, a participação no PDDE, a implementação de sistemas de avaliação, o atendimento em educação infantil e a formação superior de docentes. Entre os processos envolvendo a forma de escolha do diretor, apenas a seleção e a eleição tiveram efeito positivo nas redes de ensino que a adotaram.

3 METODOLOGIA

No intuito de avaliar o efeito do provimento do cargo de direção escolar por meio de eleição no estado da Bahia sobre o desempenho cognitivo dos alunos, foram utilizados dois métodos analíticos: diferença em diferenças (*diff-in-diff*) e dados em painel. Estas metodologias são descritas a seguir.

3.1 Diferença em Diferenças

No modelo de diferença em diferenças, as informações sobre os grupos de tratamento e controle são analisadas antes e após a implementação do processo de eleição dos diretores escolares. A estratégia de identificação do método assume que a *variação* média ao longo do tempo de desempenho dos alunos não explicada pelas variáveis de controle do modelo seria igual entre a rede estadual e municipal se não houvesse a diferença nos sistemas de escolha do diretor escolar.

Considerando que a designação das escolas para o grupo de tratamento (escolas estaduais que passaram a adotar a eleição de diretores) e de controle (escolas municipais que não selecionam e não elegem os diretores) não ocorre de forma aleatória, o uso do método de diferença em diferenças (*diff-in-diff*), com dados para os dois grupos para antes e depois da política de eleição adotada, permite identificar o impacto da eleição de diretores sobre medidas de desempenho acadêmico, evitando o problema de viés de seleção amostral.

Os indicadores de resultado utilizados são o desempenho médio em língua portuguesa e matemática das escolas baianas. Seja Y a taxa de desempenho médio da escola em uma dessas disciplinas, os sobrescritos T e C denominam, respectivamente, se a escola possui eleição de diretor (grupo de tratamento)

ou não (grupo de controle). Dado que os grupos podem ser observados antes (2007) e após a implementação da política de eleição de diretores (em 2009 e 2011), isso permite que sejam calculadas as variações no desempenho médio das escolas dos grupos de tratamento e controle entre os períodos $t = 0$ (antes da intervenção) e $t = 1$ (depois da intervenção):

$$\Delta Y^T = Y_{t=1}^T - Y_{t=0}^T. \quad (1)$$

$$\Delta Y^C = Y_{t=1}^C - Y_{t=0}^C. \quad (2)$$

Dessa forma, o método de diferença em diferenças assume que caso a política não tivesse sido implementada, as diferenças no tempo entre desempenho dos alunos das escolas do grupo de controle e de tratamento seriam iguais. Portanto, toda diferença observada entre a variação de desempenho do grupo de tratamento, ΔY^T , e a variação de desempenho do grupo de controle, ΔY^C , pode ser atribuída à mudança na política de seleção dos diretores escolares.

É possível estimar o efeito de um tratamento usando o método de diferença em diferenças por meio da seguinte regressão linear⁴:

$$Y_{it} = \alpha + \beta_1 política_{it} + \beta_2 ano_t + \beta_3 política * ano_t + A_{it}\beta_4 + D_{it}\beta_5 + P_{it}\beta_6 + E_{it}\beta_7 + \mu. \quad (3)$$

em que a variável *política* indica se a escola *i* adotou a eleição de diretores; a variável *ano* assume valor 0 quando $t = 0$ (antes da implementação da política) e valor 1 para $t = 1$ (depois da implementação da política); e a variável *política ano_t* é uma interação das variáveis anteriores que assume valor 1 apenas para as escolas do grupo de tratamento no período posterior à implementação da política. Nessa regressão, pode-se mostrar que o coeficiente β_3 , aquele associado à interação, é exatamente o impacto da eleição de diretores sobre a medida de desempenho escolar *Y*.

As demais variáveis explicativas do modelo incluídas como controles são: *i*) um vetor de características dos alunos na escola *i* no tempo *t* – *A_{it}* – que inclui gênero, cor/raça, escolaridade da mãe, presença de mais de uma televisão em casa e presença de computador em casa; *ii*) um vetor de características do diretor da escola *i* no tempo *t* – *D_{it}* –, que informa se esse diretor tem pós-graduação e se tem experiência em gestão; e *iii*) um vetor de características da escola *i* no tempo *t* – *E_{it}* – com *dummies* para captar presença de laboratório de informática e biblioteca. Ainda foram inseridas variáveis de controle para os professores – *P_{it}* –

4. Para uma demonstração desta relação entre o modelo *diff-in-diff* e regressão linear, ver, por exemplo, Wooldridge (2010).

referentes a sua experiência, formação em ensino superior e treinamentos. Essas variáveis são, em sua maioria, frequentemente utilizadas em estudos que tratam dos determinantes do desempenho educacional, como em Barros *et al.* (2001), Menezes Filho (2007), Aquino e Kassof (2011), Oshiro e Scorzafave (2011), Cadaval e Monteiro (2011), entre outros.

3.2 Dados em painel

O modelo de diferença em diferenças assume que escolas do mesmo grupo têm o mesmo desempenho médio, todas as demais variáveis constantes. Ou seja, assumimos que o intercepto α da equação (3) é o mesmo para todas as escolas. No entanto, podem existir características não observáveis das escolas que estão correlacionadas tanto com a adoção pela escola da política de seleção de diretores quanto pelo aprendizado dos alunos.

Nesta seção, adotamos uma metodologia mais geral que permite que cada escola tenha um desempenho médio ao longo do tempo, diferente das demais escolas. Foram utilizados os modelos de dados em painel com efeitos fixos e com efeitos aleatórios. Assim, é possível estimar o impacto da eleição de diretores no desempenho médio dos alunos levando em conta todas as características das escolas que não variam ao longo do tempo.

Adotou-se o seguinte modelo econométrico para o desempenho médio dos alunos da escola i no período t , Y_{it} :

$$Y_{it} = \beta_0 + \beta_1 ano_t + \beta_2 politica_{it} = A_{it}\beta_3 + D_{it}\beta_4 + E_{it}\beta_5 + P_{it}\beta_6 + \alpha_i + \mu, \quad (4)$$

em que as variáveis dependentes, nessa abordagem, continuam sendo as médias de proficiência em língua portuguesa e matemática da unidade escolar. A variável *politica_{it}* capta se a escola i adota a eleição de diretores na data t , a variável *ano* assume valor '0' quando $t = 0$ (antes da implementação da política) e valor 1 para $t = 1$ (depois da implementação da política). Os demais regressores são os mesmos utilizados no modelo de diferença em diferenças, ou seja, controla-se para características dos alunos, da infraestrutura da escola, dos professores e dos diretores das escolas.

A variável a_i capta os fatores não observados para a escola i , constantes no tempo, que afetam a variável dependente Y_{it} . No modelo de efeitos aleatórios são exigidas as hipóteses de exogeneidade estrita e de ortogonalidade entre as variáveis não observadas e o termo de erro. O modelo de efeitos fixos, por sua vez, permite que as variáveis observáveis possam estar correlacionadas com o erro e com as variáveis não observáveis (Aquino; Kassof, 2011).

A seguir descrevem-se as bases de dados utilizadas para os períodos utilizados, bem como estatísticas descritivas das variáveis do modelo.

4 DESCRIÇÃO DOS DADOS

Os dados utilizados são oriundos do Sistema de Avaliação da Educação Básica. Deste sistema, fazem parte a Avaliação Nacional do Rendimento Escolar (Anresc) – também denominada Prova Brasil – e o Sistema Nacional de Avaliação da Educação Básica (Saeb). A Prova Brasil é constituída por provas de língua portuguesa e matemática, aplicadas censitariamente aos alunos de 5º e 9º anos do ensino fundamental público, nas redes estaduais, municipais e federais, das áreas rural e urbana, em escolas que tenham no mínimo vinte alunos matriculados nas séries avaliadas. Além dos dados de desempenho, são respondidos questionários socioeconômicos por alunos, professores e diretores, possibilitando um diagnóstico do sistema educacional.

Foram utilizados os dados das escolas estaduais e municipais da Bahia que ofertam o 9º ano do ensino fundamental e que participaram da Prova Brasil nos anos de 2007, 2009 e 2011. A escolha pelas escolas do 9º ano do ensino fundamental deve-se ao fato de que a política pública é aplicada apenas às escolas que ofertam séries a partir do 5º ano, contemplando, assim, o 9º ano. As escolas estaduais constituíram o grupo de tratamento, uma vez que, depois do ano de 2008, elas passaram a adotar a política de eleição, enquanto as escolas dos municípios que não adotaram eleições em nenhum dos três anos formaram o grupo de controle.

Segundo a Lei de Diretrizes e Bases da Educação Nacional (LDB) de 1996, “os municípios são responsáveis por oferecer a educação infantil em creches e pré-escolas e, com prioridade, o ensino fundamental”. Por sua vez, os estados asseguram o ensino fundamental e oferecem, com prioridade, o ensino médio. Contudo, a LDB não estabeleceu um prazo para que escolas do ensino fundamental, que até então eram geridas pelos Estados, fossem transferidas para a gestão municipal. Na Bahia, esse processo de municipalização ainda não se consolidou: entre as escolas com turmas de 9º ano com mais de 20 alunos, ao longo do período de 2007 a 2011, apenas 24 foram municipalizadas. Além disso, há indícios de que o estado da Bahia continuou investindo na estrutura das escolas de nível fundamental de 9º ano, pois houve um aumento na proporção de escolas com bibliotecas e com laboratórios de informática, assim como na expansão de capacitação dos professores.

A vantagem de se comparar redes de ensino de um mesmo estado é uma homogeneização no perfil dos alunos (características socioeconômicas) e cultural. O uso de escolas de redes estaduais de outros estados como grupo de controle exigiria que não houvesse ocorrido eleições nestes estados após 2007. Nessa situação, encontravam-se apenas o Maranhão e Tocantins – estados com grandes diferenças culturais e socioeconômicas se comparadas à Bahia.

Uma preocupação corrente na definição dos grupos de controle e tratamento é a contaminação do grupo de tratamento, neste caso, por uma possível migração de professores da rede estadual para a municipal ou vice-versa. A legislação exige

que o professor esteja em "efetivo exercício na unidade escolar correspondente há, no mínimo, seis meses".

Apesar do tempo entre o anúncio da política e a sua implementação (dois anos), é pouco provável que um professor migre de rede apenas devido ao processo de gestão democrática, pois o ingresso na rede estadual é feito por meio de concurso público. Ressalta-se que ainda em 2010 foram convocados professores aprovados no concurso da rede estadual da Bahia, cujo edital foi publicado em 2005 e, portanto, antes da eleição do governador que adotou a política. Outro fato é que mesmo que um professor tenha sido aprovado e convocado pelo concurso, a nomeação para as escolas seria feita de acordo com a ordem de classificação do aprovado e com a necessidade da secretaria de educação da Bahia, ou seja, um candidato tem pouco controle com relação a qual escola ele será indicado.

A seguir, são descritas as bases de dados utilizadas nos modelos de dados em painel e diferença em diferenças. Esses modelos foram estimados para duas bases de dados diferentes: uma com dados de escolas que participaram da Prova Brasil em 2007 e 2009 e outra para escolas participantes em 2007 e 2011.

4.1 Base de dados para escolas de 2007 e 2009

Em 2007 e 2009, participaram da Prova Brasil 733 escolas estaduais e 686 escolas municipais da Bahia. Do grupo de tratamento, foram excluídas: *i*) 119 escolas pela ausência de dados nas respostas dos questionários referentes à proficiência dos alunos, modo de posse do diretor, experiência, pós-graduação do diretor e existência de laboratório de informática e biblioteca na escola; *ii*) 19 escolas estaduais que já tinham algum processo de eleição antes de 2008; e *iii*) outras 206 escolas que não adotaram a política de eleição. No grupo de controle, também foram excluídas 170 escolas com ausência de dados nas respostas dos questionários.

Como existe a possibilidade de redes municipais adotarem a política de eleição de diretores, optou-se por excluir todas as escolas – estaduais e municipais – de municípios com eleições. Para tanto, em vez de buscar legislações para todos os 417 municípios da Bahia, os municípios com mais de 50% de diretores eleitos foram identificados. Com esse procedimento, saíram da amostra: 187 escolas estaduais e 96 escolas municipais.

Também foram excluídas das escolas cujos diretores afirmaram terem sido eleitos, mas que estavam em municípios onde a maioria das escolas indicou não haver eleição para diretor. Por fim, foram excluídas dos dois grupos, cinco escolas que em 2007 eram de dependência administrativa do estado e que em 2009 pertenciam à rede municipal. Assim, compõe a amostra 197 escolas estaduais (10.800 alunos) e quatrocentas escolas municipais (26.857 alunos).

Na análise, é importante que as amostras finais sejam representativas das amostras completas da Prova Brasil. A tabela 1 mostra que as médias das proficiências, tanto em língua portuguesa quanto em matemática, são bem similares. A maior diferença encontrada entre os resultados da Prova Brasil e das amostras deste trabalho, para os anos de 2007 e 2009, foi de 2,52 pontos na média de proficiência de língua portuguesa das escolas estaduais, para o ano de 2007.

TABELA 1
Média e desvio-padrão da proficiência das escolas estaduais e municipais na Prova Brasil e nas amostras para os anos de 2007 e 2009

	Anos	Língua portuguesa				Diferença	Matemática				Diferença
		Prova Brasil		Amostra			Prova Brasil		Amostra		
		Média	DP	Média	DP		Média	DP	Média	DP	
Escolas estaduais	2007	218,22	16,20	215,7	15,83	2,52	226,87	14,92	224,65	14,88	2,22
	2009	224,12	17,20	223,16	17,58	0,96	226,51	15,54	225,07	15,71	1,44
Escolas municipais	2007	214,34	14,47	213,74	13,98	0,6	224,46	14,47	224,2	13,55	0,26
	2009	223,25	16,88	222,35	16,51	0,9	225,79	14,88	225,81	14,35	-0,02

Fonte: Elaboração própria, com base nos dados do Inep.

A seguir são apresentadas as estatísticas descritivas das escolas estaduais (grupo de tratamento) e municipais (grupo de controle), que compõem as amostras para os períodos 2007 e 2009.

TABELA 2
Estatísticas descritivas para os grupos de tratamento e controle, anos 2007 e 2009

Variáveis	Grupo de tratamento			Grupo de controle		
	2007	2009	Diferença	2007	2009	Diferença
Média em língua portuguesa	215,70 (15,83)	223,16 (17,58)	7,46	213,74 (13,97)	222,35 (16,51)	8,61
Média em matemática	224,65 (14,88)	225,07 (15,71)	0,42	224,20 (13,55)	225,81 (14,35)	1,61
Gênero masculino*	0,42 (0,10)	0,42 (0,090)	0,00	0,40 (0,10)	0,41 (0,09)	0,01
Cor branca*	0,18 (0,10)	0,16 (0,09)	-0,01	0,19 (0,11)	0,17 (0,11)	-0,02
Possui mais de uma TV*	0,33 (0,13)	0,37 (0,12)	0,04	0,22 (0,11)	0,27 (0,11)	0,05
Possui computador*	0,11 (0,10)	0,28 (0,15)	0,17	0,06 (0,07)	0,18 (0,13)	0,12
Mãe tem o 5º ano incompleto*	0,18 (0,10)	0,16 (0,08)	-0,02	0,25 (0,10)	0,22 (0,10)	-0,04
Mãe tem ensino fundamental incompleto*	0,30 (0,09)	0,31 (0,10)	0,01	0,34 (0,09)	0,35 (0,10)	0,02

(Continua)

(Continuação)

Variáveis	Grupo de tratamento			Grupo de controle		
	2007	2009	Diferença	2007	2009	Diferença
Mãe tem ensino médio incompleto*	0,13 (0,07)	0,14 (0,06)	0,01	0,11 (0,06)	0,11 (0,06)	0,00
Mãe tem ensino médio completo*	0,20 (0,10)	0,22 (0,10)	0,01	0,14 (0,08)	0,15 (0,09)	0,01
Mãe tem ensino superior completo*	0,06 (0,06)	0,07 (0,05)	0,01	0,04 (0,04)	0,05 (0,05)	0,01
Diretor com atualização**	0,04 (0,20)	0,05 (0,22)	0,01	0,03 (0,16)	0,02 (0,13)	-0,01
Diretor com especialização**	0,54 (0,50)	0,73 (0,45)	0,19	0,26 (0,44)	0,41 (0,49)	0,15
Diretor com mestrado**	0,03 (0,16)	0,01 (0,07)	-0,02	0,05 (0,00)	0,01 (0,07)	-0,05
Diretor com doutorado**	- -	- -		0,01 (0,07)	0,00 (0,05)	-0,01
Diretor com 2 a 5 anos de experiência**	0,09 (0,28)	0,36 (0,48)	0,27	0,38 (0,49)	0,18 (0,39)	-0,20
Diretor com mais de 5 anos de experiência**	0,30 (0,46)	0,26 (0,44)	-0,04	0,31 (0,46)	0,34 (0,48)	0,03
Computador**	0,63 (0,48)	0,88 (0,32)	0,25	0,38 (0,49)	0,67 (0,47)	0,28
Biblioteca**	0,59 (0,49)	0,71 (0,46)	0,11	0,67 (0,47)	0,76 (0,43)	0,09
Professor com ensino superior***	0,82 (0,29)	0,87 (0,23)	0,05	0,71 (0,34)	0,83 (0,27)	0,12
Professor com capacitação***	0,72 (0,33)	0,74 (0,30)	0,02	0,78 (0,29)	0,81 (0,260)	0,03
Professor com mais de 2 anos de experiência***	0,91 (0,19)	0,88 (0,21)	-0,03	0,94 (0,14)	0,94 (0,13)	0
Observações	197	197		400	400	

Fonte: Elaboração própria, com base nos dados do Inep.

Valores dos desvios-padrão em parênteses. *Valores das variáveis em proporção. ** Variáveis *dummies*. ***Para as variáveis de professor o número de observações difere, sendo 184 escolas estaduais e 380 escolas municipais.

Com relação às variáveis dependentes, pode-se observar um aumento das médias escolares, de 2007 para 2009, nas proficiências em língua portuguesa nos dois grupos. Na proficiência em matemática, a variação no grupo de tratamento é pequena, e no grupo de controle de pouco mais de um ponto. Os dois grupos apresentam proficiências com valores próximos.

As proporções de alunos do gênero masculino e de cor branca permanecem praticamente as mesmas dentro dos dois grupos estudados. Entre as características socioeconômicas dos alunos, observa-se que a proporção de alunos com mais de uma televisão não apresenta variação. O grupo de tratamento apresenta uma maior proporção de alunos com computador em casa, e está registrado um aumento na proporção de alunos que passam a ter computador, de 2007 para 2009,

dentro dos grupos de tratamento e de controle. Os dois grupos não apresentam variações quanto ao nível educacional das mães, tendo ambos os grupos uma proporção maior de mães que não completaram o ensino fundamental.

Os dados mostram que o grupo de tratamento conta com um maior número de diretores escolares que possui curso de especialização. Por sua vez, com relação à experiência do diretor, os grupos apresentam percentuais próximos de diretores com mais de cinco anos de experiência como gestores escolares, entretanto, para 2009, o grupo de tratamento tem mais diretores com menos de cinco anos de experiência.

Ambos os grupos têm um aumento no número de escolas com laboratório de informática, entre 2007 e 2009; o grupo de tratamento apresenta um percentual maior de escolas com laboratórios; e os dois contam com proporções similares de escolas com biblioteca.

Os dois grupos apresentam proporções próximas de professores com ensino superior, que tenham passado por alguma capacitação e possuam mais de dois anos de experiência. E, para os dois grupos, não existem variações ao longo do período de tempo do painel.

4.2 Base de dados para escolas de 2007 e 2011

Comparando os dados para os anos de 2007 e 2011, observa-se que participaram da Prova Brasil 641 escolas estaduais e 681 escolas municipais. Assim como para os dados do painel anterior (2007 e 2009), do grupo de tratamento, foram excluídas as escolas com ausência de dados nas respostas dos questionários (62 escolas) e as escolas estaduais que já tinham algum processo de eleição antes de 2008 (21 escolas) e não adotaram a política de eleição (212 escolas); ainda, foram excluídas as escolas estaduais localizadas em municípios que possuem política de eleição de diretores (186 escolas). O mesmo procedimento foi adotado para o grupo de controle com a exclusão das escolas com ausência de dados (em 111 escolas) e de escolas de 32 municípios com processo de eleição dos dirigentes escolares, dada a possibilidade de legislação municipal própria sobre processos eleitorais. Do painel, ainda foram excluídas 24 escolas, em cada grupo, que tiveram mudança de dependência administrativa no período de 2007 a 2011. Compuseram a amostra final 136 escolas estaduais (7.790 alunos) e 394 escolas municipais (26.462).

A tabela 3 mostra que as médias das proficiências, tanto em língua portuguesa quanto em matemática, das Provas Brasil e das amostras para 2007 e 2011, são similares. A maior diferença entre a amostra utilizada e os dados da Prova Brasil é de 2,56 pontos para a média na proficiência em língua portuguesa no ano de 2007.

TABELA 3
Média e desvio-padrão da proficiência das escolas estaduais e municipais na Prova Brasil e nas amostras para os anos de 2007 e 2011

	Anos	Língua portuguesa				Diferença	Matemática				Diferença
		Prova Brasil		Painel			Prova Brasil		Painel		
		Média	DP	Média	DP		Média	DP	Média	DP	
Escolas estaduais	2007	218,52	16,26	215,96	17,31	2,56	226,96	14,94	224,97	16,89	1,99
	2011	225,62	16,75	224,69	17,2	0,93	230,25	17,65	229,07	19,06	1,18
Escolas municipais	2007	214,49	14,39	213,09	14,27	1,4	224,62	14,12	223,63	14,31	0,99
	2011	224,66	15,65	223,06	14,59	1,6	230,42	16,65	229,24	16,25	1,18

Fonte: Elaboração própria, com base nos dados do Inep.

De acordo com os dados apresentados na tabela 4, as proficiências em língua portuguesa e matemática aumentam nos dois grupos e nos dois períodos (2007 e 2011), com aumento maior para o grupo de controle. Porém, as médias nos dois grupos são similares. Com relação às características dos alunos, as proporções de alunos do gênero masculino e de cor branca permanecem praticamente as mesmas dentro dos dois grupos estudados. Assim como nas amostras discutidas anteriormente, nas características socioeconômicas dos alunos, a proporção de alunos com mais de uma televisão e computador em casa é maior no grupo de tratamento. Ocorre um aumento na proporção de alunos que passam a ter computador em casa, em 2011, em ambos os grupos, sendo maior para o grupo de tratamento. Os dois grupos apresentam diminuição na proporção de mães que não completaram o ensino fundamental. No grupo de tratamento, a proporção de escolaridade de mães com ensino médio é maior, bem como existe uma redução em 2011 das mães que não completaram esse nível de escolaridade.

A qualificação de diretores com especialização é maior no grupo de tratamento. Quanto à experiência do diretor no cargo de gestão escolar, o grupo de tratamento apresenta um aumento no número de diretores com menos de cinco anos de experiência, de 2007 para 2011.

O aumento do número de escolas com laboratório de informática foi maior para o grupo de controle, entretanto em ambos os anos o grupo de tratamento apresentou mais laboratórios. Quanto à presença de biblioteca, os grupos apresentavam dados semelhantes em 2007, com aumento maior entre as escolas do grupo de tratamento, em 2011.

Novamente, os dois grupos apresentam proporções próximas para as características dos professores quando se compara 2007 a 2011.

TABELA 4
Estatísticas descritivas para os grupos de tratamento e controle, para 2007 e 2011

Variáveis	Grupo de tratamento			Grupo de controle		
	2007	2011	Diferença	2007	2011	Diferença
Média em língua portuguesa	215,96 (17,31)	224,69 (17,20)	8,73	213,09 (14,27)	223,06 (14,59)	9,97
Média em matemática	224,97 (16,89)	229,07 (19,06)	4,10	223,63 (14,31)	229,24 (16,25)	5,61
Gênero masculino*	0,41 (0,11)	0,43 (0,09)	0,02	0,41 (0,10)	0,42 (0,09)	0,01
Cor branca*	0,17 (0,10)	0,17 (0,10)	0,00	0,19 (0,12)	0,18 (0,11)	-0,01
Possui mais de uma TV*	0,32 (0,13)	0,42 (0,12)	0,10	0,22 (0,11)	0,32 (0,12)	0,10
Possui computador*	0,11 (0,11)	0,35 (0,18)	0,25	0,06 (0,07)	0,22 (0,15)	0,16
Mãe tem o 5º ano incompleto*	0,18 (0,10)	0,19 (0,11)	0,01	0,25 (0,10)	0,29 (0,13)	0,04
Mãe tem ensino fundamental incompleto*	0,29 (0,10)	0,17 (0,07)	-0,12	0,33 (0,09)	0,21 (0,07)	-0,12
Mãe tem ensino médio incompleto*	0,30 (0,09)	0,13 (0,06)	-0,17	0,11 (0,07)	0,11 (0,06)	0,00
Mãe tem ensino médio completo*	0,21 (0,11)	0,25 (0,11)	0,03	0,13 (0,08)	0,16 (0,09)	0,03
Mãe tem ensino superior completo*	0,00 (0,00)	0,09 (0,07)	0,09	0,04 (0,04)	0,06 (0,05)	0,02
Diretor com atualização**	0,04 (0,21)	0,03 (0,17)	-0,01	0,03 (0,16)	0,19 (0,00)	0,16
Diretor com especialização**	0,51 (0,50)	0,82 (0,38)	0,31	0,25 (0,43)	0,54 (0,50)	0,30
Diretor com mestrado**	0,02 (0,15)	0,03 (0,17)	0,01	- -	- -	-
Diretor com doutorado**	- -	- -	-	0,01 (0,07)	0,01 (0,07)	0,00
Diretor com 2 a 5 anos de experiência**	0,08 (0,27)	0,66 (0,47)	0,58	0,38 (0,49)	0,49 (0,00)	0,11
Diretor com mais de 5 anos de experiência**	0,31 (0,46)	0,32 (0,47)	0,01	0,31 (0,46)	0,48 (0,00)	0,17
Computador**	0,64 (0,48)	0,96 (0,21)	0,32	0,38 (0,49)	0,80 (0,40)	0,42
Biblioteca**	0,67 (0,67)	0,82 (0,39)	0,15	0,65 (0,48)	0,77 (0,42)	0,11
Professor com ensino superior***	0,82 (0,27)	0,91 (0,17)	0,09	0,68 (0,34)	0,90 (0,21)	0,22
Professor com capacitação***	0,71 (0,34)	0,72 (0,30)	0,01	0,78 (0,30)	0,82 (0,26)	0,04
Professor com mais de 2 anos de experiência***	0,94 (0,20)	0,87 (0,28)	-0,07	0,94 (0,15)	0,96 (0,13)	0,02
Observações	136	136		394	394	

Fonte: Elaboração própria, com base nos dados do Inep.

Valores dos desvios-padrão em parênteses.*Valores das variáveis em proporção. ** Variáveis *dummies*. ***Para as variáveis de professor o número de observações difere, sendo 111 escolas estaduais e 376 escolas municipais.

Uma última amostra, não apresentada aqui, foi construída com as escolas que participaram da Prova Brasil nas três edições realizadas em 2007, 2009 e 2011. Participaram, nos três anos, 588 escolas estaduais e 662 escolas municipais, que ofertam o 9º ano. Depois de excluídas as observações do banco de dados, de forma similar ao discutido anteriormente, a amostra resultou em 139 escolas estaduais e 330 escolas municipais. De forma similar às amostras anteriores, as médias escolares nas proficiências em língua portuguesa e matemática, nos dois grupos e nos três períodos, (2007, 2009 e 2011) são similares e aumentaram, com o grupo de tratamento apresentando médias maiores. Com relação às características dos alunos, dos diretores e das escolas, os dados repetem as constatações anteriores.

5 RESULTADOS

5.1 Diferença em Diferenças

Para analisar o impacto da adoção da política de eleição dos diretores no rendimento escolar dos alunos, estimou-se um modelo de diferença em diferenças cujo grupo de tratamento é composto pelas escolas da rede estadual da Bahia e o de controle pelas escolas da rede municipal do mesmo estado. As tabelas 5 e 6 ilustram o cálculo da diferença, ao longo do tempo, das diferenças das médias de proficiência em português e matemática, respectivamente, captando o efeito da política de eleição de diretores escolares, para o período de 2007 a 2009.

TABELA 5
Cálculo da diferença em diferenças entre os grupos de tratamento e controle para média em língua portuguesa (2007 a 2009)

	Antes do tratamento	Depois do tratamento	Mudança nas proficiências
	Ano de 2007	Ano 2009	
Escolas grupo de tratamento	215,70	223,16	7,45
Escolas grupo de controle	213,74	222,34	8,60
Diferença de média da proficiência em língua portuguesa	1,96	0,81	-1,15

Fonte: Elaboração própria, com base nos dados do Inep.

A primeira coluna da tabela 5, antes do tratamento, demonstra a diferença das proficiências dos grupos de controle e tratamento quando ambos estavam sob a política de posse de diretores por indicação. Assim, em 2007, o grupo de controle apresenta uma média em português mais baixa em 1,96 pontos. Depois, em 2009, quando as escolas do grupo de tratamento já adotaram a eleição do diretor, a diferença da média cai em 0,81 pontos. Na última coluna são apresentadas as mudanças das proficiências dos dois grupos ao longo do período entre 2007 e 2009, ambos com aumento na proficiência dentro do período. Por fim, o valor da diferença em diferenças – o qual representa a diferença, ao longo do tempo, das diferenças da média de língua portuguesa dos dois grupos – foi de 1,15 pontos negativos.

Análoga à tabela 5, a tabela 6 apresenta a diferença, ao longo do tempo, das diferenças das médias de proficiência em matemática nos dois grupos, captando o efeito da política de eleição de diretores escolares. Observa-se que o valor da diferença em diferenças é próximo ao registrado para a proficiência em língua portuguesa, sendo de 1,18 pontos, na mesma direção, com impacto negativo.

TABELA 6
Cálculo da diferença em diferenças entre os grupos de tratamento e controle para média em matemática (2007 a 2009)

	Antes do tratamento	Depois do tratamento	Mudança nas proficiências
	Ano de 2007	Ano 2009	
Escolas grupo de tratamento	224,65	225,07	0,42
Escolas grupo de controle	224,20	225,81	1,60
Diferença de média da proficiência em matemática	0,45	-0,73	-1,18

Fonte: Elaboração própria, com base nos dados do Inep.

Para conhecer se esta diferença é estatisticamente diferente de zero, faz-se a regressão apresentada na metodologia. As estimativas de interesse, obtidas por mínimos quadrados, são apresentadas na tabela 7. Diante do grande número de variáveis incluídas como controle, optou-se por apresentar apenas os resultados referentes à variável de interesse deste estudo.

TABELA 7
Impacto da política de eleição de diretores sobre a proficiência média em língua portuguesa das escolas estaduais da Bahia (2007 a 2009): estimativas por diferença em diferenças

Variáveis	Proficiência em língua portuguesa					
	(1)	(2)	(3)	(4)	(5)	(6)
Política (<i>dummy</i>)	1,962 (1,374)	-2,287* (1,292)	-2,340* (1,313)	-2,467* (1,377)	-2,088 (1,427)	-2,169* (1,270)
Ano ₂₀₀₉ (<i>dummy</i>)	8,608*** (1,116)	2,905** (1,147)	2,632** (1,171)	2,669** (1,190)	2,064* (1,227)	1,931 (1,191)
Interação (política*ano₂₀₀₉)	-1,151 (1,943)	-2,620 (1,742)	-2,604 (1,741)	-2,591 (1,809)	-2,665 (1,865)	-2,499 (1,826)
R ²	0,066	0,271	0,274	0,276	0,287	0,284
Controle (características dos alunos)	Não	Sim	Sim	Sim	Sim	Sim
Controle (características da escola)	Não	Não	Sim	Sim	Sim	Sim
Controle (características dos diretores)	Não	Não	Não	Sim	Sim	Não
Controle (características dos professores)	Não	Não	Não	Não	Sim	Sim
Número de observações	1.194	1.194	1.194	1.194	1.128	1.128

Controles: características dos alunos (gênero, cor/raça, escolaridade da mãe, mais de uma televisão em casa, possui computador em casa), características da escola (*dummies* para captar presença de laboratório de informática e biblioteca), características dos diretores (pós-graduação e experiência em gestão), características dos professores (mais de dois anos de experiência, formação em ensino superior e treinamentos).

Fonte: Elaboração própria, com base nos dados do Inep.

*** Significativo ao nível de significância de 1%; ** Significativo ao nível de 5% de significância; * Significativo ao nível de 10% de significância. Desvio-padrão robusto entre parênteses.

A primeira coluna da tabela 7 apresenta os resultados quando não é utilizada nenhuma variável de controle, ou seja, estima as diferenças das médias de proficiência em língua portuguesa, captando o efeito da política de eleição de diretores escolares, para o período de 2007 a 2009. Reflete assim o mesmo resultado calculado na tabela 5, demonstrando que a diferença não é significativa. Na segunda coluna está a estimação quando são inseridas as variáveis de controle para características dos alunos; na terceira coluna, quando se acrescentam as variáveis de controle para características da escola; na quarta e quinta coluna, os resultados quando se adicionam as variáveis de controle de características do diretor escolar e professor; por fim, levando em conta que o perfil do diretor está intrinsecamente relacionado com o processo de seleção, seja por indicação política seja por eleição, na última coluna, consideram-se todos os controles, exceto as características do diretor.

Como se pode observar, a interação entre as variáveis “Política” (*dummy* igual a um para as observações do grupo de tratamento e zero, caso contrário) e “Ano₂₀₀₉” (*dummy* igual a um para o ano de 2009), que capta o impacto da política sobre a proficiência em língua portuguesa, não se mostrou significativa em nenhuma estimação, nem mesmo quando as variáveis de controle são inseridas.

TABELA 8

Impacto da política de eleição de diretores sobre a proficiência média em matemática das escolas estaduais da Bahia (2007 a 2009): estimativas por diferença em diferenças

Variáveis	Proficiência em matemática					
	(1)	(2)	(3)	(4)	(5)	(6)
Política (<i>dummy</i>)	0,451 (1,254)	-3,081*** (1,179)	-3,047** (1,198)	-3,357*** (1,257)	-3,015** (1,305)	-2,853** (1,209)
Ano ₂₀₀₉ (<i>dummy</i>)	1,609 (1,019)	-3,098*** (1,047)	-3,272*** (1,069)	-3,413*** (1,086)	-3,674*** (1,121)	-3,598*** (1,091)
Interação (política*ano₂₀₀₉)	-1,187 (1,774)	-2,278 (1,589)	-2,276 (1,589)	-1,885 (1,651)	-2,537 (1,704)	-2,779* (1,643)
R ²	0,002	0,223	0,225	0,228	0,233	0,230
Controle (características dos alunos)	Não	Sim	Sim	Sim	Sim	Sim
Controle (características da escola)	Não	Não	Sim	Sim	Sim	Sim
Controle (características dos diretores)	Não	Não	Não	Sim	Sim	Não
Controle (características dos professores)	Não	Não	Não	Não	Sim	Sim
Número de observações	1.194	1.194	1.194	1.194	1.128	1.128

Controles: características dos alunos (gênero, cor/raça, escolaridade da mãe, mais de uma televisão em casa, possui computador em casa), características da escola (*dummies* para captar presença de laboratório de informática e biblioteca), características dos diretores (pós-graduação e experiência em gestão), características dos professores (mais de dois anos de experiência, formação em ensino superior e treinamentos).

Fonte: Elaboração própria, com base nos dados do Inep.

*** Significativo ao nível de significância de 1%; ** Significativo ao nível de 5% de significância; * Significativo ao nível de 10% de significância. Desvio-padrão robusto entre parênteses.

Assim como para a proficiência em língua portuguesa, o impacto da política de eleição, em geral, não foi significativo para a proficiência média em matemática, como demonstrado na tabela 8.

A ausência de significância estatística do parâmetro que estima o efeito da política para o período de 2007 a 2009, para ambas as proficiências, pode ser justificada pelo curto período de implantação da política, não gerando ainda impacto sobre o desempenho dos alunos.

A segunda estimação de diferença em diferenças foi realizada com os dados para 2007 e 2011. Assim como na estimação anterior, as tabelas 9 e 10 ilustram o cálculo do modelo de diferença em diferenças, isto é, a diferença, ao longo do tempo, das diferenças das médias de proficiência em língua portuguesa e matemática, entre as escolas com e sem a política de seleção e eleição de diretores, considerando os anos de 2007 e 2011.

TABELA 9
Cálculo da diferença em diferenças entre os grupos de tratamento e controle para média em língua portuguesa (2007 a 2011)

	Antes do tratamento	Depois do tratamento	Mudança nas proficiências
	Ano de 2007	Ano 2011	
Escolas grupo de tratamento	215,96	224,69	8,73
Escolas grupo de controle	213,09	223,06	9,97
Diferença de média da proficiência em língua portuguesa	2,87	1,63	-1,24

Fonte: Elaboração própria, com base nos dados do Inep.

TABELA 10
Cálculo da diferença em diferenças entre os grupos de tratamento e controle para média em matemática (2007-2011)

	Antes do tratamento	Depois do tratamento	Mudança nas proficiências
	Ano de 2007	Ano 2011	
Escolas grupo de tratamento	224,97	229,07	4,10
Escolas grupo de controle	223,63	229,24	5,61
Diferença de média da proficiência em matemática	1,34	-0,17	-1,51

Fonte: Elaboração própria, com base nos dados do Inep.

Observa-se que o valor da diferença em diferenças é negativo, para ambas as proficiências, sendo de 1,51 pontos em matemática e de 1,24 pontos em língua portuguesa. No intuito de conhecer se esta diferença é estatisticamente diferente de zero, foram estimadas as regressões por mínimos quadrados, cujos resultados estão dispostos nas tabelas 11 e 12.

TABELA 11
Impacto da política de eleição de diretores sobre a proficiência média em língua portuguesa das escolas estaduais da Bahia (2007 a 2011)

Variáveis	Proficiência em língua portuguesa					
	(1)	(2)	(3)	(4)	(5)	(6)
Política (<i>dummy</i>)	2,869*	-2,133	-2,373*	-3,258**	-4,612***	-3,751**
	(1,512)	(1,358)	(1,377)	(1,436)	(1,569)	(1,538)
Ano ₂₀₁₁ (<i>dummy</i>)	9,969***	6,884***	6,447***	6,039***	5,685***	6,042***
	(1,083)	(1,247)	(1,308)	(1,319)	(1,371)	(1,338)
Interação (política*ano₂₀₁₁)	-1,238	-3,829**	-3,716**	-3,104	-1,546	-2,153
	(2,138)	(1,861)	(1,864)	(1,951)	(2,114)	(2,038)
R ²	0,096	0,352	0,352	0,361	0,387	0,379
Controle (características dos alunos)	Não	Sim	Sim	Sim	Sim	Sim
Controle (características da escola)	Não	Não	Sim	Sim	Sim	Sim
Controle (características dos diretores)	Não	Não	Não	Sim	Sim	Não
Controle (características dos professores)	Não	Não	Não	Não	Sim	Sim
Número de observações	1.060	1.060	1.060	1.060	974	974

Controles: características dos alunos (gênero, cor/raça, escolaridade da mãe, mais de uma televisão em casa, possui computador em casa), características da escola (*dummies* para captar presença de laboratório de informática e biblioteca), características dos diretores (pós-graduação e experiência em gestão), características dos professores (mais de dois anos de experiência, formação em ensino superior e treinamentos), características dos professores (mais de dois anos de experiência, formação em ensino superior e treinamentos).

Fonte: Elaboração própria, com base nos dados do Inep.

*** Significativo ao nível de significância de 1%; ** Significativo ao nível de 5% de significância; * Significativo ao nível de 10% de significância. Desvio-padrão robusto entre parênteses.

A primeira coluna da tabela 11, sem controles, demonstra que a diferença das diferenças, já calculada na tabela 9, não é significativa. A interação que capta o impacto da política sobre a proficiência em língua portuguesa (*política**ano₂₀₁₁) é significativa apenas quando se inserem as variáveis de controle para as características dos alunos e da escola e seu impacto é negativo no desempenho em Língua Portuguesa dos alunos que estudam em escolas com processo de eleição dos diretores. Ressalta-se que quando computadas as variáveis de características dos diretores e dos professores e todas as variáveis de controle, exceto as referentes ao diretor, o parâmetro deixa de ser significativo.

Assim como para língua portuguesa, a política de eleição de diretores impacta negativamente a proficiência em matemática (tabela 12), sendo esse impacto um pouco maior e significativo mesmo com a inclusão de todas as variáveis de controle com exceção das de diretor. O impacto deixa de ser significativo quando todos os controles são considerados.

TABELA 12
Impacto da política de eleição de diretores sobre a proficiência média em matemática das escolas estaduais da Bahia (2007 a 2011): estimativas por diferença em diferenças

Variáveis	Proficiência em Matemática					
	(1)	(2)	(3)	(4)	(5)	(6)
Política (<i>dummy</i>)	1,343 (1,596)	-3,075** (1,455)	-3,286** (1,474)	-3,586** (1,538)	-4,512*** (1,707)	-4,134*** (1,510)
Ano ₂₀₁₁ (<i>dummy</i>)	5,615*** (1,143)	3,057** (1,336)	2,601* (1,401)	2,373* (1,413)	2,377 (1,491)	2,645* (1,520)
Interação (política*ano₂₀₁₁)	-1,513 (2,257)	-4,302** (1,993)	-4,205** (1,997)	-4,023* (2,090)	-3,477 (2,300)	-3,645* (2,137)
R ²	0,292	0,292	0,292	0,292	0,298	0,290
Controle (características dos alunos)	Não	Sim	Sim	Sim	Sim	Sim
Controle (características da escola)	Não	Não	Sim	Sim	Sim	Sim
Controle (características dos diretores)	Não	Não	Não	Sim	Sim	Não
Controle (características dos professores)	Não	Não	Não	Não	Sim	Sim
Número de observações	1.060	1.060	1.060	1.060	974	974

Controles: características dos alunos (gênero, cor/raça, escolaridade da mãe, mais de uma televisão em casa, possui computador em casa), características da escola (*dummies* para captar presença de laboratório de informática e biblioteca), características dos diretores (pós-graduação e experiência em gestão), características dos professores (mais de dois anos de experiência, formação em ensino superior e treinamentos).

Fonte: Elaboração própria, com base nos dados do Inep.

*** Significativo ao nível de significância de 1%; ** Significativo ao nível de 5% de significância; * Significativo ao nível de 10% de significância. Desvio-padrão robusto entre parênteses.

Da mesma forma que para as estimativas com dados de 2007 e 2009, a política de eleição agora gera impacto negativo, e significativo ao nível de 10% quando apenas as características do diretor não são incluídas. Assim, a política não aumenta significativamente o desempenho médio dos alunos mesmo quando se analisa um maior intervalo de tempo entre período pré e pós implementação da política⁵.

Em suma, nos modelos de diferença em diferenças, o efeito da eleição do diretor é, em geral, negativo e não significativo para o desempenho em português dos alunos destas escolas e negativo e significante para o desempenho em matemática, quando incluídos todos os controles com exceção das variáveis referentes aos diretores.

Alguns autores levantam mecanismos que poderiam ocasionar um efeito adverso das eleições nas notas dos alunos. Holmesland *et al.* (1989, p.128 *apud* Paro, 1996) apontam que apesar da eleição de diretores buscar a superação de práticas tradicionalistas que não consideram a meritocracia e a participação da

5. Com uma amostra de escolas menor, composta apenas pelas que participaram da Prova Brasil nos três anos (2007, 2009, 2011), foram feitas novas estimações com o método de diferença em diferenças, novamente para os períodos de 2007 a 2009 e 2007 a 2011. As estimações apresentaram resultados similares aos já apresentados. Ainda, todos os modelos foram reestimados com a opção de *cluster* por município, e os resultados não se alteraram consideravelmente.

comunidade, esta pode resultar em práticas clientelistas da comunidade escolar que pode exigir uma contrapartida do eleito ao apoio dado durante a eleição, desvirtuando o processo.

Ainda sobre isso, Luck (2011, p.96), em estudo realizado com treze secretarias estaduais no Brasil, identifica nos discursos dos gestores que o processo de eleição de diretores pode induzir a “prática de clientelismo/corporativismo ou favoritismo no ambiente escolar” e “propiciar a politização e fragmentação de grupos nas escolas, provocando um ambiente de antagonismo”.

Por sua vez, Paro (1996, p. 380) ressalta que existia uma crença de que a eleição “conseguiria eliminar o autoritarismo existente na escola e a falta de participação de professores, alunos, funcionários e pais nas decisões”, entretanto, esse cenário não foi alcançado e a participação dos segmentos escolares nas decisões não se confirmou, tendo em vista que dentro do processo democrático a participação e o envolvimento das pessoas é apenas uma possibilidade, não uma garantia. Ainda, para o autor, entre os limites da eleição de diretores, existe o corporativismo por parte dos grupos que interagem dentro da escola, em que se observa uma atitude de professores buscando favorecimento em troca de seu apoio a determinado candidato.

Outra possibilidade é que uma maior aprovação de alunos com baixo rendimento no ano anterior pode levar a um rendimento menor dos alunos que passaram. Assim, a diferença negativa de desempenho entre as escolas com eleição de diretor (estaduais) e as sem eleição (municipais) poderia ser ocasionada por variação nas taxas de aprovação entre as duas redes de ensino. Um aumento na taxa de aprovação da rede estadual poderia resultar em um pior desempenho médio dos alunos devido à inclusão de alunos com pior desempenho e não devido a um menor aprendizado por aluno. Para testar essa hipótese, foi feito um modelo de diferença em diferenças para as bases de dados de 2007-2009 e 2007-2011, com as mesmas variáveis explicativas, mas alterando a variável dependente para Taxa de Aprovação. Os resultados estão na tabela 13.

TABELA 13

Impacto da política de eleição de diretores sobre a taxa de aprovação das escolas estaduais da Bahia: estimativas por diferença em diferenças

Variáveis	Taxa de aprovação			
	2007 - 2009		2007 - 2011	
Política (<i>dummy</i>)	-4,284*** (1,308)	-4,909*** (1,328)	-3,639** (1,502)	-4,679*** (1,451)
Ano ₂₀₀₉ (<i>dummy</i>)	3,627*** (1,125)	3,390*** (1,063)	5,941*** (1,306)	5,804*** (1,322)
Interação (política*ano)	-0,88 (1,706)	-0,535 (1,728)	-4,131** (2,014)	-2,787 (2,127)
R ²	0,095	0,089	0,103	0,092

(Continua)

(Continuação)

Variáveis	Taxa de aprovação			
	2007 - 2009		2007 - 2011	
Controle (características dos alunos)	Sim	Sim	Sim	Sim
Controle (características da escola)	Sim	Sim	Sim	Sim
Controle (características dos diretores)	Sim	Não	Sim	Não
Controle (características dos professores)	Sim	Sim	Sim	Sim
Número de observações	1.115	1.115	963	963

Controles: características dos alunos (gênero, cor/raça, escolaridade da mãe, mais de uma televisão em casa, possui computador em casa), características da escola (*dummies* para captar presença de laboratório de informática e biblioteca), características dos diretores (pós-graduação e experiência em gestão), características dos professores (mais de dois anos de experiência, formação em ensino superior e treinamentos).

Fonte: Elaboração própria, com base nos dados do Inep.

*** Significativo ao nível de significância de 1%; ** Significativo ao nível de 5% de significância; * Significativo ao nível de 10% de significância. Desvio-padrão robusto entre parênteses.

A variável de interesse (de interação entre política e ano) mostrou-se significativa apenas no período entre 2007 e 2011 no modelo que inclui as características dos diretores. Isto é, as escolas estaduais, cujos diretores são eleitos, obtiveram uma taxa de aprovação inferior à das escolas municipais em aproximadamente 4,1 pontos percentuais ou 0,10 desvios-padrão. Assim, não se pode atribuir o resultado pior das estaduais, no 9º ano, a uma maior aprovação.

Outro indicador educacional a ser observado é a taxa de abandono escolar. Neste caso, quando a relação observada é negativa, trata-se, então, de um fator favorável. Na tabela 14 são apresentados os resultados para os dois períodos.

TABELA 14

Impacto da política de eleição de diretores sobre a taxa de abandono das escolas estaduais da Bahia: estimativas por diferença em diferenças

Variáveis	Taxa de abandono			
	2007 - 2009		2007 - 2011	
Política (<i>dummy</i>)	3,852*** (0,949)	3,985*** (1,053)	3,683*** (1,259)	3,894*** (1,214)
Ano ₂₀₀₉ (<i>dummy</i>)	-0,885 (0,819)	-0,790 (0,763)	-2,630*** (0,952)	-2,665*** (0,956)
Interação (política*ano)	-0,496 (1,242)	-0,616 (1,274)	-3,391** (1,503)	-3,731** (1,460)
R ²	0,110	0,105	0,134	0,125
Controle (características dos alunos)	Sim	Sim	Sim	Sim
Controle (características da escola)	Sim	Sim	Sim	Sim
Controle (características dos diretores)	Sim	Não	Sim	Não
Controle (características dos professores)	Sim	Sim	Sim	Sim
Número de observações	1.104	1.104	954	954

Controles: características dos alunos (gênero, cor/raça, escolaridade da mãe, mais de uma televisão em casa, possui computador em casa), características da escola (*dummies* para captar presença de laboratório de informática e biblioteca), características dos diretores (pós-graduação e experiência em gestão), características dos professores (mais de dois anos de experiência, formação em ensino superior e treinamentos).

Fonte: Elaboração própria, com base nos dados do Inep.

*** Significativo ao nível de significância de 1%; ** Significativo ao nível de 5% de significância; * Significativo ao nível de 10% de significância. Desvio-padrão robusto entre parênteses.

Observa-se que nas escolas estaduais, onde aconteceram eleições de diretores (variável de interação), a redução na taxa de abandono escolar foi entre 3,4 e 3,7 p.p (0,12 e 0,13 desvios-padrão) dependendo da inclusão das características dos diretores como controle no modelo. Uma redução na taxa de abandono tende a reduzir a nota média das escolas (Menezes Filho, 2011). Assim, embora a política possibilite melhoria neste indicador educacional, ele pode explicar a redução no desempenho médio nas provas de proficiência devido ao aumento de alunos com baixo desempenho dentro da escola.

5.2 Dados em painel

Conforme discutido anteriormente, o modelo de diferença em diferenças não considera características não observáveis das escolas que são constantes ao longo do tempo. Se essas características não observáveis existirem e forem correlacionadas com alguma das variáveis explicativas do modelo, temos o famoso problema de variáveis omitidas. Nesta seção, exploramos a estrutura em painel dos dados, para contornar este potencial problema: para calcular o impacto do programa, foram utilizadas técnicas para dados em painel com efeito fixo e com efeito aleatório. Com base no teste de Hausman, constatou-se que o modelo mais adequado é o de efeito fixo, uma vez que os resultados do teste rejeitaram a hipótese nula de que os erros não são correlacionados com as variáveis explicativas. Assim, são apresentadas apenas as estimativas do modelo de efeito fixo⁶.

A seguir são apresentadas as estimativas do impacto da política de eleição de diretores sobre a proficiência média em língua portuguesa e matemática (tabela 15) e para as taxas de aprovação e de abandono (tabela 16) das escolas baianas de 9º ano para os painéis de 2007-2009 e 2007-2011 do modelo de efeitos fixos com a inclusão dos controles dos alunos, da escola e dos professores. Optou-se por não incluir os controles do diretor, pois estes são decorrentes do processo de seleção do diretor adotado pela rede de ensino, além disso a sua inclusão não altera os resultados significativamente.

Assim, ao consideramos características não observáveis das escolas que estão fixas ao longo do tempo, constata-se um efeito não significativo da política de eleição de diretores no estado da Bahia para a proficiência em matemática e língua portuguesa.

6. No modelo de efeito aleatório foi encontrado um efeito significativo e negativo para as escolas que passaram a adotar a política para os dados de 2007 a 2009 e 2007 a 2011.

TABELA 15
Impacto da política de seleção e eleição de diretores sobre a proficiência média das escolas estaduais da Bahia: estimativas por dados em painel

Variáveis	2007-2009		2007-2011	
	Proficiência em língua portuguesa	Proficiência em matemática	Proficiência em língua portuguesa	Proficiência em matemática
Ano (<i>dummy</i>)	4,576*** (1,258)	-0,791 (1,030)	6,999*** (1,721)	2,589 (1,730)
Política (<i>dummy</i>)	-1,670 (1,467)	-1,699 (1,247)	-0,0401 (1,715)	-1,574 (1,541)
R ²	0,284	0,086	0,389	0,214
Número de observações	1.144	1.144	974	974

Controles: características dos alunos (gênero, cor/raça, escolaridade da mãe, mais de uma televisão em casa, possui computador em casa), características da escola (*dummies* para captar presença de laboratório de informática e biblioteca), características dos professores (mais de dois anos de experiência, formação em ensino superior e treinamentos).

Fonte: Elaboração própria, com base nos dados do Inep.

*** Significativo ao nível de significância de 1%; ** Significativo ao nível de 5% de significância; * Significativo ao nível de 10% de significância. Desvio-padrão robusto entre parênteses.

TABELA 16
Impacto da política de seleção e eleição de diretores sobre a proficiência média das escolas estaduais da Bahia: estimativas por dados em painel

Variáveis	2007-2009		2007-2011	
	Taxa de aprovação	Taxa de abandono	Taxa de aprovação	Taxa de abandono
Ano (<i>dummy</i>)	1,046 (1,127)	-0,212 (0,763)	1,740 (1,590)	-0,722 (1,110)
Política (<i>dummy</i>)	-0,839 (1,301)	-0,571 (0,915)	-3,856** (1,926)	-4,083*** (1,207)
R ²	0,033	0,052	0,086	0,133
Número de observações	1.102	1.104	952	954

Controles: características dos alunos (gênero, cor/raça, escolaridade da mãe, mais de uma televisão em casa, possui computador em casa), características da escola (*dummies* para captar presença de laboratório de informática e biblioteca), características dos professores (mais de dois anos de experiência, formação em ensino superior e treinamentos).

Fonte: Elaboração própria, com base nos dados do Inep.

*** Significativo ao nível de significância de 1%; ** Significativo ao nível de 5% de significância; * Significativo ao nível de 10% de significância. Desvio-padrão robusto entre parênteses.

Além disso, a política apresentou impactos negativos e significantes nas taxas de aprovação e de abandono para o período 2007-2011⁷: escolas que adotaram a política de eleição dos diretores obtiveram uma redução de aproximadamente 3,8 p.p. (0,09 desvios-padrão) na taxa de aprovação escolar e ao mesmo tempo obtiveram uma redução de aproximadamente 4,1 p.p. (0,15 desvios-padrão) na

7. Foi realizada uma estimação com painel de dados para os três períodos 2007, 2009 e 2011, confirmando os resultados já apresentados anteriormente.

taxa de abandono escolar. Enquanto a redução na taxa de abandono escolar é um resultado educacional positivo de tal política, a interpretação com relação à taxa de aprovação escolar não é tão clara. Em geral, apesar de indicadores de avaliação escolar, como o Índice de Desenvolvimento da Educação Básica (Ideb), considerarem desejáveis maiores taxas de aprovação, a aprovação apenas é positiva se o aluno tiver um nível adequado para acompanhar o ensino no ano subsequente.

Em resumo, a política de eleição dos diretores teve pouco (ou nenhum) efeito no desempenho escolar dos alunos: se não atrapalhou, tampouco ajudou a aumentar o rendimento dos alunos. Porém, a política trouxe consigo o benefício de manter mais alunos na escola, apesar da diminuição na taxa de aprovação.

6 CONSIDERAÇÕES FINAIS

O objetivo deste estudo foi analisar o efeito da eleição de diretores sobre o desempenho escolar dos alunos de 9º ano da rede pública estadual da Bahia – com o desempenho sendo medido pela proficiência média das escolas, nas avaliações de língua portuguesa e Matemática da Prova Brasil. Para tanto foram empregados os métodos de dados em painel e de diferença em diferenças, comparando os grupos de tratamento (escolas estaduais) e controle (escolas municipais), antes (em 2007) e após (em 2009 e 2011) a implementação da política.

Os resultados mostram que não houve aumento nas médias de proficiência em língua portuguesa e matemática dos alunos quando os diretores passaram a ser eleitos pela comunidade escolar em vez de indicados politicamente. De fato, o impacto das eleições dos diretores no desempenho dos alunos foi não significativo na maior parte dos modelos analisados e, quando significativo, foi no sentido de reduzir o desempenho médio dos estudantes

Quando são analisadas outras variáveis de desempenho educacional – as taxas de aprovação e de abandono – observa-se que a eleição de diretor gera uma diminuição na primeira taxa, o que possibilita concluir que a nota média da série não é menor devido à aprovação de alunos com pior rendimento. Entretanto, a menor taxa de abandono das escolas estaduais pode estar ocasionando redução no desempenho médio, por meio da retenção de alunos com baixo desempenho. Com efeito, o principal benefício educacional da adoção da política identificado neste trabalho foi uma redução de aproximadamente 4% na taxa de abandono escolar.

Os resultados encontrados não convergem com os resultados de trabalhos da literatura para o Brasil que estimaram esse efeito, uma vez que, nestes, quando o modo de provimento do cargo de diretor apresenta impacto sobre o desempenho dos alunos, o efeito é positivo. Porém, ressalta-se que os trabalhos anteriores utilizaram bases de dados, metodologias e períodos de tempo diferentes dos

aqui apresentados, e pouca preocupação em determinar um efeito causal desta política no rendimento dos alunos.

Possíveis explicações teóricas para a ocorrência de um efeito negativo na eleição dos diretores referem-se aos efeitos perversos que esse modo de seleção pode assumir: clientelismo entre os professores, corporativismo e politização, bem como pode decorrer da retenção de alunos com pior desempenho na escola. Por outro lado, a eleição dos diretores elimina o clientelismo político e pode fortalecer os vínculos dos membros da comunidade, principalmente dos pais dos alunos com a escola, o que poderia trazer um impacto positivo no desempenho dos alunos. Uma possibilidade é a de que a combinação desses diversos mecanismos que atuam em direções opostas equilibre-se de tal forma que não foi observado, no modelo mais geral, nem piora nem melhora no desempenho dos alunos em decorrência da mudança da política de seleção dos diretores.

Além disso, a falta de incentivos para selecionar bons gestores escolares e/ou estimular seu esforço pode, também, explicar os resultados encontrados. Luck (2011) demonstra uma baixa procura dos professores por esta atividade, com muitos dos estados com processo de eleição de diretores nos quais não existe disputa para a vaga ou nem mesmo registro de chapa única. Na Bahia, 22% das escolas não tinham candidatos e 75% das escolas apresentaram chapa única (Luck, 2011).

Como pesquisas futuras, sugere-se a aplicação da metodologia de painel para bases de dados com escolas de outros estados brasileiros que possibilitem a mesma comparação entre escolas que tiveram e não tiveram seus diretores eleitos. Em uma linha mais qualitativa, propõe-se uma investigação para analisar uma amostra aleatória de escolas antes, durante e após o período de eleições para identificar se há de fato alguma alteração no ambiente escolar devido à adoção da gestão democrática.

REFERÊNCIAS

- ALVES, F. Políticas educacionais e desempenho escolar nas capitais brasileiras. **Cad. Pesqui.**, São Paulo, v. 38, n. 134, ago. 2008.
- ANDRADE, R. J.; SOARES, J. F. O efeito da escola básica brasileira. **Estudos em Avaliação Educacional**. v. 19, n. 41, set./dez. 2008. Disponível em: <<http://www.fcc.org.br/pesquisa/publicacoes/eae/arquivos/1456/1456.pdf>>. Acesso em: 7 jan. 2013.
- AQUINO, J. M. de; KASSOUF, A. L. Ampliação da jornada escolar melhora o desempenho acadêmico dos estudantes? Uma avaliação do programa "escola em tempo integral" da rede pública do estado de São Paulo. **Rede de economia aplicada**. São Paulo, out. 2011. Disponível em: <<http://anpec.org.br/encontro/2011/inscricao/arquivos/000-1327c2b4283497c851976b4bf6125170.pdf>>. Acesso em: 5 jul. 2012.

ARAÚJO, C. H.; LUZIO, N. **Avaliação da Educação Básica**: em busca da qualidade e equidade no Brasil. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2005.

BAHIA. Decreto nº 11.218 de 18 de setembro de 2008. Regulamenta o artigo 18 da Lei nº 8.261, de 29 de maio de 2002, dispondo sobre os critérios e procedimentos do processo seletivo interno a ser realizado pela unidade escolar, requisitos para o preenchimento dos cargos de Diretor e Vice-diretor das Escolas Públicas do Estado da Bahia, e dá outras providências. **Diário Oficial do Estado**, Bahia, 2008.

_____. Lei nº 8.261 de 29 de maio de 2002. Dispõe sobre o Estatuto do Magistério Público do Ensino Fundamental e Médio do Estado da Bahia e dá outras providências, **Diário Oficial do Estado**, Bahia, 2002.

BARROS, R. P. de, MENDONÇA, R., SANTOS, D. D. dos; QUINTAES, G. **Determinantes do desempenho educacional no Brasil**. IPEA, Texto para discussão n. 834, 2001. Disponível em: <http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=4075>. Acesso em: 13 set. 2011.

BARROS, R. P.; MENDONÇA, R. S. P. O Impacto de Gestão sobre o Desempenho Educacional. *In*: **Série Documentos de Trabalho R-301**. Washington: BID, 1997.

BIONDI, R. L.; FELÍCIO, F. **Atributos Escolares e o Desempenho dos Estudantes**: uma Análise em Painel dos Dados do SAEB. Brasília: Inep, 2007.

BONELLI, R. Crescimento, Desigualdade e Educação: notas para uma resenha com referência ao Brasil. **Economia Aplicada**, n.6, p.819-873, 2002.

BRASIL. **Constituição Federal (1988)**. Constituição da República Federativa do Brasil. Brasília, DF, Senado, 1998.

_____. **Lei de Diretrizes e Bases da Educação Nacional (LDB)** Lei nº 9.394/1996. Promulgada em 20/12/1996. Brasília, 1996.

BRYK, A., CAMBURN, E., & SEASHORE LOUIS, K. Professional community in Chicago elementary schools: Facilitating factors and organizational consequences. **Educational Administration Quarterly**, 35, 1999.

CADAVAL, A. F.; MONTEIRO, S. M. M. Determinantes da qualidade da educação fundamental no Brasil: uma análise com dados do SAEB. *In*: 40º Encontro Nacional de Economia ANPEC. 2011. Disponível em: <<http://goo.gl/5Evtx>>. Acesso em: 30 out. 2012.

CREEMERS, B.P.M.; REEZIGT, G.J. School level conditions affecting the effectiveness of instruction. **School Effectiveness a and School Improvement**, 7, 197-228, 1996.

CURI, A. Z.; MENEZES FILHO, N. A. Mensalidade escolar, background familiar e os resultados do exame nacional do ensino médio (ENEM). **Pesquisa e Planejamento Econômico**, v. 43, n. 2, ago. 2013

FRANCO, C.; FERNANDES, C.; SOARES, J.F.; BELTRÃO, K.; BARBOSA, M.E.; ALVES, M.T.G. O Referencial teórico na construção dos questionários contextuais do SAEB. **Ensaio em Avaliação Educacional**, 28, pp. 39-74, 2001.

HALLINGER, P.; HECK, R. H. The principal's role in school effectiveness: An assessment of methodological progress, 1980-1995. *In*: LEITHWOOD, K. *et al.* **International handbook of educational leadership and administration** (pp. 723-783), 1996a.

_____. Reassessing the principal's role in school effectiveness: A review of empirical research, 1980-1995. **Educational Administration Quarterly**, 32 (1), 5-44, 1996b.

_____. Exploring the principal's contribution to school effectiveness: 1980-1995. **School Effectiveness and School Improvement**, 9 (2), 157-191, 1998.

HILL, P. Shaking the foundations: Research-driven school reform. **School effectiveness and school improvement**, 9 (4), 419-436, 1998.

KLEIN, R. Como está a educação no Brasil? O que fazer? **Ensaio: Avaliação de políticas públicas Educacionais**, Rio de Janeiro, v.14, n.51, p. 139-172, 2006. Disponível em: <<http://www.scielo.br/pdf/ensaio/v14n51/a02v1451.pdf>>. Acesso em: 13 set. 2012.

LEE, V.E.; BRYK, A.S.; SMITH, J. (1993). The organization of effective secondary schools. *En*: Darling-Hammond, L. (Org). **Review of Research in Education**. Washington: American Educational Research Association, n. 19, pp. 171-268. Disponível em: <<http://www.jstor.org/discover/10.2307/1167343?uid=3737664&uid=2&uid=4&sid=21102963192607>>. Acesso em: 13 fev. 2013.

LEE, V. E. What are multilevel questions and how might we explore them with quantitative methods? **Est. Aval. Educ.**, São Paulo, n. 24, dez. 2001. Disponível em <http://educa.fcc.org.br/scielo.php?script=sci_arttext&pid=S0103-68312001000200002&lng=pt&nrn=iso>. Acesso em: 13 set. 2012.

LUCK, H. Mapeamento de práticas de seleção e capacitação de diretores escolares. **Fundação Victor Civita**. Curitiba, 2011. Disponível em: <<http://www.fvc.org.br/estudos-e-pesquisas/2010/mapeamento-praticas-selecao-capacitacao-diretores-escolares-620726.shtml>>. Acesso em: 3 fev. 2013

MENDONÇA, E. F. Estado patrimonial e gestão democrática do ensino público no Brasil. **Educação & Sociedade**, ano XXII, nº 75, ago. 2001. Disponível em: <<http://www.scielo.br/pdf/es/v22n75/22n75a07.pdf>>. Acesso em: 9 jul. 2013.

MENEZES-FILHO, N. A. **Os Determinantes do Desempenho Escolar do Brasil**. Instituto Futuro Brasil, Ibmecc-SP e FEA-USP, 2007.

MEYER, L.; FEISTRITZER, E. **Better leaders for America's schools**: a manifesto. With profiles of education leaders and a summary of State certification practices. Washington, D.C. Broad Foundation and Thomas Fordham Foundation, 2003.

MYERS, J. P. Democratizing school authority: Brazilian teachers' perceptions of the election of principals Original. **Research Article Teaching and Teacher Education**, v. 24, issue 4, may 2008, pages 952-966.

NORMORE, A. H. Recruitment and selection: meeting the leadership shortage in one large Canadian school district. **Canadian Journal of Educational Administration and Policy**, n. 30, maio, 2004.

OSHIRO, C. H.; SCORZAFAVE, L. G. Efeito do pagamento de bônus aos professores sobre a proficiência escolar no Estado de São Paulo. In: 39^o Encontro Nacional de Economia. 2001. Disponível em <<http://anpec.org.br/encontro/2011/inscricao/arquivos/000-e95d077296d03faa27e429dc66192da8.pdf>> Acesso em: 20 fev. 2013.

PARO, V. H. Eleição de Diretores de Escolas Públicas: Avanços e Limites da Prática. **R. bras. Est. pedag.**, Brasília, v.77, n.186, p. 376-395, maio/ago. 1996. Disponível em: <<http://rbep.inep.gov.br/index.php/RBEP/article/viewFile/289/291>>. Acesso em: 8 ago. 2012.

RIVKIN, S. G., HANUSHEK, E. A. e KAIN, J. F. Teachers, Schools, and Academic Achievement. **Econometrica**, v. 73, n. 2, March 2005, p. 417-458.

SAES, D. A. M.; ALVES, M. L. Conflitos ideológicos em torno da eleição de diretores de escolas públicas. In: 27 a. Reunião Annual da ANPED. Caxambu, MG, 2004. Disponível em: <<http://www.anped.org.br/reunioes/27/gt05/t053.pdf>>. Acesso em: 15 dez. 2012.

SÁTYRO, N.; SOARES, S. A infraestrutura das escolas brasileiras de ensino fundamental: um estudo com base nos Censos Escolares de 1997 a 2005. **Texto para Discussão**, n.1267, Brasília: Ipea, 2007.

SEBASTIAN, J.; ALLENSWORTH, E. (2012). The influence of principal leadership on classroom instruction and student learning: A study of mediated pathways to learning. **Educational Administration Quarterly**, 48(4), 626-663. Disponível em: <<http://ccsr.uchicago.edu/publications/influence-principal-leadership-classroom-instruction-and-student-learning-study>>. Acesso em: 13 fev. 2013.

SOARES, J. F. O efeito da escola no desempenho cognitivo de seus alunos. **Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación**, v. 2, n. 2, 2004.

SOARES, T. M.; TEIXEIRA, L. H. G. Efeito do perfil do diretor na gestão escolar sobre a proficiência do aluno. **Estudos em Avaliação Educacional**, São Paulo, v. 17, n. 34, maio/ago. 2006. Disponível em: <http://www.fcc.org.br/pesquisa/publicacoes/ae/arquivos/1289/1289.pdf>. Acesso em: 1 nov. 2011.

SOUZA, A. R. de. As relações entre os resultados da avaliação e os modelos de gestão escolar. **Intermeio** (UFMS), v. 13, p. 66-83, 2008. Disponível em: <http://www.intermeio.ufms.br/revistas/25/25%20Artigo_05.pdf>. Acesso em: 30 out. 2011.

TAVARES, P. A. Os impactos de práticas de gestão escolar sobre o desempenho educacional: evidências para escolas públicas paulistas. *In*: **34o Encontro Brasileiro de Econometria**, 2012. Disponível em: <http://cmicro.fgv.br/sites/cmicro.fgv.br/files/file/WP%207_2012.pdf> Acesso em: jan. 2013.

TEKESTE, Marion. The recruitment and selection of in-school administrators in Saskatchewan: a summary of a master's thesis. **SSTA Research Center Report**, 2000

WOOLDRIDGE, J. M. **Introdução à econometria**: uma abordagem moderna. São Paulo: Cengage Learning, 2001.

(Originais submetidos em dezembro de 2013. Última versão recebida em maio de 2014. Aprovada em março de 2015.)

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/pppe/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

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.*

Livraria Ipea

SBS – Quadra 1 – Bloco J – Ed. BNDES, Térreo

70076-900 – Brasília – DF

Fone: (61) 2026 5336

Correio eletrônico: livraria@ipea.gov.br

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



ipea Instituto de Pesquisa
Econômica Aplicada

Ministério do
Planejamento

**Governo
Federal**