

2435

**BRASIL SEM IMIGRANTES:
ESTIMATIVAS DE LONGO PRAZO
BASEADAS EM MICRODADOS**

**Leonardo Monasterio
Daniel Lopes**

TEXTO PARA DISCUSSÃO



BRASIL SEM IMIGRANTES: ESTIMATIVAS DE LONGO PRAZO BASEADAS EM MICRODADOS¹

Leonardo Monasterio²
Daniel Lopes³

1. Agradecemos a Pedro Souza pelos comentários. Somos gratos também a Rodrigo Soares e Fábio Vaz pelo apoio na obtenção e no processamento de dados. Detalhes no apêndice B.

2. Técnico de planejamento e pesquisa da Diretoria de Estudos e Políticas Regionais, Urbanas e Ambientais (Dirur) do Ipea e professor da Universidade Católica de Brasília (UCB).

3. Pesquisador do Programa de Pesquisa para o Desenvolvimento Nacional (PNPD) na Diretoria de Estudos e Políticas Sociais (Disoc) do Ipea.

Governo Federal

**Ministério do Planejamento,
Desenvolvimento e Gestão**
Ministro Esteves Pedro Colnago Junior

ipea Instituto de Pesquisa
Econômica Aplicada

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 brasileiros – e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

Presidente
Ernesto Lozardo

Diretor de Desenvolvimento Institucional
Rogério Boueri Miranda

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

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

**Diretor de Estudos e Políticas Regionais,
Urbanas e Ambientais**
Alexandre Xavier Ywata de Carvalho

**Diretor de Estudos e Políticas Setoriais de Inovação
e Infraestrutura**
Fabiano Mezadre Pompermayer

Diretora de Estudos e Políticas Sociais
Lenita Maria Turchi

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

Assessora-chefe de Imprensa e Comunicação
Mylena Pinheiro Fiori

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

Texto para Discussão

Publicação seriada que divulga resultados de estudos e pesquisas em desenvolvimento pelo Ipea com o objetivo de fomentar o debate e oferecer subsídios à formulação e avaliação de políticas públicas.

© Instituto de Pesquisa Econômica Aplicada – **ipea** 2018

Texto para discussão / Instituto de Pesquisa Econômica Aplicada.- Brasília : Rio de Janeiro : Ipea , 1990-

ISSN 1415-4765

1. Brasil. 2. Aspectos Econômicos. 3. Aspectos Sociais.
I. Instituto de Pesquisa Econômica Aplicada.

CDD 330.908

As publicações do Ipea estão disponíveis para *download* gratuito nos formatos PDF (todas) e EPUB (livros e periódicos).
Acesse: <http://www.ipea.gov.br/portal/publicacoes>

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 do Ministério do Planejamento, Desenvolvimento e Gestão.

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

JEL: F22; O54.

SUMÁRIO

SINOPSE

ABSTRACT

1 INTRODUÇÃO	7
2 FONTES HISTÓRICAS	10
3 ENUMERAMENTO DOS IMIGRANTES	12
4 NOVAS ESTIMATIVAS PARA O IMPACTO DA IMIGRAÇÃO	15
5 DISCUSSÃO	22
6 CONSIDERAÇÕES FINAIS	23
REFERÊNCIAS	23
APÊNDICE A	27
APÊNDICE B	29

SINOPSE

Este trabalho estima o efeito da imigração não ibérica para o Brasil, com base em microdados históricos e contemporâneos. A base histórica engloba mais 1,7 milhão de registros de imigrantes; por sua vez, a contemporânea parte de um banco de dados com mais 165 milhões registros administrativos. O cálculo do enumeramento dos imigrantes permite afirmar que Stolz, Baten e Botelho (2013) subestimaram essa habilidade dos estrangeiros e, portanto, seu impacto no Brasil. Um algoritmo de classificação de sobrenomes categorizou a população brasileira contemporânea em grupos ancestrais. Em seguida, constroem-se então duas estimativas contrafatuais do que seria a renda *per capita* se nunca tivesse havido imigração não ibérica. Um contrafactual decorre da regressão dessa renda pelos percentuais de cada grupo ancestral nos municípios. O outro resulta da regressão de salários individuais sobre a ancestralidade de sobrenome de cada trabalhador. Os coeficientes obtidos em ambas são usados para estimar a renda *per capita* em um Brasil sem descendentes de imigrantes. Nesse caso, estimou-se que a renda *per capita* brasileira seria entre 12,6% e 17% menor que a hoje observada.

Palavras-chave: imigração; capital humano; enumeramento.

ABSTRACT

This paper estimates the effect of non-Iberian immigration to Brazil based on historical and contemporary microdata. The historical base encompasses over 1.7 million immigrant records; the contemporary has more than 165 million records. The estimation of immigrant numeracy suggest that Stolz, Baten e Botelho (2013) underestimated their skills and, therefore, their impact on Brazil. An algorithm classified the surnames of contemporary Brazilians according to their ancestral origins. Two counterfactual estimates are constructed in order to estimate the income per capita if there had never been any non-Iberian immigration. The first counterfactual is built upon the regression of income on the percentages of each ancestral group in municipalities. The second, results from the regression of individual wages on the surname ancestry workers. The coefficients are used to estimate income in couterfactuals Brazils with no descendants of immigrants. It was estimated that in the absence of non-Iberian immigrants today's income would be from 12.6 % to 17 % lower.

Keywords: immigration; human capital; numeracy.

1 INTRODUÇÃO

Em 1872, o Brasil possuía cerca de 10 milhões de habitantes. Nas seis décadas que se seguiram, mais de 4 milhões de estrangeiros entraram no Brasil (Levy, 1974). Se, por um lado, é do senso comum que essa imigração teve efeitos relevantes, por outro, faltam estimativas do real impacto econômico desse choque positivo na oferta de mão de obra. Assim, o objetivo deste trabalho é estimar os efeitos de longo prazo da imigração não ibérica.

Para construir o contrafactual que intitula este estudo, parte-se de fontes históricas de registros individuais desses imigrantes, bem como bases contemporâneas de microdados de cidadãos brasileiros. Até onde se sabe, as fontes primárias com dados de imigrantes ainda não haviam sido examinadas de forma agregada. Analisam-se aqui os registros de imigrantes que reúnem mais de 1,7 milhão de observações, de mais de 67 nacionalidades que chegaram ao Brasil entre 1840 e 1958. Igualmente, fez-se uma análise inédita de registros administrativos contemporâneos – Ministério do Trabalho e Emprego (Brasil, 2015), Ministério do Desenvolvimento Social (Brasil, 2016) e Receita Federal do Brasil (RFB)¹ –, que somam mais de 165 milhões de observações. O conjunto histórico e o contemporâneo de bases individuais permitem visões e análises inéditas sobre o impacto da imigração no Brasil.

O *paper* de Stolz, Baten e Botelho (2013) tem relação próxima com este trabalho. Os autores usaram uma amostra dos dados de imigrantes do Arquivo Nacional do Rio de Janeiro para, obtendo o censo de brasileiros e estrangeiros, estimar o impacto das imigrações no longo prazo. Eles o fazem com base em estimação econométrica por países que relaciona o censo em 1900 com o produto interno bruto (PIB) *per capita* em anos recentes. Como será visto mais adiante, a utilização de bancos de dados mais amplos de imigrantes e bases de microdados contemporâneas de grande escala nos levou a resultados bem distintos dos obtidos em Stolz, Baten e Botelho (2013).

Como se sabe, a imigração não ibérica ao Brasil foi promovida quando se tornou evidente que a escravidão como força de trabalho se tornara insustentável (Reis e Reis, 1988; Lago, 2014). Houve experimentos com a imigração subsidiada ao longo

1. Receita Federal do Brasil (RFB). Dados Abertos do Cadastro Nacional da Pessoa Jurídica (CNPJ). Disponível em: <<https://bit.ly/2Lpc6CF>>.

do século XIX, mas o fluxo mais importante concentrou-se após a emancipação dos escravos, em 1888. Entre 1872 e 1920, mais de 3,2 milhões de estrangeiros desembarcaram no Brasil (Levy, 1974).

Em 1920, 5,1% da população era de origem estrangeira ou naturalizada. Ao longo da década de 1930, especialmente durante o Estado Novo, o país foi se fechando à imigração. E, desde então, não houve mais grandes ondas migratórias para o país. Em 2010, apenas 0,3% da população nasceu no exterior, um dos menores percentuais do mundo.

A imigração subsidiada foi um choque positivo de capital humano. O nível educacional no Brasil era substancialmente baixo, até mesmo quando comparado com outros países pobres (Chaudhary *et al.*, 2012). Em 1920, apenas 23% dos brasileiros, de todas as idades, conseguiam ler ou escrever. Em geral, os imigrantes não eram altamente qualificados em seus países de origem; porém, suas dotações de capital humano eram provavelmente mais elevadas que as dos brasileiros, sendo que 52% dos estrangeiros eram alfabetizados naquele ano (Brasil, 1929).

A importância da mão de obra estrangeira para a modernização brasileira está bem assentada na literatura. Pereira (1964) apontou que, na época, 84,3% dos empresários paulistas declaram que sua origem étnica era estrangeira – em até três gerações. Estrangeiros compunham 40% e 44% da mão de obra industrial em Rio de Janeiro e São Paulo, respectivamente, em 1920 (Versiani, 1993). Monasterio e Reis (2008) também mostraram que – no Brasil como um todo – a presença de estrangeiros nos municípios esteve associada à manufatura entre 1872 e 1920, até mesmo com amplos controles estatísticos.²

Indiretamente, este trabalho se relaciona com a crescente literatura sobre a transmissão cultural, as normas e o impacto de fatores culturais sob as instituições e o desenvolvimento econômico; em particular, Bisin e Verdier (2000), Alesina, Giuliano e Nunn (2013) e Fernández e Fogli (2009).

2. Os autores também apontam que a presença de imigrantes, junto com a expansão das ferrovias, induziu forças aglomerativas que explicam a concentração espacial da industrialização brasileira.

Tabellini (2008) atribui a persistência das instituições a indicadores concernentes a crenças e valores individuais, como confiança e respeito aos outros. Especificamente, buscamos relacionar a importância da cultura para a formação de capital humano – por exemplo, Guiso, Sapienza e Zingales (2006), Figlio *et al.* (2016) e Sørensen *et al.* (2016).

Nossa análise dialoga com a vasta literatura sobre determinantes históricos das instituições locais e *performance* econômica no Brasil. Naritomi, Soares e Assunção (2012) estudam o legado colonial sob instituições; Musacchio, Martinez e Viarengo (2014) focam-se na economia política da educação após a proclamação da República. Por sua vez, Carvalho Filho e Colistete (2010) relacionam o ciclo da expansão cafeeira, a imigração europeia de trabalhadores rurais e o estabelecimento de instituições de instrução educacional no estado de São Paulo.

O estudo de Rocha, Ferraz e Soares (2017) analisa o caso dos núcleos coloniais em São Paulo e sugere evidências do impacto do capital humano dos imigrantes no desenvolvimento de longo prazo dessas regiões. Carvalho Filho e Monasterio (2012), por sua vez, exploram o estabelecimento de colônias patrocinadas pelo governo no Rio Grande do Sul e encontram efeitos persistentes sobre variáveis econômicas atuais. Finalmente, Souza (2017) mostra que o impacto das escolas fundadas por imigrantes alemães em São Paulo se dissipou ao longo do tempo, não deixando sinais nas *proxies* locais de capital humano.

Indiretamente, este trabalho também é relacionado por ampla literatura sobre os resultados educacionais dos imigrantes; em especial, a transmissão intergeracional de escolaridade; incluindo-se Algan e Cahuc (2010), Card, DiNardo e Estes (1998), Borjas, Bronars e Trejo (1992) e Borjas (1995).

De acordo com nossas estimações, caso a imigração não ibérica não tivesse ocorrido no Brasil, a renda *per capita* seria entre 12,6% a 17% menor que a observada. Essa estimativa é obtida mediante modelos econométricos elaborados a partir de regressões com dados individuais e municipais, que simulam a ausência de descendentes de tais imigrantes no Brasil.

Obviamente, toda estimativa contrafactual tem de ser entendida em seus próprios limites. Aqui, apesar de tentar-se mapear alguns dos mecanismos envolvidos, não se consideram efeitos de equilíbrio geral da imigração ou alterações institucionais que possam ter efeitos no longo prazo. Além disso, está implícito que a produtividade do trabalhador é resultado de suas características individuais – com os controles locais – e que efeitos diretamente setoriais não são relevantes.

2 FONTES HISTÓRICAS

O trabalho baseia-se em três fontes primárias ainda pouco exploradas, descritas a seguir.

1. Arquivo da Hospedaria dos Imigrantes de São Paulo – Essa base engloba 1.574.107 registros de imigrantes, que passaram pelo local entre 1882 e 1958. Há registros não somente de estrangeiros, mas também de brasileiros vindos de outros estados. Livros de registro das antigas hospedarias de imigrantes (1882 a 1958) e um conjunto de mais de 150 livros de matrícula das hospedarias de Bom Retiro (1882 a 1887) e Brás (1887 a 1958) contêm informações sobre a passagem de estrangeiros por essas instituições. No tocante aos possíveis dados encontrados, destacam-se: nome, idade, nacionalidade, data de entrada na hospedaria, profissão, parentesco e estado civil.³
2. Arquivo Público do Rio de Janeiro: 208.824 observações – Contém os registros de entrada de estrangeiros no Brasil no porto do Rio de Janeiro, com informações abrangendo o período 1875-1910. Nessas informações, é possível encontrar o nome do imigrante, o nome do navio, a data de entrada, a nacionalidade, o nível de instrução, a religião, a profissão, o estado civil, o local de procedência e o local de destino do imigrante no Brasil. A escolha das relações de passageiros recair sobre o Porto do Rio de Janeiro deveu-se a esse porto ser, na primeira fase da grande leva de imigração (de 1875 a 1910), a principal porta de entrada de imigrantes.
3. Arquivo Público do Espírito Santo: 43.293 observações – O Projeto Imigrantes do Espírito Santo fornece listas contendo os nomes dos imigrantes que entraram no Espírito Santo durante os séculos XIX e XX. Até 10 de março de 2014, foram indexados 54.155 imigrantes que entraram no Espírito Santo – sendo 46.885 no século XIX e 7.146 no século XX –, provenientes dos cinco continentes. Apesar da base cobrir mais de 73 nacionalidades dos imigrantes,

3. Cabe observar que, como esses registros estão transcritos literalmente como aparecem nos livros, é comum que a grafia de nomes e sobrenomes possa não condizer com a forma correta. Também cabe notar que os registros eram feitos em nome dos “chefes de família” – via de regra, o parente do sexo masculino, em idade produtiva, mais velho.

neste trabalho tem-se apenas os dados referentes aos italianos e alemães, que constituem 75% de todos os imigrantes.⁴

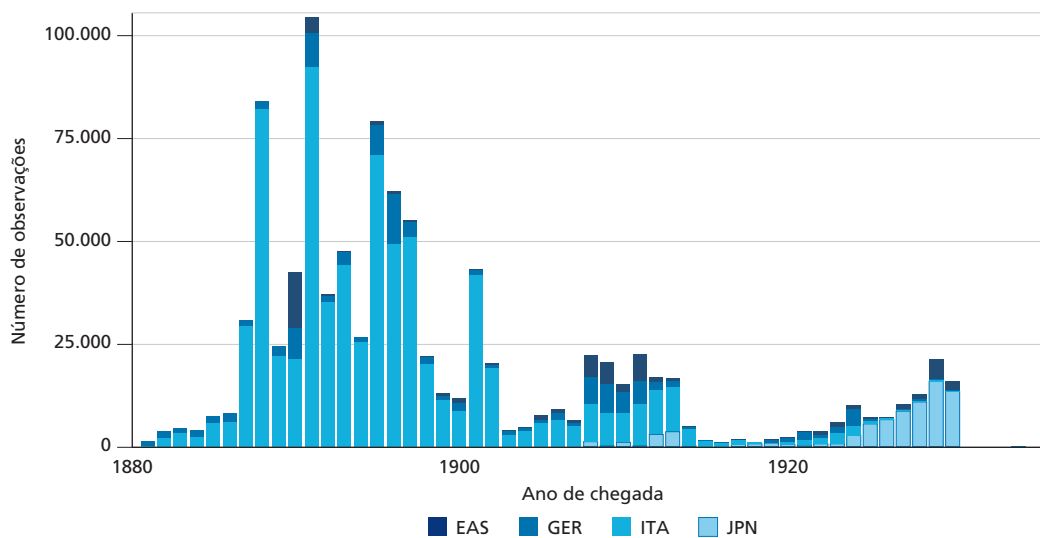
2.1 Descrição da base

O gráfico 1 mostra a chegada de imigrantes não ibéricos por ano.⁵ A comparação das ondas de imigração da nossa base com os dados de entrada apresentados por Levy (1974) é aparentemente semelhante.⁶

Como se vê, a maior parte dos registros de entrada das nossas bases são anteriores ao início da Primeira Guerra Mundial, e os italianos são predominantes. Após o fim desse conflito, a nova onda é marcadamente formada por imigrantes japoneses. O gráfico 2 mostra a distribuição dos registros da base por grupos nacionais, incluindo-se os ibéricos e a migração interna de brasileiros que passaram pela Hospedaria de Imigrantes de São Paulo.

GRÁFICO 1

Ano de entradas dos imigrantes por grupos ancestrais selecionados (1880-1937)



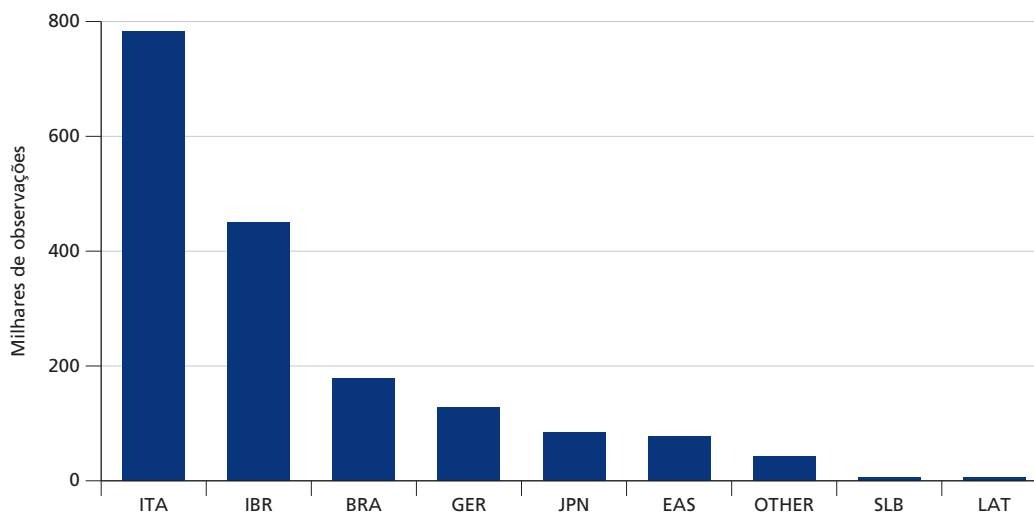
4. Atualmente, as demais nacionalidades encontram-se em processamento.

5. Há um problema no banco de dados do Arquivo Nacional. Algumas datas de chegada referentes ao século XIX foram convertidas para 1ª de janeiro de 1900. Provavelmente, trata-se de um problema causado pelo uso do Microsoft Excel. Esse *software* é incapaz de lidar com datas anteriores ao século XX. Nossos resultados neste trabalho, felizmente, independem da precisão da variável de data de chegada.

6. Em versões posteriores deste texto para discussão, será feita uma comparação mais rigorosa das bases com outras fontes de entradas de imigrantes.

GRÁFICO 2

Principais grupos nacionais nos registros dos imigrantes (1875-1958)



Fonte: Cálculos dos autores com base nos registros da Hospedaria de Imigrantes de São Paulo, do Arquivo Público do Espírito Santo e do Arquivo Nacional do Rio de Janeiro.
Obs.: BRA: brasileiros migrantes; EAS: Europa Oriental; IBR: espanhóis e ibéricos; GER: germânicos; ITA: italianos; JPN: japoneses; LAT: países latino-americanos; SLB: sírios e libaneses; OTHER: todos os demais países.

3 ENUMERAMENTO DOS IMIGRANTES

3.1 Medida de enumeramento

Como medir o capital humano por grupos de imigrantes? Utilizou-se o método proposto por Mokyr (1983) e desenvolvido por A'Hearn, Baten e Crayen (2009). Este se baseia na tendência dos indivíduos com baixa capacidade quantitativa a arredondar o último dígito de valores em 5 ou 0. O índice de Whipple mede essa tendência e é definido pelo quíntuplo da razão da soma dos indivíduos, entre 23 e 62 anos, que informam ter idades terminadas entre 0 e 5 pelo total de pessoas entre aquelas idades. Assim, o índice varia entre 100 – não haveria preferência por arredondamento – e 500:⁷

$$Whipple = 5 * \frac{\sum(n_{25}+n_{30}+\dots+n_{25}+n_{60})}{\sum_{i=23}^{62}(n_i)} \quad (1)$$

Em que n_i = número de pessoas com a idade i .

7. Ver A'Hearn, Delfino e Nuvolari (2016), no tocante a críticas quanto ao uso do arredondamento de idades como *proxy* de capital humano.

Para facilitar a interpretação do índice de Whipple, Crayen e Baten (2010) propuseram o índice ABCC, que é uma transformação linear do índice de Whipple. O ABCC varia entre 0 e 100, e quanto maior o valor, maior o enquadramento do grupo selecionado:

$$Whipple \geq 100 \Rightarrow ABCC = \left(1 - \frac{Whipple - 100}{400}\right) * 100 \quad (2)$$

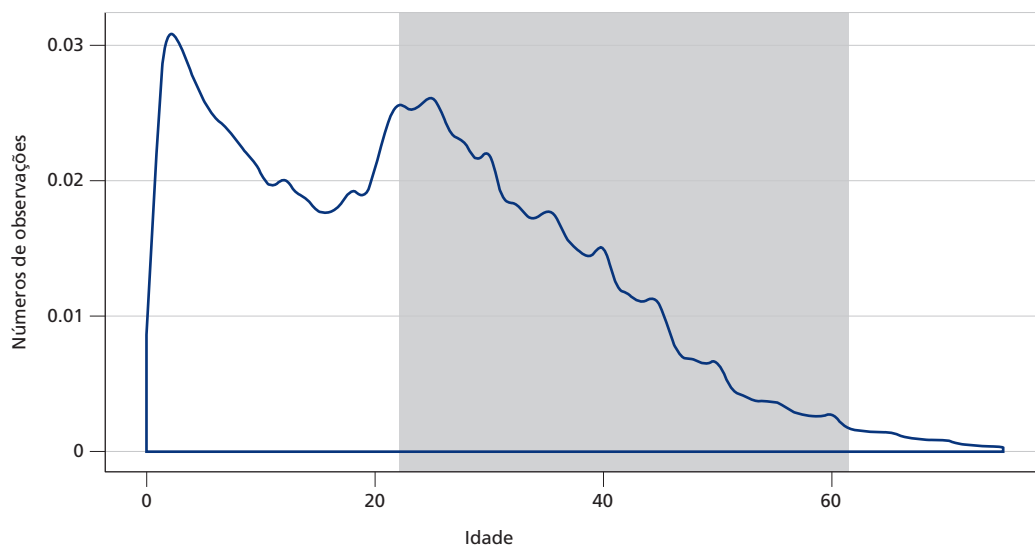
$$Whipple < 100 \Rightarrow ABCC = 100$$

3.2 Resultados do ABCC para a amostra

Stolz, Baten e Botelho (2013) estimaram o ABCC dos brasileiros e dos imigrantes, a partir de amostra de dados obtida no Arquivo Nacional do Rio de Janeiro. Nós seguimos o mesmo procedimento, utilizando nosso banco de dados. Este foi limitado apenas aos indivíduos maiores de 23 e menores de 62 anos, bem como os que chegaram ao Brasil entre 1840 e 1920. Com isso, a amostra passou a ter 774.625 observações individuais.

A distribuição de idades está no gráfico 3. A faixa de idade considerada no índice ABCC é representada pela faixa sombreada. Como se vê, até mesmo com as restrições na amostra, cobre-se parte considerável da distribuição. Vale notar que o fenômeno do arredondamento em dígitos já pode ser percebido no gráfico, nos pequenos picos nas idades terminadas em 0 e 5.

GRÁFICO 3
Distribuição das idades dos imigrantes (1840-1920)



Fonte: Cálculos dos autores com base nos registros da Hospedaria de Imigrantes de São Paulo, do Arquivo Público do Espírito Santo e do Arquivo Nacional do Rio de Janeiro.
Obs.: A área sombreada indica a faixa etária que é utilizada para o cálculo dos índices de Whipple e ABCC.

Os resultados do cálculo ABCC por coorte para os principais grupos da amostra estão representados na tabela 1, a seguir. Limitou-se à apresentação dos resultados aos pares de grupos ancestrais-coortes que tivessem ao menos cem observações. As lacunas na tabela são assim explicadas. É evidente a tendência de incremento do ABCC. A exceção é a coorte entre 1910 e 1920, em que há pequena queda. Os brasileiros, migrantes internos, têm valores baixos, conforme esperados. Em todas as coortes, exceto a de 1850, o valor do ABCC dos oriundos da Europa Oriental foi menor que o dos ibéricos, em linha com Baten, Szołtysek e Campestrini (2017).

Os japoneses passaram o maior valor do índice ABCC em todas as coortes. Os italianos ficam, em geral, na segunda posição. O enumeramento dos germânicos mostra-se maior que o dos migrantes brasileiros, mas permanece atrás do dos ibéricos na coorte de 1870, 1890 e 1910. Os sírio-libaneses possuíam níveis muito baixos do índice ABCC nas coortes em que o número de observações é superior ao mínimo.

TABELA 1
Índice ABCC para toda a amostra por grupos nacionais e coorte (1840-1910)

	Pais/região	1840	1850	1860	1870	1880	1890	1900	1910
1	BRA	69.5	65.5	70.2	78.3	83.6	87.4	94.7	
2	EAS	73.8	87.9	96.1	93.8	94.5	95.2	100.0	99.0
3	GER	90.9	95.8	96.6	94.2	98.9	99.1	100.0	92.6
4	IBR	88.9	93.8	95.8	98.0	97.8	98.3	100.0	99.4
5	ITA	95.3	97.4	97.3	97.9	98.3	100.0	100.0	98.7
6	JPN			98.7	100.0	99.6	100.0	100.0	100.0
7	LAT	81.9	80.8	85.7	94.7	100.0			
8	SLB	36.5	36.5	49.6	70.1				

Obs.: BRA: brasileiros migrantes; EAS: Europa Oriental; IBR: espanhóis e ibéricos; GER: germânicos; ITA: italianos; JPN: japoneses; LAT: países latino-americanos; SLB: sírios e libaneses; OTHER: todos os demais países.

3.3 ABCC, bases de dados e a seleção de imigrantes

Os resultados de enumeramento dos imigrantes em nossos dados são, em geral, maiores que os observados na amostra do Rio de Janeiro estudada por Stolz, Baten e Botelho (2013). Os autores encontram valor médio de 94,1 para a coorte de 1900 dos imigrantes. Por sua vez, na nossa amostra, chega-se a um ABCC de 98,3. Essa diferença de pouco mais de 4,2 pontos pode parecer desprezível, mas é a mesma que separava, por exemplo, o enumeramento dos argentinos ou norte-americanos dos hondurenhos na mesma coorte de 1900. A principal diferença da nossa base em relação à amostra de Stolz, Baten e Botelho (2013) é que na nossa há maior peso dos que passaram pela Hospedaria dos Imigrantes de São Paulo. Logo, podemos concluir que estes tinham, em média, maior enumeramento que os da base do Arquivo Nacional do Rio de Janeiro.

Vale a pena comparar os imigrantes com as populações de seus países de origem, com o objetivo de identificar em que medida houve seleção – positiva ou negativa – dos que para cá vieram. O sítio Clio Infra⁸ reúne estimativas do índice ABCC para muitos países do mundo, desde 1500, com periodicidade irregular, por coorte. Na tabela 2, comparou-se o enumeramento dos imigrantes e da população de origem. Optou-se por analisar as coortes de 1870 a 1900. Como se vê, não há padrão claro. No caso de húngaros e romenos, parece ter havido seleção negativa – ou seja, vieram aqueles com menor enumeramento. Por sua vez, no tocante aos poloneses e aos portugueses, os dados sugerem seleção positiva. Nos exemplos dos demais países, as diferenças são menos pronunciadas.

4 NOVAS ESTIMATIVAS PARA O IMPACTO DA IMIGRAÇÃO

Para Stolz, Baten e Botelho (2013), o impacto da imigração para o Brasil foi de US\$ 75 *per capita* por ano. Essa estimativa foi baseada em regressão do PIB *per capita* de 2000 sobre o índice ABCC de 1900, por um conjunto de controles geográficos e institucionais por país. Segundo os cálculos dos autores, os imigrantes teriam um ABCC de 94,1 e os nascidos no Brasil, de 89,7, para a coorte de 1900. Esse valor contrafactual – ou seja, qual seria o ABCC sem a chegada dos imigrantes – é inserido na regressão estimada por país para chegar ao valor citado. Em termos percentuais, o efeito da imigração seria de cerca de 1% do PIB *per capita*.

TABELA 2
Índice ABCC para imigrantes e para toda a amostra por grupos nacionais e coorte (1870-1900)

País	Imigrantes	População
Alemão	98.3	99.9
Húngaro	96.7	99.7
Italiano	99.0	99.8
Japonês	99.9	99.2
Poloneso	94.4	93.7
Português	98.5	93.6
Romeno	95.2	99.8
Russo	97.0	97.0
Espanhol	98.1	99,4

Fonte: Microdados desta pesquisa e Clio Infra – disponível em: <<https://bit.ly/2uP1vGU>>.

Obs.: 1. A lista de países é incompleta, pois foram omitidos os países da amostra em que não houvesse ao menos cem observações por coorte.

2. Média do índice ABCC para as coortes do período 1870-1900, no tocante à amostra e aos países de origem.

8. Clio Infra. Clio Infra Projec. Disponível em: <<https://bit.ly/2uP1vGU>>.

Para rever essa estimativa, nesta seção, criar-se-ão outros contrafatuais. Na seção anterior, mostrou-se que nossa estimativa do enumeramento dos imigrantes é sensivelmente maior que a obtida por Stolz, Baten e Botelho (2013). Agora, em vez de estimar o impacto da imigração com base em regressão por país, será feita uma regressão contrafactual, a partir de dados contemporâneos individuais e por município.

Em ambas as estratégias de construção do contrafactual, utilizou-se o algoritmo baseado em Monasterio (2017), com o objetivo de classificar sobrenomes dos indivíduos contemporâneos de acordo com a ancestralidade. O procedimento tem início com a construção de uma base de dados históricos, que associa sobrenomes a ancestralidades únicas. No artigo citado anteriormente, foram utilizadas diversas fontes: uma amostra obtida por *web scraping* no Museu da Imigração (2015); informações sobre o tráfico interprovincial de escravos (Fragoso e Ferreira, 2002); microdados dos censos históricos norte-americanos;⁹ entre muitas outras. A base foi agregada em pares únicos de sobrenome-ancestralidade. Essa base tem um duplo papel.

1. Servir para o *fuzzy matching* com a base de sobrenomes observados.
2. Calibrar o algoritmo de *machine learning*, baseado em Cavnar e Trenkle (1994), que identifica as nacionalidades com base em um padrão linguístico. Existem 464.613 sobrenomes distintos entre os trabalhadores na base da Relação Anual de Informações Sociais (RAIS) 2016, do Ministério do Trabalho (MTE), mas alguns concentram grande número de observações. Os Silva, Santos, Oliveira, Souza e Pereira somam 45% das pessoas. Assim, a maior parte dos sobrenomes únicos é constituída, na verdade, por erros de digitação e observada apenas uma ou duas vezes. Ocorreu *match* exato em 40.496 sobrenomes; o procedimento de *fuzzy matching* permitiu encontrar 336.258 sobrenomes; os demais (87.859) foram classificados com o auxílio do algoritmo de Cavnar e Trenkle (1994).

A classificação de sobrenomes tem limitações. Além dos problemas habituais de acurácia do algoritmo de classificação, o problema principal é a perda da informação matrilinear dos sobrenomes. No padrão de escolha de nomes no Brasil, o sobrenome de uma avó estrangeira, por exemplo, estará perdido já no seu neto. Contudo, caso haja endogamia entre os grupos de ancestralidade, ainda que seu sobrenome específico esteja perdido, há chances maiores de que seu neto tenha ancestrais do mesmo grupo.¹⁰

9. *Integrated Public Use Microdata Series (IPUMS). North Atlantic Population Project.* Disponível em: <<https://bit.ly/2O6n6U7>>.

10. Segundo Brasil (1916), 77% dos filhos de estrangeiros nascidos entre 1908 e 1912 em capitais selecionadas do Brasil tinham o cônjuge igualmente estrangeiro.

Nesta aplicação, de qualquer forma, o problema persiste, pois, por simplificação, será usado apenas o último sobrenome do indivíduo.

Para a classificação do perfil de grupos ancestrais de cada município, utilizaram-se as seguintes bases de dados contemporâneas: RAIS (Brasil, 2016), 67.144.598 observações; Cadastro Único – CadÚnico (Brasil, 2015), 80.957.412 observações; e a base de sócios da RFB,¹¹ 17.780.860 observações. A combinação desses dados cobre amplo espectro social do Brasil. O CadÚnico contém informações sobre os mais pobres, a maioria no setor informal da economia. A RAIS, como foi dito, as faixas intermediárias, os empregados formais. A lista de sócios, por fim, abrange não apenas os representantes dos estratos mais altos, mas também os microempreendedores individuais (MEIs).¹²

Os dados dessas três bases foram classificados conforme a ancestralidade, de acordo com o procedimento apresentado na seção anterior e agrupados por municípios. Utilizou-se a informação de cor/raça disponível na RAIS para o caso de indivíduos registrados como índio (NAT), *pardo* (MXD) e *preto* (BLK). Por sua vez, para os *brancos* e os *amarelos*, empregou-se o algoritmo de classificação de sobrenomes.

4.1 Regressão municipal

Regrediu-se a renda *per capita* domiciliar mensal observada em 2010 sobre um conjunto amplo de variáveis em nível municipal.¹³ A inovação é a inclusão de variáveis do lado direito das regressões que representem o impacto de longo prazo da imigração na população. Esse efeito será medido como as participações dos grupos ancestrais de sobrenome na base composta (CadÚnico, RAIS e sócios) por município.

Além dessas variáveis relacionadas com a ancestralidade, foram utilizadas *proxies* para o nível educacional do município, suas instituições, a população, *dummies* estaduais e um grupo exaustivo de variáveis geográficas e climáticas (altura, distância da capital do estado, pluviosidade, temperatura e aptidão dos solos).

11. Disponível em: <<https://bit.ly/2Lpc6CF>>.

12. A base da RFB não informa o município de residência, apenas o Cadastro Nacional da Pessoa Jurídica (CNPJ) da empresa da qual o indivíduo é sócio. Fez-se um cruzamento dessa base com a Relação Anual de Informações Sociais (RAIS) Estabelecimentos e atribuiu-se ao sócio o mesmo município da sede da sua firma.

13. As variáveis utilizadas e suas fontes constam do apêndice A.

A tabela 3 apresenta os resultados. A participação do grupo com ancestralidade ibérica é a variável de referência. Tomando-se a especificação mais completa e o grupo IBR como referência, tem-se que as participações dos grupos NAT, MXD, BLK, SLB estão associadas a reduções da renda *per capita* municipal; por sua vez, os grupos GER, ITA e JPN têm relação com aumentos desse valor. Podem causar surpresa inicial os valores negativos associados à participação de indivíduos com sobrenomes da Europa Oriental. Vale lembrar, contudo, que o enumeramento desses imigrantes foi menor que o dos ibéricos na maior parte das coortes. Por seu turno, o valor absoluto dos estimadores associados à participação dos japoneses é realmente impressionante.

Qual o significado preciso dos coeficientes associados aos percentuais de cada grupo ancestral de sobrenomes? Quais os canais exatos que levam a essa associação? Como os coeficientes continuam significativos até mesmo após a inserção das variáveis educacionais e institucionais, existem duas possibilidades: ou as *proxies* não captam totalmente os efeitos dessas dimensões; ou existem canais não contemplados na análise. Será evitada uma resposta definitiva neste estudo. A próxima subseção também busca avançar um pouco nessa questão, mas – no estado atual da pesquisa e diante deste objetivo – seria imprudente defender um mecanismo causal de forma categórica.

4.1.1 Análise contrafactual: dados municipais

Para estimar o impacto da imigração, criou-se um contrafactual em que foram zeradas as parcelas da população com ancestralidade não ibérica por município. Recalculadas as participações para os grupos étnicos restantes, esses valores foram inseridos na especificação mais completa da tabela 3, e então foi recalculada a renda *per capita* brasileira. A comparação entre a renda *per capita* contrafactual (sem imigrantes) e a observada representa o impacto local da imigração não ibérica.

A renda *per capita* observada em 2010 foi igual a R\$ 781,75. Por sua vez, a contrafactual, a partir da equação 4, foi de R\$ 683,18. Ou seja, a renda *per capita* brasileira seria cerca de 12,6% menor que a observada, caso não houvesse imigrantes. Isto é, trata-se de efeito bem maior que o 1% estimado por Stolz, Baten e Botelho (2013). Note-se que esse resultado não é o efeito trivial da retirada de indivíduos com salários mais altos da amostra. Afinal, a especificação preferida foi feita com base nas parcelas de cada grupo, e foram mantidos todos os controles municipais: variáveis institucionais, geográficas, população municipal e *dummies* por Unidade da Federação (UF).

TABELA 3
Regressões com dados municipais

	Renda <i>per capita</i> média por município			
	Modelo 1	Modelo 2	Modelo 3	Modelo 4
NAT (%)	-510,8 ***	-179,2 ***	-182,0 ***	-219,1 ***
	-47,9	-40,5	-43,7	-41,6
MXD (%)	-418,6 ***	-82,4 ***	-88,4 ***	-164,5 ***
	-11,2	-13,1	-13,2	-13,6
BLK (%)	-54,0	-155,7 ***	-120,9 ***	-125,5 ***
	-42,3	-37,6	-37,8	-37,3
GER (%)	201,9 ***	220,6 ***	231,6 ***	277,7 ***
	-29,4	-26,2	-27	-26,1
EAS (%)	-1.034,5***	-657,2 ***	-591,2 ***	-461,6 ***
	-89,3	-70,5	-71,3	-67,9
ITA (%)	609,4 ***	550,3 ***	552,8 ***	632,6 ***
	-30,8	-26,8	-27,8	-28,3
SLB (%)	3.671,4 ***	418,4	44	-202,3
	-603,8	-475,9	-495,3	-464,1
JPN (%)	8.898,1 ***	4.692,5 ***	3.379,4 ***	2.375,5 ***
	-604,6	-523,6	-509,2	-480,5
Educação		-12,2 ***	-10,0 ***	-7,8 ***
		-0,3	-0,3	-0,3
Governança			23,8 ***	12,6 ***
			-2,4	-2,3
Gini da distribuição da terra			-146,2 ***	-151,9 ***
			-22,1	-20,9
Acesso à justiça			44,5 ***	12,3 ***
			-2,1	-2,4
Persistência política			3,8	3,7
			-3,2	-3
Log (população)				48,9 ***
				-2,1
Dummies estaduais		Y	Y	Y
Controles geográficos				Y
Observações	5.505	5.505	4.900	4.898
R2 ajustado	0,6	0,8	0,8	0,8

Elaboração dos autores, com base nos dados da pesquisa.
Obs.: *p<0,01; **p<0,05; ***p<0,01.

4.2 Regressão individual

Segue-se aqui a mesma estratégia empírica da subseção anterior, mas com dados individuais dos trabalhadores. O problema dessa escolha é que apenas os ocupados no setor formal são examinados. Por sua vez, a vantagem do método consiste no fato de que podem ser utilizadas *dummies* municipais, que capturam os possíveis efeitos geográficos, econômicos ou institucionais de caráter local.

Mesmo contendo mais de 67 milhões de observações, a base da RAIS cobre apenas os trabalhadores formais brasileiros. Limitou-se a amostra aos trabalhadores no setor privado, brasileiros, entre 23 e 60 anos e que trabalhassem ao menos quarenta horas por semana. Com isso, restaram 23.899.675 registros de trabalhadores com Cadastro de Pessoas Físicas (CPF) único. Apenas com base nesses dados, os grupos mais presentes são IBR (48%), MXD (35%) e BLK (5,4%). As participações dos não ibéricos são as seguintes: EAS: 0,8 %; GER: 2,0%; ITA: 7,2%; e JPN: 0,6 %.

Devido a restrições computacionais, fez-se uma amostra aleatória de um milhão de observações. E, no modelo mais completo, com mais de 5.500 *dummies* municipais, a amostra foi reduzida para 100 mil observações para reduzir o tempo de processamento.

Em seguida, seguindo-se a lógica da subseção anterior, os parâmetros dessa regressão servem para estimar os salários contrafatuais em uma base sem os trabalhadores com ancestralidade não ibérica. Mais uma vez, a diferença entre os valores observados e simulados dos salários por trabalhador representa o impacto da imigração não ibérica.

A tabela 4 apresenta os resultados individuais. As variáveis de interesse são as *dummies* de ancestralidade de sobrenomes, mas os controles tiveram sinal e significância esperados. Omitiu-se dos resultados os coeficientes associados ao sexo, ao grau de instrução, à idade, à idade ao quadrado, à UF e a *dummies* municipais – conforme o caso.

Na especificação mais completa, com controles individuais e *dummies* municipais, os coeficientes associados à variável de ancestralidade medida pelo sobrenome são estatística e substantivamente significativos. O grupo de referência são os indivíduos com sobrenome IBR. Conforme esperado, os grupos BLK, MXD e NAT têm relação com os salários mais baixos. Por sua vez, os indivíduos com ancestralidade de sobrenomes JPN auferem salários 17,8% mais elevados na especificação preferida – i.e., a mais

completa. Os coeficientes dos possuidores dos demais sobrenomes não ibéricos também são positivos, mas com valores mais baixos. Surpreende que os trabalhadores com sobrenome sírio-libaneses tenham salários 10% maiores que os ibéricos e até mesmo maiores que os GER, os EAS e os ITA. Não é clara a razão para tal fenômeno.

TABELA 4
Regressão individual

	Log do salário por hora		
	Modelo 1	Modelo 2	Modelo 3
BLK	-0,095***	-0,067***	-0,094***
	-0,002	-0,002	-0,007
EAS	0,078***	0,072***	0,053***
	-0,006	-0,006	-0,018
GER	0,084***	0,077***	0,087***
	-0,003	-0,003	-0,01
ITA	0,090***	0,065***	0,072***
	-0,002	-0,002	-0,006
JPN	0,223***	0,179***	0,164***
	-0,006	-0,006	-0,02
MXD	-0,137***	-0,050***	-0,081***
	-0,001	-0,001	-0,004
NAT	-0,103***	-0,043***	-0,067*
	-0,011	-0,011	-0,036
SLB	0,086***	0,085***	0,100***
	-0,007	-0,006	-0,02
<i>Dummies estaduais</i>		Sim	
<i>Dummies municipais</i>			Sim
Observações	997.682	997.682	99.768
R2	0,374	0,407	0,454
R2 ajustado	0,374	0,407	0,432

Elaboração dos autores, com base nos dados da pesquisa.

Obs.: *p<0,01; **p<0,05; ***p<0,01.

4.3 Análise contrafactual: dados individuais

A criação do contrafactual foi feita pela substituição da variável de ancestralidade de sobrenomes dos descendentes de imigrantes não ibéricos (EAS, GER, ITA e JPN) por ibéricos (IBR). O salário médio por hora da amostra observada é de R\$ 58,04. Por sua vez, os salários contrafatuais foram de R\$ 48,29. Isso sugere que, sem a imigração não ibérica, os salários seriam cerca de 17% mais baixos.

Isso pode ser considerado o limite inferior da estimativa, pois em vez de criar um contrafactual que reproduzisse as participações observadas com outros grupos ancestrais (MXD, BLK e NAT), optou-se apenas pela simples substituição dos portadores de sobrenomes não ibéricos por ibéricos.

5 DISCUSSÃO

Supondo-se que as estimativas estejam corretas, quais seriam os canais que fizeram com que a imigração tenha tal impacto no crescimento econômico? O primeiro candidato é, claro, o capital humano, em seu sentido mais amplo. Lopes (2017) mostrou que crianças de 8 anos com ancestralidade de sobrenome não ibérica obtêm desempenho sensivelmente superior às brancas com sobrenome ibérico medido por testes padronizados, até mesmo com controles para *background* socioeconômico dos pais e efeitos fixos por sala de aula. Em termos substantivos, no caso daqueles estudantes com ancestrais japoneses, o efeito equivale a um ano a mais de escolaridade na disciplina de matemática. Ao que parece, há transmissão intergeracional do capital humano que não é capturada adequadamente pelas *proxies* de escolaridade usadas nas regressões.

Ehrl e Monasterio (2017), por sua vez, analisaram o impacto de longo prazo da imigração na produtividade via aumento da diversidade local de habilidades dos trabalhadores do Rio Grande do Sul. Os autores encontraram efeitos bastante robustos. Porém, sua abordagem não é capaz de explicar integralmente os resultados aqui encontrados. Se a diversidade fosse o canal, os controles locais nas regressões individuais deveriam capturar todo o efeito das variáveis de sobrenome. A mesma lógica indica que explicações baseadas em capital social ou instituições locais também são insuficientes.

Quais outros canais tornariam as estimativas enganosas? Se os salários mais altos forem explicados integralmente por discriminação no nível da firma em favor dos possuidores de sobrenomes não ibéricos, esses resultados econométricos seriam obtidos nas regressões em nível individual. Porém, nesse caso, seria necessário explicar o porquê de os municípios com maiores participações de não ibéricos também estarem associados a rendas *per capita* mais altas.

6 CONSIDERAÇÕES FINAIS

Este trabalho estimou o impacto da imigração não ibérica para o Brasil, com base em microdados históricos e contemporâneos. Enquanto estudo anterior de Stolz, Baten e Botelho (2013) calculou um impacto de cerca de 1%, aqui se estimou que a renda *per capita* do brasileiro seria entre 12,6% e 17% mais baixa sem a imigração.

A partir dos registros de entrada de mais de 1,7 milhão de imigrantes, calculou-se o enquadramento por país de origem e mostrou-se que, em geral, este era bem maior que o dos brasileiros e também maior do que indicavam as estimativas anteriores.

A elaboração do contrafactual – isto é, o Brasil sem descendentes de imigrantes não ibéricos – foi feita a partir de regressões de salários e renda *per capita* em nível individual e municipal sobre variáveis de ancestralidade de sobrenomes. Por mais críticas que esse o método possa sofrer, muitas destas pertinentes, este permite que se tenha um entendimento mais preciso dos efeitos de longo prazo da imigração para o Brasil.

Os próximos passos da pesquisa em curso dirigem-se em três direções: *i)* aumentar ainda mais as bases de dados históricas pela inclusão de novas fontes; *ii)* incrementar o algoritmo para classificar sobrenomes com base em origens sub-nacionais – por exemplo: oriundos do norte/sul da Itália, Okinawa, entre outras; e *iii)* buscar identificar os nexos causais que fizeram com que a imigração não ibérica tivesse os efeitos aqui identificados.

REFERÊNCIAS

A'HEARN, B.; BATEN, J.; CRAYEN, D. Quantifying quantitative literacy: age heaping and the history of human capital. **The Journal of Economic History**, v. 69, n. 3, p. 783-808, 2009.

A'HEARN, B.; DELFINO, A.; NUVOLARI, A. **Rethinking age-heaping**: a cautionary tale from nineteenth century Italy. Pisa: Institute of Economics/Scuola Superiore Sant'Anna, 2016. (Laboratory of Economics and Management Working Paper Series, n. 2016/35).

ALESINA, A.; GIULIANO, P.; NUNN, N. On the origins of gender roles: women and the plough. **The Quarterly Journal of Economics**, v. 128, n. 2, p. 469-530, 2013.

ALGAN, Y.; CAHUC, P. Inherited trust and growth. **American Economic Review**, v. 100, n. 5, p. 2060-2092, Dec. 2010.

BISIN, A.; VERDIER, T. “Beyond the melting pot”: cultural transmission, marriage, and the evolution of ethnic and religious traits. **The Quarterly Journal of Economics**, v. 115, n. 3, p. 955-988, 2000.

BATEN, J.; SZOŁTYSEK, M.; CAMPESTRINI, M. “Girl Power” in Eastern Europe? The human capital development of Central-Eastern and Eastern Europe in the seventeenth to nineteenth centuries and its determinants. **European Review of Economic History**, v. 21, n. 1, p. 29-63, 2017.

BORJAS, G. J. Ethnicity, neighborhoods, and human-capital externalities. **The American Economic Review**, v. 85, n. 3, p. 365-390, 1995.

BORJAS, G. J.; BRONARS, S. G.; TREJO, S. J. Self-selection and internal migration in the United States. **Journal of Urban Economics**, v. 32, n. 2, p. 159-185, 1992.

BRASIL. Ministério da Agricultura, Indústria e Comércio. Diretoria-Geral de Estatística. **Anuário Estatístico do Brasil, 1908-1912**. Rio de Janeiro: Tipografia da Estatística, 1916.

_____. Ministério da Agricultura, Indústria e Comércio. Diretoria-Geral de Estatística. **Recenseamento do Brasil**. Rio de Janeiro: Tipografia da Estatística, 1929.

_____. Ministério do Desenvolvimento Social. **Cadastro Único**. Brasília: MDS, 2015.

_____. Ministério do Trabalho. **RAIS Identificada**. Brasília: MTB, 2016.

CARD, D.; DINARDO, J.; ESTES, E. **The more things change**: immigrants and the children of immigrants in the 1940s, the 1970s, and the 1990s. Cambridge, MA: NBER, 1998. (NBER Working Paper Series, n. 6519).

CARVALHO FILHO, I.; COLISTETE, R. P. **Education performance**: was it all determined 100 years ago? Evidence from São Paulo, Brazil. 2010. (MPRA Paper, n. 24494). Disponível em: <goo.gl/E49AKB>.

CARVALHO FILHO, I.; MONASTERIO, L. Immigration and the origins of regional inequality: government-sponsored european migration to southern Brazil before World War I. **Regional Science and Urban Economics**, v. 42, n. 5, p. 794-807, 2012.

CAVNAR, W. B.; TRENKLE, J. M. N-gram-based text categorization. **Ann Arbor MI**, v. 48113, n. 2, p. 161-175, 1994.

CHAUDHARY, L. *et al.* Big BRICs, weak foundations: the beginning of public elementary education in Brazil, Russia, India, and China. **Explorations in Economic History**, v. 49, n. 2, p. 221-240, 2012.

CRAYEN, D.; BATEN, J. Global trends in numeracy 1820-1949 and its implications for long-term growth. **Explorations in Economic History**, v. 47, n. 1, p. 82-99, 2010.

EHRL, P.; MONASTERIO, L. Inherited cultural diversity and wages in Brazil. *In: AFD-WORLD BANK MIGRATION AND DEVELOPMENT CONFERENCE*, 10., 2017, Clermont-Ferrand. **Séminaires...** Clermont-Ferrand: IADT, 2017.

FERNÁNDEZ, R.; FOGLI, A. Culture: an empirical investigation of beliefs, work, and fertility. **American Economic Journal: macroeconomics**, v. 1, n. 1, p. 146-77, Jan. 2009.

FIGLIO, D. *et al.* **Long-term orientation and educational performance**. Cambridge, MA: NBER, 2016. (NBER Working Paper Series, n. 22541).

FRAGOSO, J. L. R.; FERREIRA, R. G. **Tráfico de escravos e relações comerciais no Sudeste do Brasil: primeira metade do século XIX**. Rio de Janeiro: Ipea, 2002. (Final report and database of the Ipea/LIPHIS project).

GUISSO, L.; SAPIENZA, P.; ZINGALES, L. Does culture affect economic outcomes? **Journal of Economic Perspectives**, v. 20, n. 2, p. 23-48, June 2006.

LAGO, L. A. C. **Da escravidão ao trabalho livre: Brasil, 1550-1900**. São Paulo: Companhia das Letras, 2014.

LEVY, M. S. F. O papel da migração internacional na evolução da população brasileira (1872 a 1972). **Revista de Saúde Pública**, v. 8, p. 49-90, 1974.

LOPES, D. A. F. **Culture, institutions and school achievement in Brazil**. 2017. Dissertação (Mestrado) – Universidade Católica de Brasília, Brasília, 2017.

MOKYR, J. **Why Ireland starved: a quantitative and analytical history of the Irish economy, 1800-1850**. London: Routledge, 1983.

MONASTERIO, L. Surnames and ancestry in Brazil. **PLOS ONE**, v. 12, n. 5, 2017. Disponível em: <[goo.gl/ZbVQFg](https://doi.org/10.1371/journal.pone.0172888)>.

MONASTERIO, L.; REIS, E. **Mudanças na concentração espacial das ocupações nas atividades manufatureiras no Brasil: 1872-1920**. Rio de Janeiro: Ipea, 2008. (Texto para Discussão, n. 1361).

MUSACCHIO, A.; MARTINEZ, A.; VIARENGO, M. **Colonial institutions, commodity booms, and the diffusion of elementary education in Brazil, 1889-1930**. Cambridge, MA: NBER, 2014. (NBER Working Paper Series, n. 20029).

MUSEU DA IMIGRAÇÃO. **Acerco digital do Museu da Imigração do Estado de São Paulo**. 2015. Disponível em: <<https://bit.ly/2CPHHGx>>.

NARITOMI, J.; SOARES, R. R.; ASSUNÇÃO, J. J. Institutional development and colonial heritage within Brazil. **The Journal of Economic History**, v. 72, n. 2, p. 393-422, 2012.

PEREIRA, L. C. B. Origens étnicas e sociais do empresário paulista. **Revista de Administração de Empresas**, v. 4, n. 11, p. 83-106, 1964.

REIS, E. J.; REIS, E. P. As elites agrárias e a abolição da escravidão no Brasil. **Dados: revista de ciências sociais**, v. 31, n. 3, p. 309-341, 1988.

ROCHA, R.; FERRAZ, C.; SOARES, R. R. Human capital persistence and development. **American Economic Journal: applied economics**, v. 9, n. 4, p. 105-36, 2017.

SØRENSEN, R. J. *et al.* **Culture and school performance: evidence from second generation immigrants to Norway**. Oslo: BI, 23 Oct. 2016. (BI Working Paper).

SOUZA, B. G. W. Immigration and the path dependence of education: the case of german-speakers in São Paulo, Brazil (1840-1920). **The Economic History Review**, v. 71, n. 2, p. 506-539, 2017.

STOLZ, Y.; BATEN, J.; BOTELHO, T. Growth effects of nineteenth-century mass migrations: “Fome Zero” for Brazil? **European Review of Economic History**, v. 17, n. 1, p. 95-121, 2013.

TABELLINI, G. Presidential address institutions and culture. **Journal of the European Economic Association**, v. 6, n. 2-3, p. 255-294, 2008.

VERSIANI, F. R. Imigrantes, trabalho qualificado e industrialização: Rio e São Paulo no início do século. **Revista de Economia Política**, v. 13, n. 4, p. 52, 1993.

APÊNDICE A

FONTES DE DADOS

Dados de imigrantes: dados obtidos em arquivos digitais, *web scraping* e reconhecimento óptico de caracteres (OCR):

- Hospedaria de São Paulo;
- Arquivo Nacional do Rio de Janeiro; e
- Arquivo Público do Espírito Santo.

Variáveis institucionais: dados fornecidos por Rodrigo Soares e utilizados em Naritomi, Soares e Assunção (2012):

- *land Gini*: coeficiente de Gini para a distribuição de terra construído com base no Censo Agrícola do Instituto Brasileiro de Geografia e Estatística (IBGE) de 1996;
- *political persistence*: número de repetições de ao menos um sobrenome de família no nome dos prefeitos nas eleições de 1996, 2000 e 2004;
- *governance*: média simples de indicadores qualitativos sobre práticas da administração municipal, calculados com dados entre 1997 e 2000;
- *access to justice*: média de variáveis binárias que indicam existência de tribunal de pequenas causas, conselho tutelar e comissão de defesa do consumidor em 2001.

Variáveis geográficas: dados obtidos no Ipeadata:¹

- chuvas: estimativas da pluviosidade em cada uma das estações do ano;
- temperatura: estimativas das médias trimestrais de temperatura em centígrados, em cada uma das estações do ano;
- distância à capital estadual em quilômetros; e
- altitude: altura da sede do município.

1. Disponível em: <<https://bit.ly/1kgRmqB>>.

Variáveis sociais e demográficas: dados obtidos no Programa das Nações Unidas para o Desenvolvimento (PNUD):²

- educação: taxa de alfabetização dos maiores de 25; e
- renda *per capita*: renda domiciliar *per capita*.

Bases cadastrais contemporâneas: arquivos digitais de larga escala:

- Brasil (2015) – Cadastro Único (CadÚnico). Arquivos digital de acesso restrito com dados individual dos beneficiados. Fábio Vaz preparou o banco de dados;
- Brasil (2016) – RAIS identificada com os dados dos vínculos dos empregados. Arquivo digital de acesso restrito; e
- Receita Federal do Brasil³ – Dados abertos do *Cadastro Nacional da Pessoa Jurídica* (CNPJ). Álvaro Justen criou um *script* em Python para *download* e limpeza dos dados da RFB. Disponível em: <<https://bit.ly/2L8cAO9>>.

REFERÊNCIA

NARITOMI, J.; SOARES, R. R.; ASSUNÇÃO, J. J. Institutional development and colonial heritage within Brazil. **The Journal of Economic History**, v. 72, n. 2, p. 393-422, 2012.

2. Programa das Nações Unidas para o Desenvolvimento (PNUD). Atlas do Desenvolvimento Humano. Disponível em: <<https://bit.ly/191tPbN>>.

3. Disponível em: <<https://bit.ly/2Lpc6CF>>.

APÊNDICE B

TABULAÇÕES COMPLEMENTARES

TABELA B.1
Distribuição das fontes por grupos de ancestralidade (1840 -1958)

País	Arq. Público – ES	Arq. Nacional – RJ	Hosp. dos Imigrantes -SP
BRA	0	0	178.065
EAS	0	37.997	37.032
GER	4.596	66.844	56.070
IBR	0	40.128	410.769
ITA	36.569	12.698	735.048
JPN	0	167	84.238
LAT	0	3.536	475
OTHER	0	18.072	21.747
SLB	0	4.874	95

Fonte: Dados da pesquisa.

Obs.: BRA: brasileiros migrantes; EAS: Europa Oriental; IBR: espanhóis e ibéricos; GER: germânicos; ITA: italianos; JPN: japoneses; LAT: países latino-americanos; SLB: sírios libaneses; OTHER: todos os demais países.

TABELA B.2
Distribuição das fontes por países (1840-1958)

País	ES	RJ	SP
Albânia	0	0	91
Arábia Saudita	0	4.084	27
Argentina	0	3.088	379
Armênia	0	137	85
Austrália	0	0	3
Áustria	0	31.940	29.378
Bélgica	0	2.177	1.605
Bolívia	0	91	29
Brasil	0	0	178.065
Reino Unido	0	0	1.510
Bulgária	0	72	195
Canadá	0	0	905
Chile	0	280	31
China	0	34	272
Colômbia	0	23	29
Croácia	0	0	15
Cuba	0	0	13
República Tcheca	0	1.163	1.122

(Continua)

(Continuação)

País	ES	RJ	SP
Dinamarca	0	0	857
Holanda	0	0	904
Egito	0	0	39
Equador	0	16	9
Estônia	0	0	346
Finlândia	0	0	35
França	0	1.332	4.233
Alemanha	0	34.904	26.692
Grécia	0	0	4.031
Hungria	0	576	3.209
Índia	0	0	300
Iraque	0	0	1
Irã	0	13	25
Irlanda	0	0	3
Israel	0	46	261
Itália	0	12.698	735.048
Japão	0	167	84.238
Coreia do Sul	0	0	4
Líbano	0	142	64
Letônia	0	0	3.163
Lituânia	0	0	4.740
República de Malta	0	0	324
México	0	92	6
Montenegro	0	28	24
Marrocos	0	0	31
Noruega	0	0	21
Outro	0	7.342	343
Palestina	0	0	9
Paraguai	0	97	2
Peru	0	108	12
Polônia	0	4.684	10.896
Portugal	0	13.145	155.643
Romênia	0	388	1.924
Rússia	0	30.651	7.736
Sérvia	0	233	4
Eslováquia	0	0	5
Eslovênia	0	0	1
Espanha	0	26.983	255.126
Sem país	0	0	1.521
Suécia	0	0	1.414

(Continua)

(Continuação)

País	ES	RJ	SP
Suíça	0	0	2.035
Síria	0	4.732	31
Tunísia	0	2.244	581
Turquia	0	0	1
Ucrânia	0	230	1.531
Uruguai	0	355	78
Estados Unidos	0	0	143
Venezuela	0	21	9
Iugoslávia	0	0	2.132

Fonte: Dados da pesquisa.

TABELA B.3
ABCC por país e coorte (1840-1910)

País	1840	1850	1860	1870	1880	1890	1900	1910
Argentina		84.7	87.0	93.8	99.2			
Áustria	90.3	96.1	96.8	93.0	97.5	99.5	95.9	
Brasil	69.5	65.5	70.2	78.3	83.6	87.4	94.7	
República Tcheca		94.7	95.3	100.0		97.8	100.0	95.7
Alemanha	91.7	95.3	96.2	95.0	99.5	98.8	100.0	98.4
Hungria		97.4	96.4	95.6	96.0	95.1	100.0	94.1
Itália	95.4	97.4	97.3	97.8	98.4	99.9	100.0	98.8
Japão			98.7	100.0	99.6	100.0	100.0	100.0
Letônia			90.9	94.6	93.2	93.0	100.0	97.9
Lituânia				92.8	96.2	93.8	99.6	100.0
Polônia	78.6	89.5	94.6	94.2	87.3	96.2	100.0	100.0
Portugal	93.1	96.0	98.1	98.5	98.2	97.5	100.0	99.2
Romeno				91.9	94.8	94.3	100.0	
Rússia	72.4	86.6	97.3	92.9	97.1	98.0	100.0	98.3
Espanha	85.5	92.4	94.6	97.8	97.5	100.0	97.2	99.8
Síria	37.5	36.8	49.3	70.8				
Ucrânia							100.0	99.9
Iugoslávia					99.1	95.3	100.0	98.7

Fonte: Dados da pesquisa.

Ipea – Instituto de Pesquisa Econômica Aplicada

Assessoria de Imprensa e Comunicação

EDITORIAL

Coordenação

Cláudio Passos de Oliveira

Supervisão

Everson da Silva Moura

Leonardo Moreira Vallejo

Revisão

Ana Clara Escórcio Xavier

Camilla de Miranda Mariath Gomes

Clícia Silveira Rodrigues

Idalina Barbara de Castro

Olavo Mesquita de Carvalho

Regina Marta de Aguiar

Reginaldo da Silva Domingos

Alice Souza Lopes (estagiária)

Amanda Ramos Marques (estagiária)

Isabela Monteiro de Oliveira (estagiária)

Isabella Silva Queiroz da Cunha (estagiária)

Lauane Campos Souza (estagiária)

Lynda Luanne Almeida Duarte (estagiária)

Polyanne Alves do Santos (estagiária)

Editoração

Aeromilson Trajano de Mesquita

Bernar José Vieira

Cristiano Ferreira de Araújo

Danilo Leite de Macedo Tavares

Herlyson da Silva Souza

Jeovah Herculano Szervinsk Junior

Leonardo Hideki Higa

Capa

Danielle de Oliveira Ayres

Flaviane Dias de Sant'ana

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

Tel.: (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.

ipea Instituto de Pesquisa
Econômica Aplicada

MINISTÉRIO DO
**PLANEJAMENTO,
DESENVOLVIMENTO E GESTÃO**



ISSN 1415-4765

