

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

BRUNO ZANOTTO VIGNA

**DIVERSIDADE, IMIGRAÇÃO E
DESENVOLVIMENTO**

Rio de Janeiro
2019

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

BRUNO ZANOTTO VIGNA

**DIVERSIDADE, IMIGRAÇÃO E
DESENVOLVIMENTO**

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Economia da Indústria e Tecnologia, Instituto de Economia, Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Doutor em Economia.

Orientador: Rudi Rocha de Castro (IE/UFRJ)

Rio de Janeiro
2019

FICHA CATALOGRÁFICA

V678 Vigna, Bruno Zanotto
Diversidade, Imigração e Desenvolvimento / Bruno Zanotto Vigna. - 2019.
192 p.; 31 cm.

Orientador: Rudi Rocha de Castro

Tese (doutorado) – Universidade Federal do Rio de Janeiro, Instituto de Economia, Programa de Pós-Graduação em Economia da Indústria e Tecnologia, 2018.

Bibliografia: f. 138 – 150.

1. Imigração. 2. Diversidade. 3. Desenvolvimento. I. Castro, Rudi Rocha de, orient.
II. Universidade Federal do Rio de Janeiro. Instituto de Economia. III. Título.

CDD 325.1

BRUNO ZANOTTO VIGNA

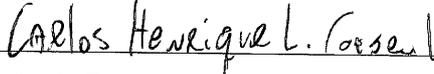
Diversidade, Imigração e Desenvolvimento

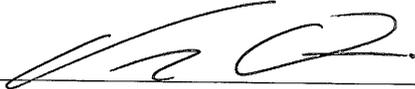
Tese de Doutorado apresentada ao Programa de Pós-Graduação em Economia da Indústria e Tecnologia, Instituto de Economia, Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Doutor em Economia.

Aprovado em: Rio de Janeiro, 28 de junho de 2019.


Prof. Dr. Rudi Rocha de Castro (IE/UFRJ) (Orientador)


Prof. Dr. Arthur Amorim Bragança (PUC-Rio)


Prof. Dr. Carlos Henrique Leite Corseuil (IPEA)


Prof. Dr. Francisco Junqueira Moreira da Costa (EPGE-FGV)


Prof. Dr. Romero Cavalcanti Barreto da Rocha (IE/UFRJ)

Rio de Janeiro
2019

RESUMO

Vigna, Bruno Zanotto. **Ensaio sobre Diversidade, Imigração e Desenvolvimento**. 2019. 56 f. Tese (Doutorado em Economia) - PPGE, Instituto de Economia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2019.

Esta tese é composta por três ensaios complementares sobre diversidade, imigração e desenvolvimento de longo prazo.

O primeiro documenta os padrões de ocupação do território paulista no contexto da imigração em massa para o Brasil do fim do século XIX e início do XX. Mais especificamente, documentam-se os determinantes da dispersão da diversidade de origem dos imigrantes, com base na combinação de dois conjuntos de informação histórica: (i) o fato de que o fluxo de imigrantes teve grande flutuação no período 1880-1920 em sua composição/diversidade, e (ii) a expansão das linhas férreas, uma vez que os recém-chegados eram alocados nas terras recém-conectadas pela ferrovia.

Os resultados evidenciam a importância das ferrovias para a distribuição dos fluxos imigratórios com destino a São Paulo entre 1880 e 1920. Mostram, ainda, que a ocorrência de eventos idiossincráticos, específicos de cada nacionalidade, tais como conflitos na Alemanha e Portugal, e alterações da legislação emigratória na Itália e Japão, condicionaram a composição de origens dos fluxos imigratórios neste período. É importante prever esta dispersão pois ela está dissociada de quaisquer outros fatos ligados à expansão da economia local. Com isso, isola-se a parte causal da correlação entre diversidade e o desempenho econômico dos municípios. A diversidade predita resultante do primeiro artigo é insumo para o seguinte.

O segundo artigo investiga os efeitos de longo prazo da diversidade de local de nascimento. Em específico, testa se os municípios de população mais diversa, posteriormente à onda imigratória registrada entre 1880 e 1920, tiveram melhor desempenho econômico de longo prazo. A estimação por OLS resulta em coeficientes viesados, uma vez que, (i) pode haver causalidade reversa, ou seja, as cidades mais desenvolvidas podem ter atraído diversidade; (ii) há inúmeras variáveis omitidas, não-observáveis, que explicam renda de longo prazo e são potencialmente correlacionadas com o fluxo imigratório, e; (iii) há potencial erro de medida na variável independente. Faz-se uso, então, de uma estratégia empírica que isola a variação trazida pela diversidade com o uso de um instrumento, construído no primeiro artigo, a partir dos fluxos anuais de imigrantes observados ao longo de 40 anos.

Mostra-se que a diversidade de origem dos imigrantes tem efeito positivo na renda per capita de longo prazo. Municípios com um desvio-padrão a mais de diver-

sidade em 1920 apresentam, em média, renda per capita de 12% a 21% mais alta em 2000, a depender do instrumento utilizado. Testes de robustez foram implementados para maior segurança de que os resultados não são influenciados por outros fatores (inclusão de controles, erros-padrão por *bootstrap* e o uso de instrumentos alternativos). Também são documentados os mecanismos que são subjacentes a esse impacto de longo prazo: na média, os municípios mais diversos em 1920 tinham (i) maior proporção de ocupados na indústria e serviços do que no setor agrícola, em 1920 e em 2000, o que confirma a hipótese dos benefícios da diversidade para a economia regional e urbana de Jacobs (1969) e Glaeser et al.(1992); (ii) maior diversidade ocupacional (emprego industrial ou total), em 1920 e em 2000, conforme já fora sugerido por estudos teóricos de Kim (2007) e Lazear (1998) e estimado empiricamente por Alesina, Harnoss e Rapoport (2015); (iii) maior gasto municipal com educação, já na década de 1920, o que evidencia impacto positivo a respeito da controvérsia de Alesina e La Ferrara (2005) sobre o efeito da diversidade na provisão de bens públicos, e; (iv) mais escolas e professores por criança em idade escolar, maior proporção de crianças matriculadas em 1920 e em 2000, e mais anos de escolaridade em 2000, corroborando os resultados encontrados por Rocha, Ferraz e Soares (2017).

Finalmente, o terceiro ensaio investiga se diferenças regionais do país de origem dos imigrantes influenciam os salários dos descendentes no país de destino. Para isso, utiliza dados dos emigrantes italianos com destino a São Paulo durante o período de imigração em massa (1880-1920) e identifica os sobrenomes dos emigrantes nos microdados de mercado de trabalho atuais (RAIS). Estima-se, aqui, a importância do desenvolvimento das regiões italianas de origem nos rendimentos do trabalho dos descendentes de italianos atualmente. O artigo encontra elasticidade-ponto dos ganhos dos descendentes em relação às regiões de origem dos ancestrais de 0,023 a 0,05, a depender dos controles incluídos, sempre estatisticamente significativos a 1%. Estes resultados confirmam a persistência da relação entre região de origem dos descendentes e seus rendimentos do trabalho contemporâneos. O principal elemento novo no presente trabalho é o fato de se vincular pessoas através de mais de duas gerações. Os resultados encontrados são consistentes com os de outros autores que examinaram a mobilidade socioeconômica de longo prazo (e.g., Collado et al., 2013; Clark e Cummins, 2014; Barone e Mocetti, 2016).

Entender o impacto da imigração e da diversidade populacional dela resultante nas localidades é de suma importância. Compreender as consequências de movimentos históricos é fundamental para nortear o redesenho de políticas públicas no presente. Os custos no curto prazo devem ser relativizados pelos benefícios de longo prazo, que começam a ser quantificados graças à disponibilidade dos microdados e do avanço do ferramental de análise na ciência econômica.

Palavras-chave: Diversidade, Imigração, Desenvolvimento de Longo Prazo, Variáveis Instrumentais.

ABSTRACT

Vigna, Bruno Zanotto. **Ensaio sobre Diversidade, Imigração e Desenvolvimento**. 2019. 56 f. Tese (Doutorado em Economia) - PPGE, Instituto de Economia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2019.

This thesis is composed of three complementary essays on diversity, immigration and long-term development.

The first essay documents the patterns of occupation of São Paulo in the context of mass immigration in the late 19th and early 20th centuries. More specifically, it documents the determinants of the dispersion of immigrant birthplace diversity, based on the combination of two sets of historical information: (i) the fact that the influx of immigrants had great fluctuation in the period 1880-1920 in its composition / diversity, and (ii) the expansion of railways, once newly arrived immigrants were allocated to the lands just connected to the rail network. The results show the importance of railroads for the territorial distribution of immigration flows. They also show that the occurrence of idiosyncratic events specific to each nationality, such as conflicts in Germany and Portugal, and changes in emigration legislation in Italy and Japan, conditioned the composition of origins of those flows. Predicting this dispersion is important because it is dissociated from any other facts linked to the expansion of the local economy. It isolates the causal part of the correlation between diversity and the economic performance of municipalities. The predicted diversity resulting from this first article is input to the following.

The second article investigates the long-term economic effects of birthplace diversity. Specifically, it tests the hypothesis that the municipalities of the State of São Paulo with a more diverse population after the immigration wave recorded, between the 1880s and 1920s, had a better long-term economic performance. The OLS estimation yields biased coefficients, since (i) there may be reverse causality, that is, more developed cities may have attracted diversity; (ii) there are numerous omitted variables that explain long-term income and are potentially correlated with the immigrant flow; (iii) there is potential measurement error in the independent variable. The article adopts then an empirical strategy to isolate the variation brought by diversity with the use of an instrument, constructed in the first article, based on the annual flows of immigrants observed over the course of 40 years. These streams of distinct compositions of origin are distributed in the territory as the railroads advanced. The variation of the diversity of origin over time, interacted with a vector of expansion of the railways towards the west of the state, allowed the creation of a measure of predicted diversity for each municipality, which was used

as an instrument for the diversity of nationalities observed in 1920.

The results show that the diversity of origin of immigrants has a positive effect on long-term per capita income. Municipalities with one more standard deviation of diversity in 1920 have, on average, from 12% to 21% higher per capita income in 2000, depending on the instrument used in the estimates. Robustness checks were implemented to assure that the results are not influenced by other factors (including controls, standard errors by bootstrap and the use of alternative instrument).

The mechanisms underpinning this long-term impact are also documented: on average, the most diverse municipalities in 1920 had (i) a larger proportion of those employed in industry and services in 1920 and in 2000, which confirms the hypothesis of the benefits of diversity for the regional and urban economy of Jacobs (1969) and Glaeser et al. (1992); (ii) greater occupational diversity (industrial or total employment) in 1920 and in 2000, as suggested by theoretical studies by Kim (2007) and Lazear (1998) and by empirical estimates by Alesina, Harnoss, and Rapoport (2015); (iii) higher municipal spending on education, as early as the 1920s, which brings empirical evidence to the controversy in Alesina and La Ferrara (2005) on the effect of diversity on the local provision of public goods; (iv) more schools and teachers per school age child, a higher proportion of children enrolled in 1920 and in 2000, and more years of schooling in 2000, corroborating the results found by Rocha, Ferraz and Soares (2017).

Finally, the third essay investigates whether the regional differences in the immigrants' country of origin influence the labor market outcomes of the descendants in the country of destination. It uses data from Italian emigrants bound for São Paulo during the period of mass immigration (1880-1920) and identifies the emigrants' surnames in the current labor market microdata in Brazil (RAIS). The article then estimates the importance of the development of the Italian regions of origin in the Italian descendants' labor income today. The article finds a point elasticity relative to the ancestral home regions of 0.023 to 0.05, depending on the controls included in the specifications, always statistically significant at 1%. These results confirm the persistence of the relationship between the region of origin of the descendants and their contemporary labor income. The main new element in the present work is the fact of linking people through more than two generations. The results are consistent with those of other authors who examined long-term socioeconomic mobility (eg, Collado et al., 2013, Clark and Cummins, 2014, Barone and Mocetti, 2016).

Understanding the impact of diversity on localities is of paramount importance. Discerning the consequences of historical movements is fundamental to guide the redesign of public policies. Short-term costs must be relativized by the long-term benefits, which are finally being quantified thanks to the availability of microdata and the advancement of analytical tools in economic science.

Keywords: Diversity, Immigration, Development, Instrumental Variables.

AGRADECIMENTOS

Agradeço, primeiramente, à família. Pelo suporte em todos os momentos e incentivos de todas as maneiras. Compreensão nas semanas de mais ansiedade. Os valores que me foram transmitidos são a base de tudo. A vocês, dedico este trabalho.

Agradeço ao Prof. Rudi Rocha. Desde a visita ao IE, em 2014, quando queria apenas frequentar as aulas como ouvinte. O aprendizado nestes anos foi muito grande. Algumas coisas só vou entender daqui a uns meses. Em muitas ocasiões foi assim. A mensagem recebida só era totalmente assimilada depois de um tempo. Obrigado pela paciência.

Agradecimento especial ao Prof. Luigi Minale, que me recebeu de braços abertos durante o semestre de visiting.

Agradeço ao BNDES pela oportunidade de terminar o doutorado com a licença. Sem ela, o projeto não teria saído do papel.

Por fim, mas não com menor importância, sou grato aos colegas de turma pelos conhecimentos compartilhados. Aos professores das disciplinas que cursei, aos funcionários do IE e da Secretaria de Pós, sempre dispostos a ajudar.

LISTA DE TABELAS

Tabela 1.1: Estatísticas Descritivas	54
Tabela 1.2: Linhas Férreas por Município	55
Tabela 1.3: Quantidades e proporções de municípios cruzados por ferrovias - evolução por quinquênio (1880-1920)	56
Tabela 1.4: Eventos de Repulsão de Imigrantes	57
Tabela 1.5: Painel - Número de imigrantes por país de origem, por ano e por município de destino - Dados do Museu da Imigração (1882-1920)	58
Tabela 1.6: Painel - Número de imigrantes por país de origem, por ano e por município de destino - Dados do Museu da Imigração (1882-1920)	59
Tabela 2.1: Estatísticas Descritivas	103
Tabela 2.2: Painel - Número de imigrantes por país de origem, por ano e por município de destino - Dados do Museu da Imigração (1882-1920)	104
Tabela 2.3: Primeiro Estágio	105
Tabela 2.4: Mínimos Quadrados Ordinários (MQO)	106
Tabela 2.5: Resultado da Estimação em dois estágios - 2SLS	107
Tabela 2.6: Resultado da Estimação em dois estágios - 2SLS - Instrumento: Diversidade Preditada Apenas pelo Tratamento	108
Tabela 2.7: Resultado da Estimação em dois estágios - 2SLS - Erros Padrão dos coeficientes calculados por bootstrap	109
Tabela 2.8: Efeito da Diversidade na Renda per capita, condicional nas iden- tidades nacionais e religiões	110
Tabela 2.9: Os Efeitos de Longo Prazo da Diversidade de Países de Origem no Emprego e na Transformação Estrutural	111
Tabela 2.10: Os Efeitos de Longo Prazo da Diversidade de Países de Origem na Diversidade Ocupacional Industrial	112
Tabela 2.11: Os Efeitos de Longo Prazo da Diversidade de Países de Origem nos Gastos Municipais com Educação	113
Tabela 2.12: Os Efeitos de Longo Prazo da Diversidade de Países de Origem em Variáveis Educacionais	114
Tabela 3.1: Estatísticas Descritivas	132
Tabela 3.2: Frequência das Regiões de Prevalência dos Sobrenomes dos Imi- grantes Italianos (ordem decrescente de número de descendentes)	133
Tabela 3.3: Salários em São Paulo x Renda per capita regiões italianas	134
Tabela 3.4: Salários em São Paulo x Renda per capita regiões italianas e In- terações	135

SUMÁRIO

1	DIVERSIDADE VIA IMIGRAÇÃO: FATORES EXÓGENOS DE REPULSÃO E A DISPERSÃO TERRITORIAL DA DIVERSIDADE DOS FLUXOS IMIGRATÓRIOS EM SÃO PAULO (1880-1920)	13
1.1	Introdução	13
1.2	Dados	17
1.2.1	Estatísticas descritivas	17
1.2.2	Ocupação territorial e infraestrutura ferroviária	21
1.3	Contexto Histórico	24
1.3.1	Fatores de Atração	24
1.3.2	Fatores de Repulsão	26
1.4	Análise Empírica	41
1.4.1	Modelo	42
1.4.2	Resultados sobre imigrantes	43
1.4.3	Resultados sobre a diversidade	47
1.5	Conclusões	48
2	DIVERSIDADE E DESENVOLVIMENTO DE LONGO PRAZO	60
2.1	Introdução	60
2.2	Revisão da Literatura Relacionada	64
2.2.1	Literatura Teórica	64
2.2.2	Literatura Empírica	67
2.2.3	Contribuição do Artigo	69
2.3	Contexto Histórico	70
2.4	Dados	72
2.4.1	Variáveis de interesse	73
2.4.2	Variáveis de resultado e mecanismos	77
2.4.3	Variáveis de controle	78
2.4.4	Estatísticas Descritivas	80
2.5	Estratégia Empírica	81
2.6	Resultados	87
2.6.1	Mínimos Quadrados Ordinários (MQO)	87
2.6.2	Estimações com Instrumentos	88
2.7	Robutez	90
2.7.1	Estimação em 2SLS: diversidade predita apenas pelo tratamento	90
2.7.2	Estimação em 2SLS: erros padrão por <i>bootstrap</i>	91
2.7.3	Cultura e Religião	92

2.8	Canais de Transmissão	93
2.8.1	Transformação Estrutural	93
2.8.2	Diversidade Ocupacional	95
2.8.3	Gastos do Governo Municipal	97
2.8.4	Escolaridade	98
2.9	Conclusões	100
3	IMIGRAÇÃO E PRODUTIVIDADE: CEM ANOS E UM OCEANO DE DISTÂNCIA	115
3.1	Introdução	115
3.2	Contexto Histórico da Emigração na Itália e Dispersão Territorial dos Italianos em São Paulo	119
3.3	Dados e Estratégia Empírica	122
3.3.1	Dados	122
3.3.2	Estratégia Empírica	124
3.4	Resultados	125
3.5	Considerações Finais e Próximos Passos	127
4	CONCLUSÃO	1
	REFERÊNCIAS	2
	APÊNDICE A	

1 DIVERSIDADE VIA IMIGRAÇÃO: FATORES EXÓGENOS DE REPULSÃO E A DISPERSÃO TERRITORIAL DA DIVERSIDADE DOS FLUXOS IMIGRATÓRIOS EM SÃO PAULO (1880-1920)

1.1 Introdução

O objetivo do presente artigo é documentar a dispersão territorial da diversidade dos fluxos imigratórios para São Paulo no período 1880-1920. A demanda por mão-de-obra para o cultivo do café, em forte expansão no final do século XIX e início do XX, atraiu grandes quantidades de imigrantes de dezenas de nacionalidades. A composição de origens destas correntes de trabalhadores que chegavam ao Brasil foi, no entanto, resultado da ocorrência de choques exógenos nos países de origem dos imigrantes (p.ex., conflitos, epidemias, secas e alterações nas legislações imigratórias dos países emissores e dos destinos concorrentes do Brasil à época, a dizer, Estados Unidos, Canadá e Argentina, para ficar nos mais importantes). A dispersão destes fluxos pelo território paulista, por sua vez, foi condicionada pela expansão da rede ferroviária, através de oito linhas que cruzavam o Estado, de leste a oeste. A construção das ferrovias iniciou-se em 1867, ligando o porto de Santos, principal responsável pelo escoamento das exportações de café do país, à cidade de Jundiaí, no coração da área produtora dos grãos à época. Em 1920, as linhas cortavam 74% dos 202 municípios. Este fato histórico e a coincidência temporal com o período de maior fluxo imigratório permitiram a construção de um modelo com boa capacidade de previsão da diversidade de origem dos imigrantes nos municípios de destino.

A literatura de história econômica documenta a chegada dos imigrantes de

maneira fragmentada, analisando características e episódios de cada nacionalidade ou etnia isoladamente. Dada sua importância relativa em termos quantitativos, a imigração italiana é a mais frequente (Rosoli, 1992; Sori, 1979; Benedini e Arquilla, 2007; Cenni, 2009). Em volume de publicações, destaca-se também a literatura sobre a imigração portuguesa (Leite, 1987; Baganha, 1991), mas a imigração desta origem em específico está relacionada a todo o período colonial e não dá tratamento especial para o período escopo deste artigo. A imigração espanhola no período da imigração em massa do final do século XIX e início do XX é pormenorizada por Souza (2006). Muito embora haja detalhamento das estimativas de quantitativos por reinos de origem e das regiões de destino em São Paulo, a obra dá ênfase ao engajamento dos espanhóis nos movimentos grevistas de 1917, principalmente na capital do Estado. Já Bellotto (1992) e Oliveira (2015) se restringem ao período da Guerra Civil, na segunda metade da década de 1930. Sobre a imigração alemã, os trabalhos se concentram nos fluxos com direção ao sul do país (e.g., Schäffer, 1984; Mauch e Vasconcellos, 1994). A dispersão geográfica da imigração japonesa foi alvo de maior documentação, pois quando passou a ser volumosa, funcionava por meio de companhias de desenvolvimento das colônias, responsáveis pela promoção, recrutamento, transporte e, inclusive, pelo assentamento das famílias nos destinos (Takeuchi, 2007; Hashimoto, 2008; Carneiro e Takeuchi, 2010). Ainda que, no caso dos japoneses, haja maior documentação da ocupação territorial dos fluxos imigratórios, a literatura se restringe a esta origem, não analisando o contexto da dispersão das outras nacionalidades ou à luz dos padrões observados na sequência temporal dos choques exógenos que os determinam. O mesmo ocorre com, por exemplo, Knowlton (1961), Truzzi (2005) e Mizrahi (2005) que tratam de maneira isolada a imigração sírios, libaneses e judeus, respectivamente.

Dos trabalhos que não se restringem a uma nacionalidade, etnia ou religião específicas, cabe citar Moura e Paiva (2008), que documentam a importância da Hospedaria dos Imigrantes. No entanto, não há detalhamento do processo de aloca-

ção territorial dos recém-chegados. Bassanezi (2008) traz, no "Atlas da Imigração Internacional em São Paulo, 1850–1950", um extenso registro da dispersão dos fluxos imigratórios e do papel das ferrovias, mas não faz qualquer conexão com os fatores de repulsão das distintas origens.

A historiografia econômica também estuda a expansão ferroviária, mas frequentemente a relacionando apenas à sua importância econômica derivada da relação com o capital excedente da cafeeira e, posteriormente, dos seus efeitos para o processo industrializante, já nas primeiras décadas do século XX (e.g., Dean, 1969, 1977; Netto, 1981; Cardoso de Mello, 1984; Abreu e Bevilaqua, 1996; Gonçalves, 2009).

Desta forma, o artigo contribui para a literatura existente de duas maneiras. Em primeiro lugar, diferentemente dos trabalhos já existentes, documenta a dispersão geográfica de cada uma das principais nacionalidades de forma unificada e sequencial. Mostra, ainda, que a expansão da ferrovia para o oeste e norte do território foi o vetor de difusão desta composição de origens dos imigrantes. Em segundo lugar, faz uma análise da importância de choques exógenos ocorridos nos países emissores para a determinação do espalhamento da diversidade dos fluxos imigratórios com destino a São Paulo no período em análise (1880-1920).

O artigo traz sequências de mapas do território paulista que retratam a expansão territorial do povoamento por nacionalidade a cada quinquênio a partir de 1880 (vide Apêndice, para mapas por nacionalidade de origem e por quinquênio, além de mapas que ilustram o alcance das ferrovias a cada década). Nunn, Qian e Sequeira (2017) utilizam estratégia empírica similar para os Estados Unidos e mostram que regiões vizinhas entre si receberam distintos volumes de imigrantes a depender de quando a localidade foi conectada à ferrovia, que à época avançava em direção ao oeste americano. Se um condado se conectava à rede durante períodos

de alta (baixa) imigração, tendia a ter mais (menos) assentamentos de imigrantes.

Nota-se a importância das ferrovias, pois o desenvolvimento dos municípios mais a oeste se deu à medida que as linhas férreas avançavam (Saes, 1981, 1996; Lamounier, 2000; Stefani, 2007; Grandi, 2013). A diferença entre o padrão de ocupação geográfica dos italianos e espanhóis, por exemplo, em relação a dos japoneses e russos, é resultado do período em que os imigrantes destes países chegaram ao Brasil.¹ Enquanto os imigrantes da Itália e Espanha chegaram já a partir de 1880, os russos e japoneses só emigraram em maior número a partir de 1905 e 1912, respectivamente. Neste artigo, utiliza-se um modelo em painel que explora eventos exógenos (para estas duas origens, em específico, a Guerra Russo-Japonesa / Revolução de 1905 e as alterações nas legislações imigratórias relativas aos japoneses de 1911-13) e permite a predição das composições anuais dos fluxos imigratórios para os municípios paulistas. Uma vez que os imigrantes recém-chegados eram alocados para as fronteiras agrícolas por meio das ferrovias, aproveita-se este padrão de ocupação para se prever a alocação das distintas composições de imigrantes nos municípios recém-conectados à malha ferroviária. Os mapas mostram que russos e japoneses, por exemplo, por chegarem em um segundo momento, ocupam porções mais afastadas das zonas tradicionais de cultivo do café, como Jundiaí, Campinas e Itú.

Compreender como se deu a distribuição geográfica e temporal da diversidade de origens introduzida pela imigração na virada do século XIX é importante também para se isolar a parte causal da correlação entre diversidade e o desempenho econômico dos municípios. A variável de resultado deste artigo é a diversidade predita ao nível dos municípios e é insumo para o artigo seguinte da tese.

O restante do artigo se desenvolve em mais três seções, além da conclusão. A

¹A Tabela 1.4 resume os eventos de repulsão, discriminados por país, considerados no artigo.

Seção 1.2 descreve o conjunto de dados utilizados e discute as potenciais medidas de diversidade. A Seção 1.3 documenta os fatores de atração e de repulsão dos imigrantes, investigando, em suas subseções, os casos de cada uma das principais origens dos imigrantes. Finalmente, a Seção 1.4 traz a análise empírica, com o detalhamento do modelo utilizado e dos resultados sobre imigrantes e sobre a dispersão geográfica da diversidade.

1.2 Dados

1.2.1 Estatísticas descritivas

O artigo utiliza dois conjuntos de dados. Primeiro, os registros coletados no momento da chegada dos imigrantes para matrícula na Hospedaria do Brás, em São Paulo. A partir deles, elencaram-se as quantidades anuais de imigrantes por país de nascimento para cada município de destino.² Em segundo lugar, com base em pesquisas de fontes históricas,³ foi construída a série temporal do avanço de cada linha férrea que cruzava o território paulista em 1920, a partir do cronograma de inaugurações das estações ferroviárias.

O crescimento econômico e o adensamento populacional ao longo do anos aumentou a divisão administrativa no Estado de São Paulo. Enquanto em 1920 o território era dividido em 202 municípios, atualmente, há 645. Desta forma, as análises aqui apresentadas utilizam a divisão administrativa de 202 municípios de 1920. Os 645 municípios atuais foram agrupados de modo a recompor as fronteiras de 1920. Com este procedimento, um município de 1920 pode corresponder a diversos

²As Tabelas A.1 e A.2, no Apêndice, trazem estatísticas das nacionalidades e municípios de destino mais comuns nos registros da Hospedaria.

³O portal www.estacoesferroviarias.com.br compila informações sobre as linhas e estações ferroviárias, suas datas de inauguração e coordenadas geográficas, entre outras.

da divisão atual, mas os territórios passam a ser comparáveis no tempo, a partir de 1920 até o presente.

O Painel A da Tabela 1.1 apresenta um resumo das estatísticas descritivas baseadas na divisão administrativas de 1920, com 202 municípios. Ele traz os dados sobre diversidade e a população municipal. O índice de diversidade apresentado refere-se à diversidade intragrupo, análogo ao de fracionalização, comumente utilizada na literatura, entretando independente dos nativos.

[TABELA 1.1 AQUI]

A maior parte da literatura empírica utiliza como medida de diversidade o índice de fracionalização (e.g., Alesina e La Ferrara, 2005; Beach e Jones, 2017; Ottaviano e Peri, 2006; Ager e Brückner, 2013):

$$Frac_{m,1920} = 1 - \sum_p (s_{m,p,1920})^2$$

Onde: $s_{m,p}$ é a proporção de imigrantes provenientes do país de nascimento ‘p’, no município ‘m’ em 1920; $s_{m,1}$, em particular, se refere aos habitantes nativos do município ‘m’. Inspirado no índice de concentração de Herfindahl-Hirschmann (IHH), a fracionalização mede a probabilidade de que dois habitantes de um município, sorteados aleatoriamente, tenham dois países de nascimento diferentes. A fracionalização zero denota diversidade zero, ou seja, ocorre quando todos os residentes do município são de um mesmo país de nascimento. Fracionalização igual à unidade, por sua vez, se dá quando cada habitante do município tem um país de nascimento diferente.

O índice de fracionalização, no entanto, tem duas dimensões e é importante separá-las (Alesina, Harnoss e Rapoport, 2016). A primeira é a proporção de es-

trangeiros⁴ $(1 - s_1)$, que independe do país de origem e capta a escala dos fluxos de imigrantes. A segunda é a variedade intragrupo de imigrantes. Em outras palavras, a proporção de estrangeiros mede a diversidade na dimensão intergrupos (entre nativos e não-nativos) e a variedade de estrangeiros mede a diversidade na dimensão intragrupo (apenas entre estrangeiros). Ao agruparmos todos os estrangeiros $(1 - s_1)$, independentemente de seu país de origem, resumindo o somatório em apenas dois termos, s_1 e $(1 - s_1)$, o índice de fracionalização pode ser simplificado da seguinte forma:

$$Frac = 1 - \sum_p (s_p)^2 = s_1 * (1 - s_1) + (1 - s_1) * s_1 = 2 * s_1 * (1 - s_1)$$

Temos, assim, que a fracionalização (ou diversidade intergrupos nativos e imigrantes) é função de $(1 - s_1)$. Utilizar a fracionalização pode nos levar a confundir o tamanho ou escala do fluxo com a variedade do intragrupo (apenas entre estrangeiros).

A diversidade intragrupo tem cálculo análogo ao da fracionalização, mas independente dos nativos. Define-se s_j como a proporção de imigrantes de origem j entre o total de estrangeiros: $s_j = (s_p)/(1 - s_1)$. A diversidade (ou variedade) de imigrantes, independente do tamanho da parcela de estrangeiros, então, fica definida conforme a equação 1.1, abaixo, e será a principal variável de interesse do artigo:

$$Div_{m,1920} = 1 - \sum_j (s_{m,j,1920})^2 \quad (1.1)$$

Ou seja, o índice de diversidade utilizado no artigo reflete a diversidade apenas entre os estrangeiros. Assim como no caso do índice de fracionalização, ela assume valores entre zero e um. Zero denota diversidade zero e valor unitário, para o caso de diversidade máxima, no caso limite em que cada estrangeiro do município

⁴Aqui, s_1 é a proporção de nativos.

tem uma origem distinta. Entre os 202 municípios do Estado de São Paulo em 1920, a média dos índices de diversidade observado é de 0,431. Aproximadamente um quarto dos municípios não recebeu imigrantes e, sendo assim, apresenta diversidade igual a zero. O município com maior diversidade entre estrangeiros apresentava índice de 0,821. A média do índice de diversidade predito é de 0,593 e o valor máximo para esta variável foi de 0,799.

Livro de Matrículas da Hospedaria dos Imigrantes

O artigo utiliza dados coletados no momento da chegada dos imigrantes para registro na Hospedaria do Brás, em São Paulo. O banco de dados, fornecido pelo Museu da Imigração de São Paulo,⁵ contém informações sobre aproximadamente 1,6 milhão de imigrantes que chegaram ao Brasil entre 1882 e 1958. Apesar das estimativas encontradas na historiografia apontarem para o recebimento de cerca de 2,5 milhões de imigrantes na Hospedaria (Holloway, 1984; Vangelista, 1991; Vasconcellos, 1994; Gonçalves, 2012), só há registro sistematizado de 65% das matrículas totais.⁶ Os registros trazem os nomes e sobrenomes dos recém-chegados, bem como seu gênero, estado civil, idade, nacionalidade, profissão, religião, relações de parentesco intrafamiliares, seu último país de residência, destino no Brasil e informações sobre o financiamento de sua viagem (se vieram com subsídios dos governos receptores – central ou estaduais; se via sociedades promotoras de imigração; ou se a imigração foi espontânea ou dita autopatrocinada). Com base neste livro de matrículas, construiu-se o banco de dados que descreve as quantidades anuais de imigrantes por país de nascimento para cada município de destino.

⁵O edifício da Hospedaria do Brás foi tombado pelo CONDEPHAAT (Conselho de Defesa do Patrimônio Histórico, Arqueológico, Artístico e Turístico) em 1982 e, atualmente, abriga o Museu da Imigração, sob administração da Secretaria da Cultura do Governo do Estado de São Paulo. O Museu disponibiliza seu acervo digitalizado por meio de sua página na internet: <http://www.museudaimigracao.org.br/>.

⁶De acordo com o Centro de Preservação, Pesquisa e Referência (CPPR) do Museu da Imigração, nem todos os livros de matrícula foram digitalizados por questão de conservação.

A Tabela A.1, no Apêndice, mostra as quantidades anuais de imigrantes dos principais países de origem mais frequentes. A Tabela A.2 traz as quantidades anuais de imigrantes dos principais municípios de destino. Quando o destino do imigrante era uma estação de trem ou fazenda, foi feita pesquisa sobre a sua localização para que fosse atribuído o município de destino corretamente. Os registros cujas estações ou fazendas não puderam ser localizadas foram descartados.⁷

1.2.2 Ocupação territorial e infraestrutura ferroviária

A ocupação do território paulista deu-se de leste para o oeste e acompanhou a construção das ferrovias. As primeiras linhas férreas, naturalmente, ligaram as principais regiões produtoras de café ao porto exportador de Santos. No entanto, à medida em que avançavam em direção aos rios Paraná e Grande, que fazem fronteira com os Estados vizinhos passou a ocorrer o inverso, ou seja, cidades nasciam com a ferrovia (Saes, 1981; Stefani, 2007; Grandi, 2013). Portanto, o *timing* da conexão à rede ferroviária das localidades prediz a onda de imigrantes que ali vai se estabelecer. As Figuras de A.4 a A.60, no Apêndice, trazem mapas que ilustram a expansão da ocupação populacional pelos imigrantes a cada quinquênio (total e por nacionalidade) e da expansão das estações ferroviárias a cada década.

O artigo utiliza dados detalhados sobre a ampliação da infraestrutura ferroviária paulista a partir do último quarto do século XIX. O deslocamento da fronteira agrícola e da urbanização no Estado esteve intimamente ligada à proximidade de meios de transporte necessários para o escoamento da produção exportável (Dean, 1969). A expansão da rede de ferrovias muitas vezes foi financiada com excedente

⁷O Livro de Matrículas da Hospedaria dos Imigrantes contém 1.264.204 registros até o ano de 1920, dos quais 689.161 são *non-missing* na variável destino. Destes, 2.863 (0,4%) são localidades fora do Estado de São Paulo e 1.303 (0,2%) não puderam ser localizados a partir do que constava no campo descritivo da base de dados digitalizada. A amostra final é de 684.995 observações.

de capital do café, tanto que entre os acionistas das companhias ferroviárias era frequente figurarem as famílias dos cafeicultores (Gonçalves, 2008). Construído a partir de informações históricas sobre as companhias e datas de construção das linhas e estações de trem,⁸ foram utilizadas dummies de presença de ferrovias nos municípios em 1920, bem como efeitos fixos de linhas férreas, uma vez que cada conjunto de municípios era cortado por uma linha férrea diferente, com motivações, acionistas e datas de construção distintas.

O Painel B da Tabela 1.1 resume os dados sobre a infraestrutura ferroviária de 1920 e a presença de ferrovias nos municípios em São Paulo. Para determinar se um município estava conectado à rede ferroviária, atribuiu-se variável indicadora igual à unidade se o limite do município era cruzado por pelo menos uma linha ferroviária. Adicionalmente, para se capturar características das distintas regiões cortadas pelas diferentes linhas, criaram-se efeitos fixos de linhas férreas. O território de São Paulo era cortado por 8 diferentes linhas em 1920 (Central, São Paulo-Jundiaí, Paulista, Sorocabana, Mogiana, Noroeste, São Paulo-Minas, Araraquara). Desta forma, para cada município atribuiu-se um valor entre 1 e 8, a depender de qual linha cruzou pela primeira vez os limites do município.

Enquanto 52 municípios, ou 26% dos 202 municípios existentes no Estado de São Paulo em 1920, não tiveram seus limites cruzados por qualquer linha férrea, a rede ferroviária atinge os 74% restantes já em 1920. O fato de que, em 1872, apenas 2% dos municípios tinham seus territórios cortados por linhas férreas, evidencia a velocidade da expansão deste meio de transporte, o principal naquela época, tanto para escoar mercadorias quanto para os deslocamentos populacionais.

A Tabela 1.2 traz as quantidades de municípios cortados por cada uma das oito companhias de transporte ferroviário operantes em 1920. As linhas Central

⁸Dados disponíveis em www.estacoesferroviarias.com.br.

e Inglesa (SP Railway, que ligava Santos a Jundiaí) foram as primeiras a serem construídas. Em seguida, novas companhias, tais como a Paulista, a Sorocabana e a Mogiana, foram criadas com capital dos fazendeiros para expansão da malha rumo às novas fronteiras agrícolas. Num segundo momento, mais a oeste e norte, e já atingindo as fronteiras do Estado de São Paulo com os vizinhos Mato Grosso e Minas Gerais, respectivamente, houve a construção da Estrada de Ferro Noroeste, a partir do município de Bauru, e da linha SP-Minas.

[TABELA 1.2 AQUI]

A Tabela 1.3, por sua vez, resume a evolução das proporções de municípios cruzados pelas linhas a cada quinquênio a partir de 1880. Ela coincide com os períodos de maior expansão das lavouras de café, conforme documentado pela literatura de história econômica (Saes, 1996; Lamounier, 2000; Grandi, 2013). O Apêndice traz, na Tabela A.3, informações sobre as primeiras estações construídas em cada município, que incluem data da inauguração, linha a que pertencia e coordenadas geográficas.

[TABELA 1.3 AQUI]

A Figura 1.3 mostra a relação positiva entre os anos de inauguração das estações e a sua posição geográfica, cada vez mais a oeste.

[FIGURA 1.3 AQUI]

1.3 Contexto Histórico

1.3.1 Fatores de Atração

No segunda metade do século XIX, o Brasil e, em maior grau, São Paulo, enfrentavam a falta de mão-de-obra necessária para o atendimento da forte demanda por exportações de café. Em 1850, o tráfico de escravos fora banido e, em 1871 e 1888, as Leis do Ventre Livre e Áurea, respectivamente, restringiram ainda mais a oferta de trabalho dos africanos, a mais comumente utilizada até então. Manter escravos se tornou cada vez mais arriscado e custoso. Em 1881, em São Paulo, houve imposição de nova tributação, desta vez para a transferência de escravos vindos de outras províncias. Neste contexto, o apoio à imigração subsidiada tornou-se uma posição pragmática dos fazendeiros paulistas. Estimativas do IBGE apontam para a entrada no Brasil de mais de 4,1 milhões de imigrantes entre 1872 e 1929. Para efeito de comparação, o Censo de 1872 registra população residente total de 9,9 milhões no país, dos quais 383 mil (3,8%) eram estrangeiros ou naturalizados. Já em 1900, a população residente total atingira 17,4 milhões, porém os estrangeiros e naturalizados já somavam 1,279 milhão, ou 7,3% (IBGE, 2007).

A imigração estrangeira para São Paulo deu-se com forte intervenção estatal, em resposta às necessidades dos produtores do café, produto com 70% de participação na pauta de exportações do país em 1920 (Abreu e Bevilaqua, 1996). As viagens dos imigrantes eram subsidiadas e famílias e homens adultos eram preferidos. Como instrumento de atração, estabeleceram-se quantias indenizatórias aos imigrantes. O ordenamento e a promoção da imigração ficaram a cargo da Secretaria da Agricultura, Comércio e Obras Públicas (SACOP), criada no mesmo ano da promulgação da primeira Constituição do Estado (1891), que também recebeu a atribuição de gerir a “Hospedaria dos Imigrantes”, localizada, inicialmente, no bairro do Bom Retiro

e, alguns anos depois, no Brás. A função da Hospedaria era servir como abrigo aos recém-chegados nos seus primeiros dias no Brasil. A permanência dos imigrantes durava o período necessário para o despacho aduaneiro das bagagens, desinfecção, numa espécie de quarentena,⁹ e indicação do destino e conseqüente espera por condução, na ampla maioria dos casos, ferroviária, uma vez que a linha da São Paulo Railway tinha estação contígua ao seu prédio principal.

Do total de imigrantes que vieram ao Brasil, algo por volta de 58% foram acolhidos por São Paulo (Almeida, 2014). A preferência por este local pode estar vinculada às facilidades concedidas pelo governo local, tais como passagem e alojamento, somadas às oportunidades de trabalho de uma economia em expansão (Fausto, 2009). Apenas entre 1880 e 1900, a província de São Paulo recebeu aproximadamente 940 mil imigrantes (Vasconcelos, 1994, p. 227). Para se ter uma ideia da dimensão deste fluxo, o Censo de 1872 aponta população total da província em cerca de 837 mil habitantes. A maioria dos imigrantes chegados a São Paulo tinha a viagem subsidiada¹⁰ e já sabia o destino final no interior. Também era possível assinar contrato com fazendeiros que procuravam mão-de-obra na Hospedaria (Colistete e Lamounier, 2011; Gonçalves, 2009).

O fluxo imigratório sofre interrupção em 1914, com o início da Primeira Grande Guerra. Como fim do conflito, em 1919, e a política de sustentação de preços levada a cabo pelo governo brasileiro nos anos 1920, a lavoura recupera-se e a demanda por mão-de-obra atrai novo fluxo imigratório. Em 1927, encerram-se os subsídios para a vinda de novos imigrantes. E, em 1930, com a crise provocada pela superprodução do café, o governo Vargas impõe restrições à imigração, que só

⁹Muitos eram os casos de epidemias (e.g., cólera, pelagra, varíola, peste) contraídas ainda nos portos da Europa, em razão das condições precárias na espera pelo embarque, ou mesmo durante a travessia, que chegava a durar três meses (Telarolli Jr., 1996).

¹⁰Segundo os dados do Museu da Imigração, 60% dos imigrantes chegados à Hospedaria entre 1889 e 1915 tiveram sua viagem subsidiada pelo governo brasileiro. A parcela restante é classificada como espontânea ou autofinanciada.

voltaria a crescer após a Segunda Guerra Mundial.

1.3.2 Fatores de Repulsão

Diversos fatores explicam a emigração em massa dos europeus a partir de meados do século XIX. Em primeiro lugar, nesta época houve um crescimento demográfico jamais visto na história da humanidade. As inovações e ganhos de produtividade advindos da revolução industrial provocaram o rompimento da lógica malthusiana. A população europeia cresceu de 144 para 486 milhões entre 1800 e 1920 (Arruda, 2007). Houve, neste processo, uma alteração da relação do camponês com a terra, agora destinada à produção em larga escala. A terra passou a ser fator de produção, não mais meio de vida tradicional do camponês. A falta de trabalho no campo e a fome eram combustíveis para revoltas e levantes que colocavam a ordem social em risco. Os governos viam a emigração como solução parcial redutora de pressão interna. Parte do êxodo foi, em razão disso, incentivado pelas nações emissoras (Gonçalves, 2008).

Assistiu-se, desde o século XVII, a migração sazonal de trabalhadores do sul mediterrâneo para os países vizinhos do norte. A título de exemplo, era comum na Itália recém-unificada, as populações das províncias do norte (Vêneto, Piemonte, Trento e Lombardia) buscarem trabalho na França, Alemanha e Império Austro-Húngaro. No século XVIII e XIX, as dificuldades na economia rural alteraram a duração do expatrio, de sazonal para plurianual. O campo sofria as consequências do avanço do mundo industrial e sofria concorrência dos grãos norte-americanos e argentinos. Os excedentes de população rural, especialmente na Espanha, Portugal e Itália, ao enfrentar uma queda na demanda por sua mão-de-obra, optam pela emigração para a América. Contribuíram, ainda, para o êxodo: (i) redução do custo

dos transportes (e do tempo de viagem), pela introdução dos vapores e ferrovias,¹¹ e; (ii) estabelecimento de redes, viabilizadas pela popularização dos serviços postais e do telégrafo. À medida que as notícias relativas ao sucesso dos primeiros paisanos na América iam chegando, mais e mais emigrantes se encorajavam.

Este era o pano de fundo no contexto histórico dos principais países de onde partiram os imigrantes. O artigo, porém, explora uma série de eventos desconexos entre si, particulares de cada país emissor, que impulsionaram contingentes ainda maiores de imigrantes para o êxodo (por exemplo, conflitos bélicos, condições climáticas de seca extrema, ocorrência de epidemias e alterações nas legislações imigratória e emigratória). A sucessão desordenada destes eventos resulta em fluxos migratórios de composição de nacionalidades idiossincrática do ponto de vista da relação que se pretende estimar, no artigo seguinte da tese, entre a diversidade e o desenvolvimento de longo prazo.

O presente ensaio, desta forma, explora os diversos choques exógenos que afetaram a composição de origens do fluxo anual de imigrantes para São Paulo de modo a eliminar vieses de seleção (ou fontes de endogeneidade) inerentes ao processo de expansão da fronteira agrícola rumo ao oeste da província que atraía os imigrantes. A intenção é ficar apenas com a diversidade resultante dos fatores exógenos, pois esta não é endógena ao ciclo econômico da região receptora dos fluxos. A seguir, os fatores de repulsão específicos de cada um dos principais países emissores são detalhados:

Italianos

A importância da emigração italiana pode ser ilustrada em poucos números. A Itália foi o país que mais exportou mão-de-obra para o Brasil. Na verdade, no

¹¹A duração média da viagem desde a Europa reduziu-se de três meses para duas semanas com o advento dos navios a vapor (Lesser, 2013).

período 1850-1914, foi também a maior exportadora do mundo. Rosoli (1992) estima que 18 milhões tenham deixado a península entre 1870 e 1930. Sori (1979) trabalha com estimativas de 20 milhões de emigrantes para o período entre 1861 e 1940. Segundo Benedini e Arquilla (2007), nos cem anos contados a partir de 1870, registrou-se a partir da península itálica o maior êxodo populacional da história moderna: cerca de 20 milhões de pessoas, aproximadamente um terço da população atual do país, deixaram a Itália com destino a outros países. O Brasil foi o terceiro a receber mais italianos, atrás de Estados Unidos e Argentina, a despeito de ter tido, entre os receptores, a política de atração mais ativa (Gonçalves, 2008). As dificuldades na economia rural ampliaram a duração do expatrio para os países vizinhos, inicialmente sazonal para plurianual. Ao mesmo tempo, o campo italiano sofria as consequências do avanço do mundo industrial e sofria concorrência no comércio internacional dos grãos norte-americanos e argentinos.

A falta de trabalho no campo, a fome e a ocorrência de algumas epidemias, como o cólera, a varíola e a pelagra, eram combustíveis para revoltas e levantes que colocavam a ordem social em risco. Na década de 1870, houve greves e agitações em resposta à taxaço da farinha, que resultaram no confisco de propriedades diante do seu não recolhimento (Vendrame, 2016). De acordo com Rosoli (1992), entre os censos de 1871 e 1880, o número de expatriados mais que dobrou (de 450 mil para aproximadamente 1 milhão). Os excedentes de população rural, ao enfrentar uma queda na demanda por sua mão-de-obra no Império Austro-Húngaro e Alemanha, principalmente, optam pela emigração para a América.

No caso italiano, o fluxo crescente de emigrantes ao longo de mais de duas décadas trouxe outros benefícios para a economia, não imaginados quando do início do movimento. Parte do *catch-up* de salário real logrado pela Itália em relação às economias mais maduras da Europa se deve à maior escassez de mão-de-obra resultante da emigração. Como resultado de uma redução acumulada de 39% na força de

trabalho italiana entre 1870 e 1910, provocada apenas pela emigração, O'Rourke e Williamson (1997) estimam que os salários reais cresceram 49,3% em 1910 (op.cit., p.169). Além disso, a própria mediação, feita por agentes e subagentes de recrutamento nas províncias, bem como o encaminhamento para o porto de Gênova e o transporte propriamente dito dos emigrantes, a cargo das companhias de navegação, eram atividades extremamente lucrativas (Gonçalves, 2008). Diante disso, o Parlamento elabora e discute, ao longo do ano de 1888, uma maneira de regular o processo, garantindo boas condições no transporte dos emigrantes, sem prejudicar a continuidade do fluxo. O processo culmina na aprovação da Lei n.5866, de 30 de dezembro. O artigo primeiro da lei vai direto ao ponto: "*L'emigrazione è libera (...)*", salvo para os militares. Nos artigos seguintes ela impôs controles à ação dos agentes e regulamentou a atividade dos armadores e da marinha mercantil.

Na prática, no entanto, pouco mudou. Com o incremento do fluxo emigratório, intensificou-se a mediação clandestina e apareceu a figura do fretador, que fazia o transporte com embarcações alugadas (Gonçalves, 2008, p.58). Passados mais de dez anos, houve a necessidade de nova legislação, aprovada em 1901, desta vez para regular os preços das passagens, que tiveram altas de 40 a 50% em dois anos (de 110 para 160 liras no trecho Gênova-São Paulo, por exemplo). Foi em 1902, contudo, que outra alteração na legislação italiana provocou desdobramentos importantes para os quantitativos e composição de origem dos fluxos de imigrantes para o Brasil: o Decreto Prinetti. A medida, na verdade, foi uma portaria, emitida pelo Ministério do Exterior italiano. Ela suspendeu as licenças das quatro companhias de navegação previamente autorizadas ao transporte subsidiado de colonos (Cenni, 2009). Também foi proibido o recrutamento de emigrantes nas províncias. O ato foi uma resposta à denúncia de maus tratos aos colonos e de ocorrência de escravidão por dívidas nas fazendas de café, confirmadas pelo chamado "Relatório Rossi", elaborado por um enviado do governo italiano ao Brasil, Adolfo Rossi.

Esta alteração na legislação emigratória italiana, em 1902, teve consequências importantes para a composição de origem do fluxo de imigrantes para o Brasil. A partir da portaria, o número de italianos com destino ao Brasil caiu drasticamente e o Brasil passa a receber imigrantes de outras origens (russos e gregos, por exemplo, a partir de 1905; letões, a partir de 1906; japoneses, búlgaros e estonianos, a partir de 1908; turcos e malteses, a partir de 1912). A chegada de espanhóis e portugueses aumentou consideravelmente após a proibição: a média anual de espanhóis salta de 5,7 mil, nos dez anos que precedem o Decreto, para 15,3 mil no decênio que o sucede; quanto aos portugueses, a média anual, nas mesmas janelas, cresce de 3,3 mil para 6,1 mil. Pela primeira vez desde 1882, os italianos não foram maioria na Hospedaria: entre 1904 e 1913, os espanhóis foram a nacionalidade modal. Especificamente nos anos de 1912 e 1913, houve mais matrículas de portugueses do que italianos, que ficaram então com a terceira posição. Em 1914, o fluxo total caiu 84% em razão da Grande Guerra.

O fluxo de italianos voltou aos níveis pré-Prinetti apenas em 1907, quando o mal estar diplomático foi contornado pelo governo brasileiro com a entrada em vigor de uma série de dispositivos legais de proteção ao imigrante. Dentre as principais, pode-se destacar a fundação, em 1906, da Agência Oficial de Colonização e do Trabalho, destinada a facilitar a assimilação no mercado de trabalho rural ou urbano, e a aprovação da Lei Cardoso de Almeida, 1906, que regulamentou as relações de dívidas financeiras e hipotecárias entre os fazendeiros e os colonos (Cenni, 2009; Reis-Vieira, 2012).

Neste contexto, o artigo estima qual o impacto deste fator exógeno ao ciclo econômico do café de São Paulo na diversidade de origem do fluxo imigratório. É esperado que os municípios que estavam na fronteira agrícola de São Paulo entre os anos de 1902 e 1907, representada pelas então extremidades oeste e norte das linhas férreas que cruzavam o território, teriam recebido, em razão desta legislação

italiana, parcela de italianos menor que em outros períodos. Neste período, o fluxo continha proporcionalmente mais imigrantes de outras origens. É esperado que os municípios localizados nas pontas das linhas férreas neste período tenham recebido maior diversidade que outros. A Seção 1.4 traz os resultados da estimação.

Espanhóis

A Espanha notadamente participou de um processo industrializante tardio quando comparado ao restante do continente europeu. Enquanto a Europa como um todo registrou crescimento médio do salário real da ordem de 1,4% ao ano entre 1870 e 1913, a Espanha decepcionou com taxa de 0,4%. Trata-se de um desempenho pífio, principalmente quando comparado com os outros países da periferia europeia que, neste período, iniciavam o processo de *catch-up* relativamente aos líderes (Reino Unido, França, Suíça, Alemanha, Bélgica e Holanda) e obtiveram performance médio de 2,6% ao ano na Escandinávia (Dinamarca, Suécia, Noruega e Finlândia) ou mesmo 1,8% na Irlanda e Áustria (O'Rourke e Williamson, 1997).

A estrutura fundiária, principalmente no sul, mantinha as populações em regimes senhoriais de produção. A maioria analfabeta não tinha condições de reivindicar mudanças nas zonas urbanas. Neste contexto, o principal fator de repulsão na Espanha foi a pobreza no campo (Souza, 2006). Cerca de 3,3 milhões de espanhóis dirigiram-se à América entre 1880 e 1930. A grande maioria preferiu a Argentina, o Uruguai e Cuba como destino, pelos laços culturais e linguísticos. Estima-se que o Brasil recebeu cerca de 17% deste fluxo, ou o equivalente a 570 mil imigrantes. O Brasil foi um dos principais destinos para os espanhóis porque existia a política de subvenções para as passagens de navio (Fausto, 1999).

As condições da população deterioravam-se ainda mais quando o território passava por estiagens prolongadas, com prejuízos à produção de cereais, videiras,

frutas e gado (Liria e Marcén, 1981). Duas das secas mais longas já registradas no norte da península, nos vales do Ebro (foz a leste) e do Douro (foz a oeste) ocorreram entre os anos de 1893 e 1897. Estes períodos secos só foram superados em 1978 (op.cit., p.64). De acordo com o Catálogo de Secas Históricas de autoria do Ministério da Agricultura espanhol, especialmente o biênio 1895-96 foi classificado como de seca extrema. Especificamente para os anos 1893, 1894 e 1897, como seca moderada ou intensa (Espanha, 2013). Os mapas do relatório histórico referentes a este período estão representados na Figura A.3, no Apêndice.

De fato, o fluxo de imigrantes espanhóis para São Paulo se intensifica entre 1888 e 1897, quando a origem se consolida como a segunda mais frequente na Hospedaria, atrás apenas da Itália e pouco a frente de Portugal. Nestes dez anos de imigração mais intensa, a média anual de matrículas foi de 8,1 mil. Essa média só foi superada no decênio entre 1904 e 1913, quando 15,3 mil por ano passaram pela Hospedaria do Brás.¹²

O artigo explora, portanto, a variação na imigração decorrente deste choque climático, nos moldes do realizado por Nunn, Qian e Sequeira (2017). Há evidências de forte ligação entre o clima e a produção agrícola na Europa durante a Era da Migração em Massa. Por exemplo, Solomou e Wu (1999) estudam Grã-Bretanha, França e Alemanha de 1850 a 1913 e concluem que entre um terço e dois terços da variação total na produção agrícola é explicada por choques climáticos. Também é motivada por descobertas existentes de uma forte relação entre choques climáticos e migração internacional no período contemporâneo dentro dos países em desenvolvimento (Feng, Krueger e Oppenheimer, 2010).

¹²Embora haja registro na historiografia de grande emigração de espanhóis em razão da Guerra Civil entre 1936 e 1939 (Bellotto, 1992; Oliveira, 2015), neste intervalo, a Hospedaria não registra vinda numerosa de imigrantes da Espanha. Possivelmente, por ser um movimento mais politizado e relativamente mais recente, o fluxo causado por este conflito já não passou pela Hospedaria.

Testa-se o impacto deste fenômeno climático, que expulsou grandes contingentes populacionais da Espanha, um fator exógeno ao ciclo econômico paulista, na diversidade de origem dos fluxos de imigrantes. Os municípios que estavam nas extremidades das linhas férreas durante a seca de 1893-1897, teriam recebido parcela de espanhóis maior do que em outros períodos. A Seção 1.4, a seguir, traz os resultados da estimação.

Portugueses

Em um contexto similar ao do vizinho ibérico, Portugal participou timidamente do processo industrializante na Europa do século XIX. Essencialmente agrícola, a numerosa população não encontrava emprego, nem no campo nem na cidade (Reis, 1987). Além disso, outros fatores de repulsão dos portugueses foram a instabilidade política, a reforma do serviço militar obrigatório e a iminência de conflitos na Europa e nas colônias africanas, além das restrições à liberdade de imprensa.

Em fevereiro de 1908, o Rei Dom Carlos e o príncipe herdeiro Luís Filipe foram assassinados em uma praça de Lisboa. O atentado, embora não tenha sido levado a cabo pelos republicanos, terminou abreviando a monarquia. O novo monarca, Dom Manuel II, que tinha apenas 18 anos ao assumir o trono, foi obrigado a abandonar Lisboa menos de 18 meses depois. Os combates em plena capital deixaram cerca de 50 mortos e a família real buscou exílio. O regime republicano, recém-implantado em outubro de 1910, promoveu reformas na estrutura do Exército, herdada da monarquia. Logo em março de 1911, o governo provisório publicou decreto determinando profundas alterações no sistema de recrutamento militar, que passou a não permitir a remissão em dinheiro do serviço militar (Sena, 2010).

Em paralelo, a iminência de conflitos, tanto na Europa quanto nas colônias africanas, levou mais portugueses a emigrar (Baganha, 1991). Além disso, uma nova

lei da imprensa foi aprovada ainda em outubro de 1910, com a intenção de se restituir a liberdade de expressão. As dificuldades que o novo regime enfrentava, contudo, levaram o governo provisório a impor um conjunto de medidas que justificaram a apreensão de publicações que ultrapassem as instituições republicanas e a segurança do Estado. Como consequência, as medidas provocaram o fechamento de dezenas de jornais cujos proprietários foram presos e deportados (Franco, 1993).

Esta sequência de eventos históricos, em conjunção com a pobreza que assolava a Europa mediterrânea, ainda alijada do processo de industrialização em marcha nas porções mais setentrionais do continente, provocaram grande fluxo emigratório de portugueses, mas não só para o Brasil, mas sobretudo para os Estados Unidos e Argentina (Leite, 1987).

Ainda que seja documentado entre os historiadores uma alta propensão para a emigração clandestina (Baganha, 1991), o fluxo de origem portuguesa é o terceiro maior da Hospedaria. A maior onda de matrículas de portugueses registrada pela Hospedaria dos Imigrantes nos seus quase oitenta anos de funcionamento ocorre justamente entre 1909 e 1913. Neste quinquênio, 54,4 mil portugueses entraram no Brasil pela Hospedaria, o que equivale a 34% do total de portugueses matriculados no período 1882-1958. A média nestes seis anos, de pouco mais de 9 mil imigrantes por ano, foi praticamente seis vezes maior que a dos seis anos anteriores (aproximadamente 1,5 mil imigrantes de nacionalidade portuguesa).

Neste contexto, foi utilizada uma variável dummy no painel anual de número de imigrantes por nacionalidade de origem, município de destino $Imig_{(p,t,m)}$ que assumiu valor unitário para as linhas em que o imigrante era de nacionalidade portuguesa e, concomitantemente, em anos de chegada entre 1909 e 1913. É esperado que os municípios situados nas franjas das ferrovias neste período tenham recebido maior proporção de portugueses. A Seção 1.4 traz os resultados da estimação.

Japoneses

A emigração japonesa tem início no contexto da profunda reforma promovida pelo imperador Meiji a partir de 1873. O novo governo visava a modernização da sociedade, o que incluía abertura ao estrangeiro e alteração da estrutura agrária. A redistribuição de terras e elevação dos impostos resultou em maior endividamento das famílias. A emigração japonesa teve início na década de 1880, incentivada pelo governo. Ela era tida como solução para se reduzir a pressão interna, em função do aumento da densidade demográfica (Carneiro e Takeuchi, 2010).

Os japoneses enviaram colonos, principalmente, para os Estados Unidos (Havaí), México e Peru. A imigração dos japoneses aos Estados Unidos teve início após a assinatura de um tratado em 1894, no qual os americanos garantiam aos japoneses os mesmos direitos que os cidadãos dos EUA. Em 1906, no entanto, a Junta de Educação de São Francisco passa a segregar crianças asiáticas em escolas separadas. Somente em fevereiro de 1907 o governo federal concordou em pressionar San Francisco a retirar a medida, e o governo japonês prometeu restringir a numerosa imigração de trabalhadores aos Estados Unidos (Daniels, 1999). Estes acordos de reconciliação, conhecidos como os *Gentlemen Agreements*, não resolveram a questão e o tratamento dos residentes japoneses nos EUA continuou a causar tensão entre as duas nações. A Lei de Terras Estrangeiras (*Alien Land Act*) de maio de 1913, em adição, proibiu os japoneses de possuir ou arrendar terras por mais de três anos (Masuda, 2009). As dificuldades impostas pelos Estados Unidos aos japoneses, por outro lado, levou as autoridades japonesas a pressionar o Brasil por melhores condições. Havia também restrições à imigração asiática no Canadá. A América do Sul era vista pelos orientais como alternativa interessante, onde igualmente havia possibilidade de colonização.

A chegada dos japoneses ao Brasil só foi autorizada com a aprovação da Lei de Imigração e Colonização, em 1907 (Hashimoto, 2008). A legislação regularizou a entrada de imigrantes independentemente de sua origem, o que removeu as restrições do Decreto n.528 de 1890, que proibia a entrada de africanos e asiáticos. Por muitos anos se propagou no Brasil o antiniponismo. A política de imigração visava a civilizar e branquear a população (Schwarcz, 1993; Skidmore, 1976; Seyferth, 1985) e, nesse sentido, a vinda de asiáticos não era desejável. Os japoneses, em específico, eram vistos como de difícil assimilação, por seus costumes e religião.

A imigração japonesa, nesse sentido, enfrentou condições mais restritivas que outras nacionalidades. Só eram permitidos embarques de famílias. Japoneses sozinhos não podiam emigrar, o que não acontecia com outras etnias. Outra característica particular foi o agenciamento e promoção da imigração por empresas privadas, as *kaisha*. O surgimento das companhias de desenvolvimento exterior foi, no início, desordenado, mas a sua atuação tinha o apoio do governo japonês, interessado em promover a emigração. Com o final da Guerra Russo-Japonesa, em 1905, a crise em consequência dos altos custos do conflito e o regresso dos soldados ao arquipélago ampliaram a tensão social que os desempregados poderiam causar no país (Takeuchi, 2007, p.15). Posteriormente, no entanto, as companhias tiveram papel relevante no sucesso das colônias ultramarinas japonesas. Elas eram responsáveis pela promoção, recrutamento, transporte e, inclusive, pelo assentamento das famílias nos destinos. O governo japonês teve papel importante no sucesso do modelo das *kaishas*, tanto que mais adiante, em 1917, estatizou as empresas, unificando-as para ter melhor controle (Takeuchi, 2007). Em reação às medidas restritivas impostas pelos Estados Unidos e, para garantir o sucesso das colônias no Brasil, o governo japonês costurou, em 1911, novos termos para a fracassada Lei de Imigração e Colonização de 1907. O Decreto nº 9.081, de 3 de Novembro de 1911, regulamentou em nível federal a atuação de empresas de viação estrangeiras e companhias particulares responsáveis por expedição de títulos, distribuição e comercialização de lotes, rurais e urbanos.

Com esta mudança, os japoneses estavam autorizados a se tornarem proprietários de terra. Surgiram as primeiras colônias de povoamento, como a de Monção e as de Registro, Sete Barras e Katsura, no Vale do Ribeira (Takeuchi, 2007, p.31).

Outro período de maior imigração japonesa para o Brasil foi entre 1925 e 1930, como um resultado das severas restrições impostas pelos Estados Unidos com a aprovação do *Emergency Quota Act*, em 1921, e do *Immigration Act*, em 1924. Com o objetivo explícito de “preservar o ideal da homogeneidade”¹³, a legislação americana instituiu cotas por país origem, de acordo com as proporções de imigrantes no país. Canadá e Austrália também promulgaram leis similares. Entre 1908 e 1924, a média anual de japoneses registrados na Hospedaria de São Paulo passou de cerca de 900 para uma média anual de mais de 10 mil entre 1925 e 1930.

Diante deste contexto, o artigo estima o impacto deste conjunto de alterações nas legislações japonesa, americana e brasileira, ocorridas entre novembro de 1911 e maio de 1913, na composição de origem dos fluxos imigratórios com destino a São Paulo a partir deste período. Para isto, utiliza-se uma dummy que assume valor unitário para as linhas do painel nas quais a nacionalidade de origem era japonesa e, concomitantemente, o ano de chegada era superior a 1912. Esta dummy foi interagida também com as dummies de presença ferroviária, para se medir o impacto deste choque exógeno na chegada dos imigrantes às franjas do território paulista à época. O Apêndice traz mapas que ilustram a chegada dos japoneses a cada quinquênio. Nota-se que os primeiros só chegam a partir de 1908 e que estes se dirigem aos municípios mais distantes, relativamente, da capital. Como os imigrantes japoneses chegaram pelo menos 20 anos depois dos primeiros vindos de outras nações, estes foram levados a municípios mais a oeste. Os resultados das estimações estão nas Tabelas 1.5 e 1.6 e são analisados em detalhe na Seção 1.4.

¹³De acordo com o *United States Office of the Historian*, disponível em: <https://history.state.gov/milestones/1921-1936/immigration-act>.

Alemães

A imigração alemã para o Brasil tem início na década de 1820, com destino, principalmente, à serra fluminense e ao Vale do Rio dos Sinos, no Rio Grande do Sul. O fluxo, ainda que tímido em comparação com os do final do século, foi perene e, na década de 1840, se dirigiu a outras regiões, tais como Santa Catarina e Espírito Santo (Mauch e Vasconcellos, 1994). Em 1859, no entanto, os prussianos tiveram a emigração para o Brasil proibida pelo governo. O *Decreto Heydt* foi uma reação direto à revolta de Ibicaba, uma fazenda de café em São Paulo. Imigrantes de diversas nacionalidades europeias revoltaram-se contra as condições de trabalho nas plantações de propriedade do Senador Vergueiro. Com a Unificação da Alemanha, em 1871, a proibição foi estendida para todo o país. O Decreto é considerado pela historiografia como determinante para que o Brasil buscasse novas fontes de imigrantes em regiões mais pobres da Europa, como Itália e Polônia (Lesser, 2013).

Os registros da Hospedaria de Imigrantes de São Paulo só estão disponíveis a partir de 1882. Este corte temporal, no entanto, permite a análise do período de maior fluxo (Schäffer, 1994). Os dados da Hospedaria registram um forte aumento na chegada de imigrantes alemães entre 1889 e 1893. Nestes cinco anos foram matriculados, em média, 1708 imigrantes teutos, contra médias anuais de apenas 207 e 416 nos quinquênios imediatamente anterior e posterior, respectivamente. Este surto emigratório é imediatamente posterior à tripla troca no trono no Império Alemão, em 1888. Os *kaisers* Guilherme I, Frederico III e Guilherme II se sucederam em um período menor que três meses. A queda do chanceler Bismark, estadista à frente da Prússia desde 1862 e um dos responsáveis diretos pela unificação, resultou em um período de extrema instabilidade, política e econômica. O evento ficou conhecido na historiografia alemã como o *Dreikaiserjahr*, ou o Ano dos Três Imperadores (Blackbourn e Eley, 1984; Berghahn, 2005). Outro período no qual a imigração se intensificou foi a primeira metade da década de 1920, a partir da instabilidade polí-

tica dos primeiros anos da República de Weimar (1919-1933). Entre 1920 e 1924, a Hospedaria registrou média anual de matrículas da ordem de 1732 alemães, ante uma média de apenas 34 ao ano no quinquênio que o antecedeu, pois este coincide com a Primeira Grande Guerra. A média anual de alemães na Hospedaria diminuiu para 202 no quinquênio 1925-1929.

Estes três períodos conturbados (Decreto Heydt, a tripla troca de imperadores e a transição para a República de Weimar) causaram forte fluxo emigratório na Alemanha, no entanto, o corte temporal da amostra de registros da Hospedaria de São Paulo nos restringe a testar apenas a hipótese dos anos 1889-1893. O artigo estima, neste contexto, o impacto deste conflito na diversidade de origem do fluxo imigratório para os municípios pertencentes à fronteira agrícola de São Paulo no período em questão (de 1889 a 1893). É esperado que, nos anos do conflito, haja um aumento do fluxo de alemães com destino aos municípios da fronteira agrícola paulista. As Tabelas 1.5 e 1.6 trazem os resultados do painel, que são analisados adiante, na Seção 1.4.

Fatores de outras nacionalidades

Ainda que a maioria dos imigrantes, cerca de 3 dos 4,1 milhões estimados pelo IBGE, tenha vindo dos cinco maiores países emissores (Itália, Portugal, Espanha, Japão e Alemanha), outros 600 mil eram de outras 70 nacionalidades (Bassanezi et al., 2008; Museu da Imigração, 2016). Numerosos foram os austríacos. Há registros da chegada de 30,2 mil. Em menor número, chegaram ao Brasil os russos, como consequência da Guerra Russo-Japonesa (1904-5) e das Revoluções de 1905 e 1917, franceses, belgas, suecos, suíços, sírios e libaneses. Nas décadas de 1910 e 20 são registradas entradas de romenos, lituanos, sírios, iugoslavos, húngaros, gregos, poloneses, entre outros (Moura e Paiva, 2008).

Para os austríacos, franceses e poloneses, em específico, o artigo testou os efeitos das condições climáticas nos volumes de imigração registrada na Hospedaria de São Paulo. Solomou e Wu (1999) mostram que entre um e dois terços da variação total na produção agrícola é explicada por choques climáticos e Feng, Krueger e Oppenheimer (2010) documentam forte relação entre choques climáticos e migração internacional no período contemporâneo. Para a determinação dos períodos de cada choque, o artigo utiliza dados históricos de temperatura e precipitação de Luterbacher et al. (2004) e Pauling et al. (2006), nos moldes do utilizado por Nunn, Qian e Sequeira (2017).

Para os casos da imigração russa e grega, testou-se a hipótese de que as Guerras Russo-Japonesa contemporânea à Revolução de 1905 (Neto, 2006) e o conflito conhecido como "Luta Macedônica" (Goumaris, 1990; Hanioglu, 2010), respectivamente, tenham provocado emigração para o Brasil nos anos subsequentes aos conflitos.

Outros países de origem que merecem destaque são a Síria e o Líbano. Dos cerca de 50,2 mil libaneses e sírios que chegaram ao Brasil até 1920, segundo estimativas de Knowlton (1961), 38% se destinaram ao Estado de São Paulo. O livro de matrículas da Hospedaria dos Imigrantes do Brás não confirma este quantitativo, registrando 597 imigrantes destas nacionalidades. A primeira matrícula de sírios é de 1891 e, de libaneses, em 1905. As hipóteses da historiografia para a não passagem pela Hospedaria são as de que estas comunidades tenham se auto organizado e, relativamente a outras origens, não participaram dos subsídios (Truzzi, 2005). "As redes sociais são instrumentos utilizados pelos imigrantes para alcançar melhores ocupações, status econômico e social na sociedade hospedeira". Além disso, "as conexões com os primeiros imigrantes proporcionam recursos que reduzem os riscos e os custos da migração para os que os sucedem" (Vilela, 2011, p.157). Uma vez que o grande fluxo de sírios e libaneses não está espelhado nos dados utilizados, este artigo

não testou qualquer hipótese determinante de emigração a partir destes países de origem em específico.

1.4 Análise Empírica

O artigo tem como objetivo documentar os padrões de ocupação territorial resultante do fluxo imigratório em massa ocorrido entre 1880 e 1920. Em particular, identificar em que medida a ocorrência de choques exógenos nos países emissores de mão-de-obra e a contemporânea expansão da rede ferroviária no território contribuem para explicar a dispersão da diversidade destes fluxos.

Como visto na Seção 1.3, na qual se detalham os fatores de repulsão em cada um dos principais países emissores de imigrantes, a composição de origem dos fluxos de imigrantes para o Brasil respondeu a diversos fatores independentes da demanda por mão-de-obra nos municípios onde havia expansão no cultivo do café. Em resumo, entre os principais fatores, estão alterações de políticas emigratórias ocorridas nos distintos países de origem (e.g., Itália e Japão), eventos idiossincráticos, tais como revoluções e guerras civis de escala local (Portugal e Alemanha são os exemplos mais claros), além de catástrofes naturais, como a forte seca que assolou o território espanhol entre 1893 e 1897. Os eventos considerados para cada nacionalidade estão listados na Tabela 1.4. Resulta desta sequência desordenada de choques que, ao longo do período em que há disponibilidade dos dados do Museu da Imigração (1882-1920), a composição de nacionalidades do fluxo imigratório teve grande variação.

[TABELA 1.4 AQUI]

A Figura A.2, no Apêndice, traz as quantidades anuais de imigrantes e os

índices anuais de diversidade dos imigrantes, de acordo com os dados das matrículas dos imigrantes na Hospedaria de São Paulo. Em particular, como se pode observar, os períodos de fluxos mais diversos (em termos dos países de origem dos imigrantes) não coincidem com os de maior fluxo quantitativo de imigrantes.

No mesmo período em que ocorre esta flutuação na variedade de origens da corrente imigratória, houve grande expansão das linhas férreas em direção ao oeste e ao norte. As primeiras linhas férreas ligaram as principais regiões produtoras de café, mas à medida que avançavam em direção aos rios Paraná e Grande, passou a ocorrer o inverso, ou seja, cidades nasciam com a ferrovia (Saes, 1981, 1996; Lamounier, 2000; Stefani, 2007; Grandi, 2013). Portanto, o *timing* da conexão à rede ferroviária das localidades determina a onda de imigrantes que ali vai se estabelecer, enquanto que fatores exógenos determinam a sua composição.

1.4.1 Modelo

Por meio de um painel de dados ao nível da célula município de destino-ano de chegada-nacionalidade de origem, regride-se o número de imigrantes $Imig_{p,t,m}$ da nacionalidade de origem p no ano $t \in \{1882, \dots, 1920\}$ com destino ao município m sobre uma interação entre (i) uma dummy $D_{n,t}$ indicando eventos exógenos de repulsão de imigrantes da nacionalidade n em um determinado período de tempo t ; (ii) uma dummy indicando presença de ferrovias $Ferr_{m,t}$, e; (iii) a interação destas duas dummies. Mais especificamente, a equação a ser estimada é definida por:

$$Imig_{p,t,m} = \alpha_p + \theta_t + \rho_m + \beta_1 D_{n,t} * Ferr_{m,t} + \beta_2 Ferr_{m,t} + \beta_3 D_{n,t} + u_{n,t,m} \quad (1.2)$$

Os termos α_p , θ_t e ρ_m são os efeitos fixos de país de origem, de ano de chegada e de município de destino, respectivamente. O termo $D_{p,t}$ é uma dummy para os anos nos quais ocorreram eventos exógenos no país de origem p e que provocaram

fluxos de emigração. Por fim, $Ferr_{m,t}$ indica presença de ferrovia no município m no ano t . A partir dos valores preditos para $\hat{Imig}_{p,t,m}$, pode-se acumular o número de imigrantes preditos com origem p e que chegaram no município m ao longo de todo o período de 1872 a 1920. A partir disso, calcula-se um índice de diversidade de origem predito em 1920 para o município m , definido por $\hat{Div}_{m,1920}$, a ser utilizado no próximo capítulo da presente tese.

1.4.2 Resultados sobre imigrantes

De maneira geral, o modelo apresentou boa capacidade preditiva. Os sinais dos coeficientes estimados estão, em geral, de acordo com o esperado. A Tabela 1.5 traz os resultados dos painéis rodados com diferentes amostras de países. Nela, todas as colunas reportam resultados da especificação incluindo os efeitos fixos de ano de chegada, de município de destino e de país de origem do imigrante, além dos tratamentos e das interações dos tratamentos com a dummy de presença de ferrovias em 1920. Na primeira coluna, o painel contém 472.680 observações, por reportar os resultados do painel com 60 países. Das 76 nacionalidades de origem para as quais há registros nos livros de matrícula da Hospedaria de São Paulo, sete foram reclassificadas¹⁴ e, para outras nove, só há registros depois de 1920. A coluna (2), por sua vez, tem 126.048 observações, pois o painel contém os 16 países de origem mais frequentes.¹⁵ A coluna (3) tem 78.780 observações, pois reporta resultado do painel com os 10 países para os quais foram investigados choques exógenos de repulsão

¹⁴Os imigrantes de nacionalidade registrada como "Russos-Letões"(635) foram reclassificados como letões. Analogamente, os "Russo-Estonianos"(198) foram considerados estonianos. A imigrante registrada como "Russa-Polaca"(1) foi reclassificada como polonesa. O imigrante "Russo-Esloveno"(1) foi reclassificado como esloveno. Os declarados "Austro-Húngaros"(29) foram reclassificados como austríacos. Os "Tiroleses"(90) também como austríacos. E, finalmente, os de nacionalidade "Polonesa-Ucraniana"(77) como ucranianos. Os números entre parênteses são as respectivas quantidades de imigrantes reclassificados.

¹⁵O corte em dezesseis nacionalidades baseou-se nas quantidades de imigrantes; mantiveram-se as nacionalidades com registro de ao menos 1.000 imigrantes entre 1882 e 1920.

(Itália, Portugal, Espanha, Japão, Alemanha, Áustria, Rússia, Polônia, França e Grécia), listados na Tabela 1.4. Finalmente, a coluna (4) tem 39.390 observações, por trazer os resultados do painel com os 5 países de origem mais frequentes nos registros da Hospedaria, e para os quais há medidas de sua representatividade nas populações municipais no Censo de 1920 (Itália, Portugal, Espanha, Japão e Alemanha).

O resultado do Teste-F parcial restrito aos coeficientes da dummy de presença de ferrovia, aos dos tratamentos (choque exógenos) e aos das interações entre ferrovias e tratamentos (ou seja, o teste exclui todos os coeficientes dos efeitos fixos de nacionalidade de origem, de ano de chegada e de município de destino) é estritamente crescente à medida que se diminui a amostra para os principais países. Ele atinge 10,4 na amostra com cinco países e, por isso, esta especificação é analisada em maior profundidade na Tabela 1.6. Nota-se que a coluna (4) da Tabela 1.5 é idêntica à coluna (5) da Tabela 1.6.¹⁶

A Tabela 1.6, portanto, foi construída apenas com os resultados do painel com as cinco nacionalidades mais frequentes. Desta forma, sempre a partir de 39.390 observações (202 municípios x 39 anos x 5 nacionalidades). As colunas (1) e (2) incluem apenas os efeitos fixos de ano de chegada, de país de origem e de município de destino do imigrante e servem apenas para referência. As colunas (3) e (4) incluem os tratamentos ou choques exógenos de repulsão de imigrantes, para cada um dos cinco países. As colunas (5) e (6), finalmente, trazem as interações dos tratamentos com a dummy de presença de ferrovias.

¹⁶As colunas de 1 a 3 da Tabela 1.5 trazem os resultados das estimações para as cinco nacionalidades menos frequentes para os quais foram testados os impactos de choques exógenos. A hipótese de que um período de estiagem prolongada tenha causado emigração de contingentes populacionais da Áustria para São Paulo não é confirmada neste estudo. Também não se confirmam as hipóteses dos conflitos na Rússia e Grécia. Os coeficientes são não-significativos ou tem sinal invertido em relação ao esperado. Os coeficientes encontrados são positivos e significativos para as secas de Polônia e França. Optou-se por não utilizá-los na Tabela 1.6 porque o Teste F parcial restrito aos coeficientes da dummy de presença de ferrovia, aos dos tratamentos (choque exógenos) e aos das interações entre ferrovias e tratamentos do modelo com 7 nacionalidades ficou em 7,1, menor do que os 10,4 do modelo com os 5 principais.

Na Tabela 1.6, os coeficientes das dummies de presença de ferrovia são positivos e significantes a 1% (colunas 1, 3 e 5). Conclui-se que, a cada ano, em média, os municípios que eram cruzados por ferrovias receberam 21,1 imigrantes a mais do que as localidades que não tinham seus limites cortados por linhas férreas. Isso equivale a 1,38 vezes a média anual de imigrantes por município chegados no período 1882-1920. Quanto aos coeficientes dos efeitos fixos de linhas férreas (colunas 2, 4 e 6), a significância estatística é de, ao menos, 10%, sendo que, na maioria dos casos, é a 1%, o que denota a importância de características específicas das regiões cortadas por cada linha. Diante do documentado na literatura de histórica econômica, esperava-se impacto positivo da presença de ferrovia no município na chegada de imigrantes (Saes, 1996; Lamounier, 2000; Grandi, 2013). Desta forma, os resultados confirmam que os imigrantes se dirigiam às fronteiras do território.

Nas colunas (3) e (4), nas quais foram incluídos os termos referentes aos choques exógenos específicos de cada nacionalidade, tem-se que, com exceção do sinal do coeficiente do fator de repulsão da seca na Espanha, de 1893-7, os demais confirmaram as expectativas. Era de se esperar que a forte seca registrada no reino espanhol entre 1893 e 1897 provocasse maior êxodo populacional, mas os efeitos desta dummy em específico não foram capturados pelo modelo. Foram testadas, sem sucesso, outras janelas temporais, mais amplas, para capturar um efeito mais tardio da seca, como a 1893-1900 e a 1894-1898. A suspensão da imigração subsidiada na Itália, resultado do Decreto Prinetti, em 1902-7, por exemplo, tem efeito negativo de, em média, -27,096 imigrantes italianos por ano/município, nestes seis anos de vigência. Esta queda no fluxo de italianos equivale a 67,4% do fluxo anual médio (de 40,197 imigrantes italianos por ano e município) de italianos para São Paulo. Os conflitos em Portugal (1910-13) e Alemanha (1889-1893) provocaram forte onda emigratória e os coeficientes das dummies (+35,688 e +10,833, equivalentes a 214% e 1124% em relação às médias anuais por município, respectivamente) evidenciam o aumento destes fluxos em resposta aos fatores de repulsão. A alteração na legislação

relativa à imigração japonesa, em 1912, também teve impacto positivo no número de imigrantes para o Brasil (coeficiente +10,825, equivalente a 741% do fluxo anual médio por município).

Por último, o sinal dos coeficientes das interações entre as dummies de presença de ferrovia e as dummies de choques exógenos confirma que os imigrantes se dirigiam às fronteiras do território. Para o caso dos imigrantes italianos, tem-se que, a despeito de o fluxo ter diminuído 67% no período de vigência da portaria Prinetti, entre 1902 e 1907, continuavam a chegar imigrantes italianos às franjas da rede ferroviária. Para estes municípios com presença de ferrovias, a queda foi, inclusive marginal, de menor intensidade que se comparada ao território como um todo: os coeficientes da dummy de presença de ferrovias (21,140), do Decreto Prinetti (-33,508) e o de sua interação com a dummy de presença de ferrovias (11,293), somam -1,087, o que equivale a uma queda 2,7% no fluxo médio anual de italianos em decorrência da alteração da legislação. No que se refere aos portugueses, o coeficiente denota concentração do fluxo para as áreas atendidas pela rede ferroviária, tanto que o coeficiente específico da dummy do choque exógeno passa de 35,688 para -9,286, quando comparam os coeficientes das colunas (4) e (6). Para estes municípios com ferrovias, o fluxo de portugueses foi de 78,927 (dummy de presença de ferrovias= 21,140; + dummy da Revolução Republicana= -9,286; + interação com a dummy de choque com dummy de presença de ferrovias=67,073), o que equivale a quase 7 vezes o fluxo médio anual por município de 11,342 imigrantes por ano por município.

Conclui-se que, de maneira geral, a ocorrência dos eventos específicos de cada nacionalidade teve efeito significativo na composição de fluxos imigratórios das diversas origens para o território paulista entre 1880 e 1920. As dummies de presença de ferrovias e suas interações, por sua vez, mostram a importância da expansão da rede de linhas para a distribuição dos recém-chegados no território. A relevância de

se predizer bem esta dispersão se dá pelo fato de que ela está dissociada de quaisquer outros fatos ligados à expansão da economia local.

1.4.3 Resultados sobre a diversidade

A Figura 1.1 traz uma comparação de barras que representam, por um lado, as quantidades anuais de imigrantes registrados nos livros de matrículas da Hospedaria de São Paulo e, por outro, os índices de diversidade de origem anuais no período entre os anos de 1882 e 1920. As barras ilustram como os períodos de maior fluxo imigratório não coincidem com os de fluxo mais diverso. A Figura 1.2 mostra as proporções de imigrantes de cada uma das principais origens (Alemanha, Espanha, Itália, Japão, Portugal e uma última categoria na qual todas as outras 55 estão agrupadas) a cada ano nestas quase quatro décadas de análise.

Convém destacar o papel de cada choque exógeno na diversidade de cada fluxo anual. Em ordem cronológica, destacam-se, em primeiro lugar, o conflito na Alemanha, entre 1889 e 1893, que eleva a parcela de alemães nos fluxos deste período específico. O índice de diversidade vai de 0,13 nos três anos anteriores ao conflito para 0,42 nos quatro posteriores ao conflito que provocou a repulsão em massa de alemães. Em segundo lugar, pode-se verificar o impacto da alteração na legislação emigratória italiana de 1902. O Decreto Prinetti proíbe o recrutamento e o transporte subsidiado de imigrantes ao Brasil até 1907. Os índices de diversidade se elevam de 0,4 para 0,6 quando se comparam os quatro anos anteriores à legislação como os quatro posteriores. As proporções de imigrantes de outras origens ganham importância relativa, na realidade, para todo o período amostral subsequente a este choque. A média anual do índice de diversidade sobe de 0,4 para 0,66 na comparação 1882-1902 contra 1903-1920. Em terceiro lugar, a Figura 1.2 evidencia o aumento relativo da parcela de portugueses no período de conflitos relativos à

Revolução Republicana, entre 1911 e 1913. O grande exôdo de portugueses que fugiram dos conflitos faz com que esta origem seja responsável por 27% de todo o fluxo de imigrantes para São Paulo. Nos anos anteriores ao evento, a média anual da proporção de portugueses era de 13%. Os índices de diversidade chegam a inéditos 0,72 em 1913, com média de 0,7 entre 1911 e 1913. Por último, as barras evidenciam o ganho de importância relativa da imigração japonesa a partir de 1912, quando a conjunção de alterações nas legislações imigratórias do Japão, Brasil e Estados Unidos deram maior impulso para a vinda dos nipônicos à América do Sul. A permissão para comprar terras, recém-cassada nos Estados Unidos, elevou a proporção anual média de japoneses no fluxo imigratório para 10% do total a partir de 1912. O fluxo se iniciara em 1908, mas era tímido, não atingindo mais que 2% do fluxo anual médio (1908-1912), em razão das imposições aos japoneses até 1912, como a restrição a famílias inteiras e a impossibilidade de se constituir colônias por meio de arrendamento ou compra de terras (Takeuchi, 2007; Hashimoto, 2008).

1.5 Conclusões

As variações de quantitativo anual de imigrantes ao longo dos quase 40 anos de registros de matrícula na Hospedaria, bem como as variações de suas origens, foram resultado de inúmeros fatores ligados ao ciclo econômico. Por outro lado, os fluxos anuais também responderam a diversas alterações na legislação imigratória (brasileira e de outros receptores) e emigratória (dos países emissores).

Neste sentido, o presente artigo documenta os padrões de ocupação do território paulista e, mais especificamente, os determinantes da dispersão da diversidade de origem dos imigrantes, com base na combinação de dois conjuntos de informação histórica. Em primeiro lugar, o fato de que o fluxo de imigrantes teve grande flutuação no período 1880-1920 em sua composição/diversidade. Em particular, períodos

de maior diversidade dos fluxos não coincidem com maior fluxo quantitativo de imigrantes. Em segundo lugar, explora-se a expansão das linhas férreas. Uma vez que os imigrantes recém-chegados eram alocados nas terras recém-conectadas pela ferrovia, utiliza-se o cronograma de inaugurações das estações ferroviárias como mecanismo de predição da alocação anual dos fluxos imigratórios.

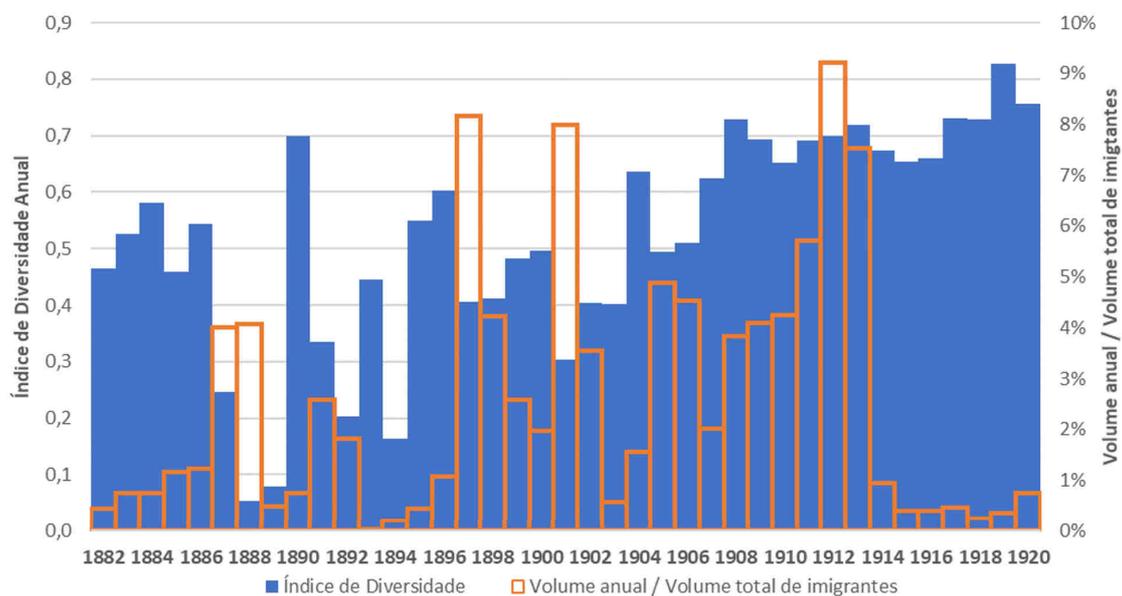
De maneira geral, o modelo apresenta boa capacidade preditiva. Especificamente para as cinco origens mais frequentes (Itália, Espanha, Portugal, Japão e Alemanha), os coeficientes encontrados tem os sinais esperados e, na maioria das vezes, significativos a 1%. As dummies de presença de ferrovias confirmam que os imigrantes se dirigiam às fronteiras do território. Nas especificações nas quais foram utilizados os efeitos fixos de linhas férreas, os coeficientes, em sua maioria, são estatisticamente significantes, o que evidencia a importância de especificidades das partes do Estado cortadas por elas.

As dummies dos choques exógenos de cada nacionalidade de origem também apresentaram coeficientes com o sinal esperado. A alteração da legislação emigratória italiana de fato reduz significamente o fluxo de imigrantes desta nacionalidade. Os conflitos em Portugal e Alemanha tem forte impacto na chegada de mais imigrantes destes países. As alterações nas legislações imigratórias referentes aos japoneses, conforme esperado, tem efeito positivo no fluxo dos asiáticos a partir de 1912. O modelo, no entanto, não capta os efeitos do choque ocorrido na Espanha. Era de se esperar que a forte seca entre 1893 e 1897 no reino espanhol provocasse maior êxodo populacional, mas os efeitos desta dummy em específico não foram capturados pelo painel. Por último, os coeficientes das interações entre as dummies de presença de ferrovia e as dummies de choques exógenos também apresentaram o sinal esperado. Este resultado mostra que, de fato, a ocorrência dos eventos específicos de cada nacionalidade teve efeito na composição de origens dos fluxos e, conseqüentemente, na dispersão da diversidade de origem dos imigrantes no território paulista, que

obedeceu o cronograma de construção das ferrovias entre 1880 e 1920.

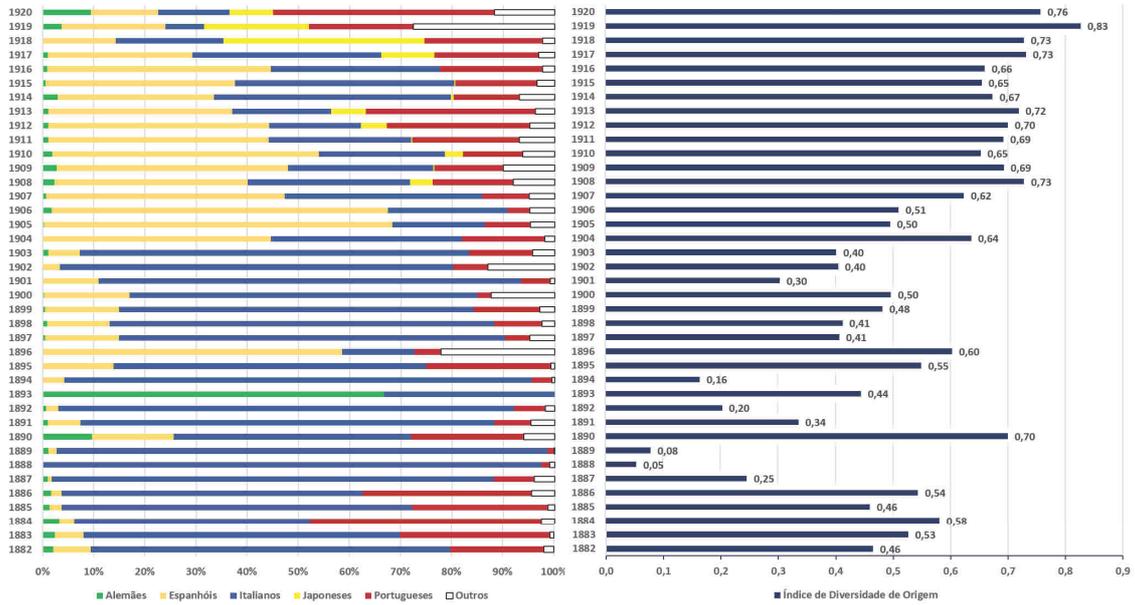
A importância de se prever bem esta dispersão se dá pelo fato de que ela está dissociada de quaisquer outros fatos ligados à expansão da economia local. Com isso, isola-se a parte causal da correlação entre diversidade e o desempenho econômico dos municípios. A diversidade predita resultante do presente artigo é insumo para o seguinte nesta tese.

Figura 1.1: Quantidades x Diversidade de Origem dos Imigrantes (1882-1920)



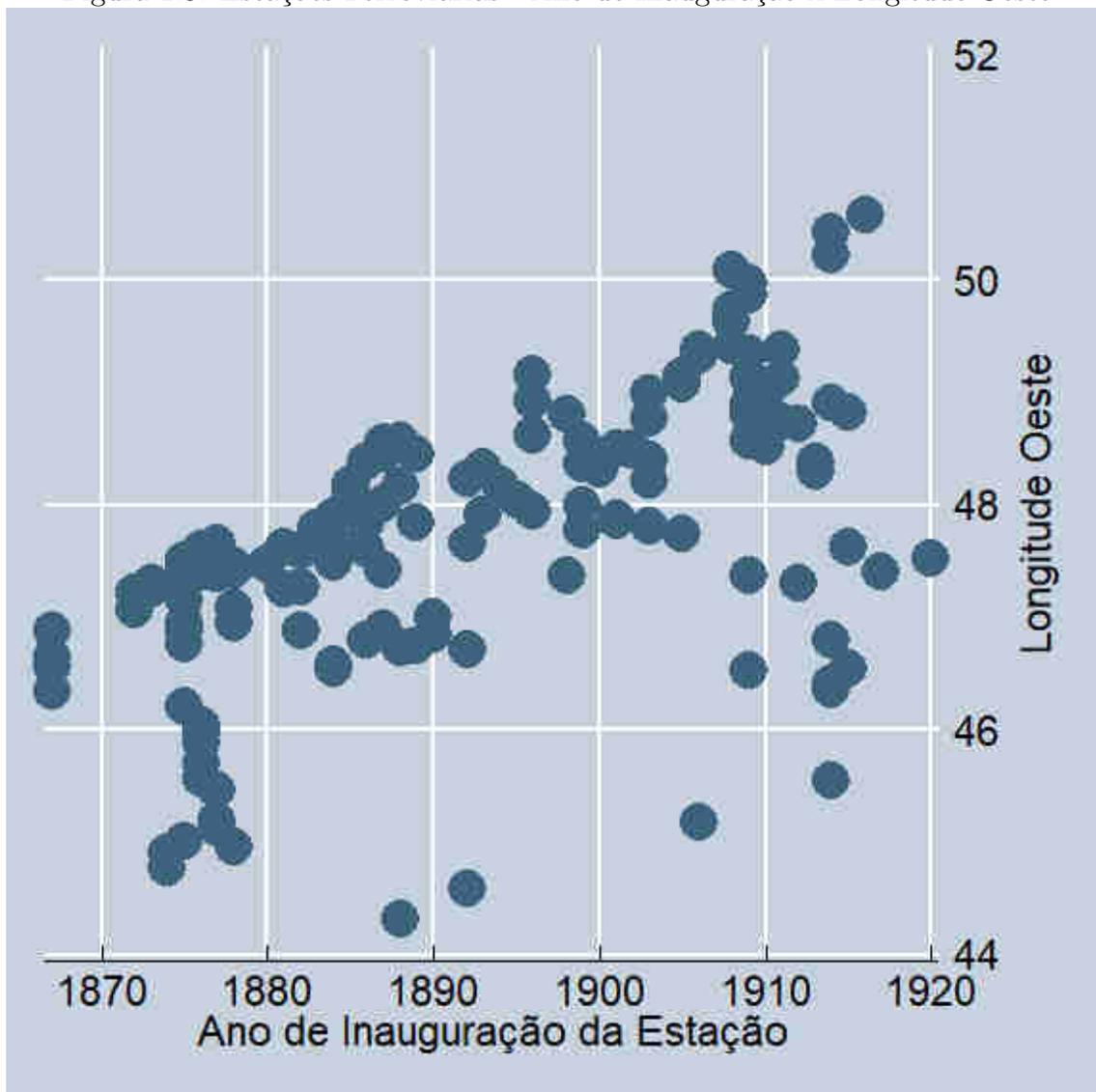
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

Figura 1.2: Diversidade de Origem dos Imigrantes (1882-1920)



Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

Figura 1.3: Estações Ferroviárias - Ano de Inauguração x Longitude Oeste



Fonte: www.estacoesferroviarias.com.br. Elaboração do autor.

Tabela 1.1: Estatísticas Descritivas

VARIÁVEIS	(1) média	(2) desv-pad	(3) min	(4) max	(5) N
A. Índices de Diversidade e População Municipal em 1920					
Índice de Diversidade (observado)	0,431	0,284	0,000	0,821	202
Índice de Diversidade (predito)	0,593	0,104	0,497	0,799	202
População em 1920 (em milhares)	183,329	818,392	2,867	11.086,8	202
B. Variáveis sobre Infraestrutura Ferroviária em 1920					
Dummy de presença de ferrovia 1872	0,024	0,156	0,000	1,000	202
Dummy de presença de ferrovia 1920	0,738	0,441	0,000	1,000	202
EF Linhas 1920	2,708	2,094	0,000	8,000	202

Nota: a amostra consiste em 202 municípios considerando as fronteiras de 1920. O indicador de presença de ferrovia para cada município foi construído com base no conteúdo do site www.estacoesferroviarias.com.br.

Tabela 1.2: Linhas Férreas por Município

LINHAS FÉRREAS	(1) período	(2) núm. municípios
Sem ferrovia		52
EF Linha 1 (Central)	1874-1915	18
EF Linha 2 (SP Railway)	1867-1915	8
EF Linha 3 (Paulista)	1872-1917	45
EF Linha 4 (Sorocabana)	1872-1920	41
EF Linha 5 (Mogiana)	1875-1912	27
EF Linha 6 (Noroeste)	1906-1909	4
EF Linha 7 (SP-Minas)	1909	1
EF Linha 8 (Araraquara)	1899-1911	6

Nota: a amostra consiste em 202 municípios considerando as fronteiras de 1920. O indicador de presença de ferrovia para cada município foi construído com base no conteúdo do site www.estacoesferroviarias.com.br.

Tabela 1.3: Quantidades e proporções de municípios cruzados por ferrovias - evolução por quinquênio (1880-1920)

Período	Número de municípios	Proporção acumulada
até 1880	40	20%
entre 1880 e 1884	14	27%
entre 1885 e 1889	18	38%
entre 1890 e 1894	9	42%
entre 1895 e 1899	11	47%
entre 1900 e 1904	11	52%
entre 1905 e 1909	21	62%
entre 1910 e 1914	19	71%
entre 1915 e 1919	5	73%
após 1920	2	74%
sem ferrovias	52	26%

Tabela 1.4: Eventos de Repulsão de Imigrantes

Anos	País de Origem	Descrição do Evento
1903-1907	Itália	Decreto Prinetti
1911-1913	Portugal	Revolução Republicana de 1910
1893-1897	Espanha	Choque climático (seca)
1912-1920	Japão	Legislação emigratória
1889-1893	Alemanha	Dreikaiserjahr (Ano dos Três Imperadores)
1908-1913	Áustria	Choque climático (seca)
1905-1911	Rússia	Guerra Russo-Japonesa e Revolução de 1905
1890-1891	França	Choque climático (seca)
1889-1893	Polônia	Choque climático (seca)
1905-1906	Grécia	Conflito "Luta Macedônica"

Fontes: Itália: Cenni (2009) e Reis-Vieira (2012); Portugal: Sena (2010), Baganha (1991) e Franco (1993); Espanha: Liria e Marcén (1981) e Espanha (2013); Japão: Takeuchi (2007), Masuda (2009) e Hashimoto (2008); Alemanha: Blackburn e Eley (1984) e Berghahn (2005); Áustria, França e Polônia: Luterbacher et al. (2004) e Pauling et al. (2006); Rússia: Neto (2006); Grécia: Gounaris (1990) e Hanioglu (2010).

Tabela 1.5: Painel - Número de imigrantes por país de origem, por ano e por município de destino - Dados do Museu da Imigração (1882-1920)

VAR. DEP.: Núm. Imigrantes	60 PAÍSES (1)	16 PAÍSES (2)	10 PAÍSES (3)	5 PAÍSES (4)
Dummy de Ferrovia	1,786 (0,264)***	7,218 (0,982)***	10,438 (1,313)***	21,140 (2,584)***
Dummy Prinetti (ITA)	-41,598 (7,023)***	-38,839 (6,560)***	-37,566 (6,236)***	-33,058 (5,402)***
Dummy Seca (ESP)	-21,985 (3,755)***	-18,347 (3,150)***	-16,154 (2,700)***	-9,653 (1,712)***
Dummy Revolução (POR)	-6,705 (1,762)***	-8,416 (2,280)***	-8,789 (2,763)***	-9,262 (4,131)**
Dummy Legislação (JAP)	1,521 (0,239)***	6,987 (0,999)***	12,132 (1,746)***	25,056 (3,670)***
Dummy Conflito (ALE)	1,084 (0,240)***	7,047 (1,231)***	11,765 (2,200)***	22,857 (4,369)***
Dummy Seca (AUS)	-0,313 (1,187)	-0,935 (1,378)	-0,728 (1,650)	
Dummy Conflito (RUS)	1,037 (1,014)	0,334 (1,194)	-0,283 (1,486)	
Dummy Seca (FRA)	1,511 (0,264)***	6,884 (1,180)***	11,126 (2,001)***	
Dummy Seca (POL)	1,929 (0,354)***	8,183 (1,473)***	12,999 (2,449)***	
Dummy Conflito (GRE)	-0,112 (0,179)	-0,635 (0,727)	-2,687 (1,252)**	
Dummy Ferr x Prinetti (ITA)	41,108 (6,797)***	33,239 (6,012)***	27,303 (5,463)***	11,281 (4,615)**
Dummy Ferr x Seca (ESP)	31,534 (5,705)***	23,402 (4,718)***	16,962 (4,077)***	0,218 (3,878)
Dummy Ferr x Revolução (POR)	93,589 (37,522)**	86,368 (36,949)**	81,428 (36,434)**	67,062 (35,209)*
Dummy Ferr x Legislação (JAP)	3,863 (2,016)*	-2,526 (1,959)	-6,778 (2,125)***	-19,477 (3,147)***
Dummy Ferr x Conflito (ALE)	-0,796 (0,461)*	-8,849 (1,747)***	-15,284 (2,979)***	-31,919 (6,153)***
Dummy Ferr x Seca (AUS)	1,351 (1,335)	-6,169 (1,413)***	-11,340 (1,896)***	
Dummy Ferr x Conflito (RUS)	0,624 (1,999)	-6,990 (1,833)***	-12,168 (2,025)***	
Dummy Ferr x Seca (FRA)	-1,039 (0,625)*	-9,187 (1,525)***	-15,684 (2,740)***	
Dummy Ferr x Seca (POL)	-2,639 (0,514)***	-10,692 (2,039)***	-17,128 (3,296)***	
Dummy Ferr x Conflito (GRE)	2,563 (3,694)	-5,353 (3,550)	-11,277 (3,691)***	
Observações	472.680	126.048	78.780	39.390
R-quadrado	0,057	0,091	0,117	0,180
Teste F-Parcial (Dummy Ferr, Tratamentos e Interações)	6,11	6,27	6,44	10,38
EF Ano	SIM	SIM	SIM	SIM
EF Município	SIM	SIM	SIM	SIM
EF País de Origem	SIM	SIM	SIM	SIM
Dummy de Ferrovia	SIM	SIM	SIM	SIM

Nota: Erros padrão robustos em parênteses, clusterizados nas fronteiras municipais de 1920. Todas as colunas reportam os resultados das regressões em MQO, nas quais a variável dependente é o número de imigrantes do país de origem 'p', a chegar no município 'm', no ano 't'. A Dummy de Presença de Ferrovia assume valor unitário se o município era cruzado por ferrovias em 1920. Todas as incluem os efeitos fixos de ano, município e país de origem do imigrante, além dos tratamentos e das interações dos tratamentos com a dummy de presença de ferrovias em 1920. A coluna (1) tem 472.680 observações por reportar os resultados do painel com 60 países. A coluna (2) tem 126.048 pois o painel contém os 16 países de origem mais frequentes. A coluna (3) tem 78.780 observações uma vez que reporta resultado do painel com 10 países. A coluna (4) tem 39.390 observações por trazer os resultados do painel com 5 países. Níveis de significância: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Tabela 1.6: Painel - Número de imigrantes por país de origem, por ano e por município de destino - Dados do Museu da Imigração (1882-1920)

VAR. DEP.: Núm. Imigrantes	EFs (1)	EFs (2)	TRAT (3)	TRAT (4)	TRAT+ (5)	TRAT+ (6)
Dummy de Ferrovia	20,868 (2,438)***		20,856 (2,438)***		21,140 (2,584)***	
EF Linha 1 (Central)		7,518 (4,328)*		7,525 (4,326)*		7,992 (4,368)*
EF Linha 2 (SP Railway)		24,850 (9,730)**		24,856 (9,728)**		25,455 (9,685)***
EF Linha 3 (Paulista)		29,698 (4,323)***		29,706 (4,325)***		30,003 (4,460)***
EF Linha 4 (Sorocabana)		15,965 (3,871)***		15,907 (3,877)***		16,183 (3,964)***
EF Linha 5 (Mogiana)		19,936 (3,810)***		19,948 (3,812)***		20,287 (3,873)***
EF Linha 6 (Noroeste)		5,945 (2,257)***		5,948 (2,257)***		5,986 (2,543)**
EF Linha 7 (SP-Minas)		3,989 (1,661)**		3,991 (1,661)**		4,001 (2,102)*
EF Linha 8 (Araraquara)		26,047 (9,132)***		26,050 (9,133)***		26,144 (9,264)***
Dummy Prinetti (ITA)			-27,081 (5,108)***	-27,096 (5,110)***	-33,508 (5,402)***	-33,078 (5,412)***
Dummy Seca (ESP)			-9,510 (2,526)***	-9,519 (2,527)***	-9,653 (1,712)***	-9,652 (1,711)***
Dummy Revolução (POR)			35,704 (20,617)*	35,688 (20,618)*	-9,262 (4,131)**	-9,286 (4,139)**
Dummy Legislação (JAP)			10,843 (2,490)***	10,825 (2,488)***	25,056 (3,670)***	25,054 (3,667)***
Dummy Conflito (ALE)			10,775 (1,983)***	10,833 (1,983)***	22,857 (4,369)***	22,925 (4,374)***
Dummy Ferr x Prinetti (ITA)					11,281 (4,615)**	11,293 (4,617)**
Dummy Ferr x Seca (ESP)					0,218 (3,878)	0,192 (3,878)
Dummy Ferr x Revolução (POR)					67,062 (35,209)*	67,073 (35,225)*
Dummy Ferr x Legislação (JAP)					-19,477 (3,147)***	-19,499 (3,134)***
Dummy Ferr x Conflito (ALE)					-31,919 (6,153)***	-31,944 (6,153)***
Observações	39.390	39.390	39.390	39.390	39.390	39.390
R-quadrado	0,175	0,175	0,178	0,178	0,180	0,180
Teste F-Parcial (Ferr, Trat, Int)	73.28	12.31	15,65	8,97	10,38	7,66
EF Ano	SIM	SIM	SIM	SIM	SIM	SIM
EF Município	SIM	SIM	SIM	SIM	SIM	SIM
EF País de Origem	SIM	SIM	SIM	SIM	SIM	SIM
Dummy de Ferrovia	SIM	NÃO	SIM	NÃO	SIM	NÃO
EF Linha Férrea	NÃO	SIM	NÃO	SIM	NÃO	SIM

Nota: Erros padrão robustos em parênteses, clusterizados nas fronteiras municipais de 1920. Todas as colunas reportam os resultados das regressões em MQO, nas quais a variável dependente é o número de imigrantes do país de origem 'p', a chegar no município 'm', no ano 't'. A Dummy de Presença de Ferrovia assume valor unitário se o município era cruzado por ferrovias em 1920. Os Efeitos Fixos de Linha Férrea assumem valores de 1 a 8 a depender de qual linha férrea cruza o município. As colunas (1) e (2) não incluem tratamentos e trazem os coeficientes dos efeitos fixos de ano, município e país de origem do imigrante e servem apenas para referência. As colunas (3) e (4) incluem os tratamentos. Apenas as colunas (5) e (6) trazem as interações dos tratamentos com a dummy de presença de ferrovias. As colunas pares utilizam a dummy de ferrovia, enquanto as ímpares utilizam efeitos fixos de linha férrea. Níveis de significância: *** p<0.01, ** p<0.05, * p<0.1.

2 DIVERSIDADE E DESENVOLVIMENTO DE LONGO PRAZO

2.1 Introdução

As grandes movimentações de contingentes demográficos alteraram a composição populacional das localidades mundo afora (Easterly e Levine, 1997; Putterman e Weil, 2010; Ashraf e Galor, 2013). A imigração tem elevado a diversidade das populações (Lazear, 1998; Putnam, 2007), assim, compreender as consequências de movimentos históricos é fundamental para nortear o desenho (ou redesenho) de políticas públicas relativas à imigração no presente. A literatura acerca da diversidade e suas consequências políticas e econômicas tem crescido na última década ressaltando seus efeitos positivos ou negativos no curto e longo prazos. Se, por um lado, alguns artigos mostram que a diversidade eleva a produtividade (Ottaviano e Peri, 2005; Kemeny, 2013), a inovação e o empreendedorismo (Beach e Jones, 2013), por outro, a diversidade pode trazer redução na provisão de bens públicos, resultado, basicamente, de diferenças nas preferências entre grupos (Easterly e Levine, 1997; Alesina, Baqir e Easterly, 1999) e mais corrupção, pois a falta de coordenação entre os tomadores independentes de propina de diferentes origens elevaria a propina por unidade de produto (Shleifer e Vishny, 1993; Mauro, 1995).

Este artigo investiga os efeitos econômicos de longo prazo da diversidade de local de nascimento ao explorar o fenômeno da imigração em massa para o Brasil do fim do século XIX e início do XX. Mais especificamente, busca responder se os municípios do Estado de São Paulo de população mais diversa, posteriormente à onda imigratória registrada a partir da década de 1880, tiveram melhor desempenho econômico de longo prazo. A principal variável dependente em análise é a renda

per capita em 2000. A estimação por meio do método de mínimos quadrados ordinários resulta em coeficientes viesados, uma vez que, (i) pode haver causalidade reversa, ou seja, as cidades mais desenvolvidas podem ter atraído diversidade; (ii) há inúmeras variáveis omitidas, não-observáveis, que explicam renda de longo prazo e são potencialmente correlacionadas com o fluxo imigratório, e; (iii) há potencial erro de medida na variável independente. O artigo faz uso, então, de uma estratégia empírica que isola a variação trazida pela diversidade com o uso de um instrumento construído a partir dos fluxos anuais de imigrantes observados ao longo de 40 anos. Eventos históricos ocorridos nos países de origem e, assim, independentes do ciclo econômico do café no Estado de São Paulo, alteraram ano a ano a composição de origem dos fluxos de imigrantes. Estes fluxos de distintas composições de origem a cada ano são distribuídos no território do Estado, à medida que as ferrovias avançavam. A variação da diversidade de origem ao longo do tempo, interagida com um vetor de expansão das ferrovias em direção ao oeste do estado, permitiu a criação de uma medida de diversidade predita para cada município, que foi utilizada como variável instrumental para a diversidade de nacionalidades observada em 1920.

O artigo mostra que a diversidade de origem dos imigrantes tem efeito positivo na renda per capita de longo prazo. Testes de robustez foram implementados para que se tivesse segurança em afirmar que os resultados não são influenciados por outros fatores. Em primeiro lugar, os resultados são relativamente estáveis à inclusão de controles geográficos, fixos no tempo, de características socioeconômicas de 1872, prévias ao fluxo de imigrantes, e de dummies que denotam a presença de ferrovias. Também permanecem estáveis quando se controla pelas proporções de estrangeiros em 1872 e 1920. A manutenção dos resultados após a inclusão destes controles é importante porque não se pode confundir o efeito da variedade de origens com o efeito da escala da onda imigratória. A intenção é eliminar a possibilidade de a diversidade capturar a variação que, na verdade, seria resultado da escala do fluxo de imigrantes. A variedade de origens continua explicando as diferenças de

performance, mesmo quando se consideram os volumes relativos de estrangeiros no município. Em segundo lugar, os resultados se mantêm quando se controla por proxies de normas sociais ou culturais associadas às nacionalidades mais comuns (as que constam do Censo de 1920: italiana, portuguesa, espanhola, japonesa e alemã) e religiões (católica e protestante). Por último, os resultados permanecem positivos, mesmo quando se usa um instrumento alternativo, calculado apenas com as interações entre os fatores de repulsão exógenos e a expansão das ferrovias, ou seja, o instrumento líquido dos efeitos fixos de município, ano e nacionalidades de origem do fluxo predito de imigrantes.

Dentre os mecanismos que são subjacentes a esse impacto de longo prazo, o artigo documenta quatro: na média, os municípios mais diversos em 1920 possuíam (i) maior proporção de ocupados na indústria e serviços do que no setor agrícola, em 1920 e em 2000, o que confirma a hipótese dos benefícios da diversidade para a economia regional e urbana de Jacobs (1969) e Glaeser et al. (1992); (ii) maior diversidade ocupacional (emprego industrial ou total), em 1920 e em 2000, conforme já fora sugerido por estudos teóricos de Lazear (1998) e Kim (2007) e estimado empiricamente por Alesina, Harnoss e Rapoport (2015); (iii) maior gasto municipal com educação, já na década de 1920, o que evidencia impacto positivo em relação à controvérsia de Alesina e La Ferrara (2005) sobre o efeito da diversidade na provisão de bens públicos, e; (iv) mais escolas e professores por criança em idade escolar, maior proporção de crianças matriculadas em 1920 e em 2000, e mais anos de escolaridade em 2000, corroborando os resultados encontrados por Rocha, Ferraz e Soares (2017).

O artigo contribui para a literatura empírica sobre diversidade de quatro maneiras. Em primeiro lugar, diferentemente de parte dos artigos existentes, que utilizam *cross-section* de países (e.g., Alesina, Harnoss e Rapoport, 2016; Bove e Elia, 2017), restringe a análise ao experimento natural da imigração em massa com destino ao Brasil, mais especificamente ao Estado de São Paulo. Ao limitar a análise

aos municípios desta unidade territorial relativamente uniforme, elimina eventuais diferenças institucionais e de dotação de recursos naturais que podem explicar caminhos divergentes entre regiões. Em segundo lugar, em contraste com alguns artigos recentes (e.g., Ottaviano e Peri, 2005; Ager e Brückner, 2013; Menyhert, 2018), estima os efeitos de longo prazo, dando mais tempo para a assimilação da diversidade de origens, ao menos quatro gerações para que se observem os efeitos das complementaridades entre elas. Em terceiro, a estratégia empírica, que faz uso de variáveis instrumentais, permite estabelecer uma forte relação causal do impacto da diversidade de origem, e não apenas correlações como em parte da literatura. Por último, o artigo não deixa de documentar os mecanismos através dos quais a diversidade de origem se propagou no tempo e afetou o produto, conforme omitem Ager e Brückner (2013) e Bove e Elia (2017), que não aprofundam a análise de seus canais de transmissão.

O artigo também se relaciona com uma literatura mais geral sobre o impacto da diversidade em pequenos grupos, times ou organizações. Esse impacto tem sido estudado em economia, mas também em psicologia e administração de recursos humanos (McLeod et al., 1996; O'Reilly, Williams e Barsade, 1997; Ben-Ner, Licht e Park, 2014). Neste contexto microeconômico, os resultados apresentados pela literatura são dúbios. Ben-Ner, Licht e Park, (2014), por exemplo, encontram evidência de ambos os efeitos, conciliando resultados mistos. A diversidade pode, por um lado, levar a diminuição de performance causada por uma maior dificuldade em se atingir consensos. Por outro lado, os resultados de um time mais diverso podem melhorar em razão de uma conjunção de competências e ideias variadas que se complementam (Bouwens, 2010). Entre os resultados documentados aqui, está o fato de que municípios mais diversos em 1920 apresentam maior diversidade ocupacional (geral e industrial) em 1920, mas não só, já que estes efeitos persistem até 2000.

O restante do texto se desenvolve em mais sete seções, além da conclusão.

A Seção 2.2 traz uma breve revisão da literatura relacionada, teórica e empírica. A Seção 2.3 resume o contexto histórico no qual ocorre o choque de diversidade investigado. A Seção 2.4, por sua vez, descreve o conjunto de dados utilizados e discute as potenciais medidas de diversidade, elegendo a que melhor se aplica ao contexto histórico. A Seção 2.5 explica em detalhe a estratégia empírica adotada. A discussão dos resultados e alguns exercícios de robustez estão nas Seções 2.6 e 2.7, respectivamente. Por fim, a Seção 2.8 traz a investigação dos canais de transmissão pelos quais a diversidade se transformou em prosperidade no longo prazo.

2.2 Revisão da Literatura Relacionada

2.2.1 Literatura Teórica

O efeito da diversidade no desenvolvimento econômico é comumente atribuído na literatura teórica a quatro fatores. Em primeiro lugar, uma população mais diversificada em idade ativa estaria associada a uma maior variedade de competências, que resulta em mais inovação (Acemoglu, 2011). O autor desenvolve um modelo teórico que elucida o efeito da diversidade nas inovações. O progresso tecnológico de equilíbrio resulta em pouca inovação na medida em que os incentivos do atual sistema de patentes tendem a premiar as de aplicação imediata. Uma invenção de exploração comercial de longo prazo tem retorno incerto e demorado, podendo a patente expirar em 20 anos. O planejador social deve, portanto, investir na diversidade de pesquisadores, para que se tenha um portfólio de linhas de pesquisa mais diverso e menos dependente da liderança tecnológica.

Já Alesina e La Ferrara (2005) argumentam que uma maior diversidade populacional permite a produção de uma maior variedade de bens e serviços públicos. Eles apresentam um modelo teórico simples para mostrar, no entanto, que os efeitos

esperados da diversidade na produção de bens públicos não são triviais:

$$Y_i = N_i * f(L; D_i)$$

onde Y_i é o produto do município 'i' ou de uma localidade específica; N_i é a população total do município 'i'; L é a quantidade fixa de trabalho empenhado, igual para todas as pessoas, independentemente do local de nascimento; e, D é a diversidade de países de origem dos trabalhadores (ou a quantidade de países de origem distintos) no município 'i'. Assumem-se as derivadas parciais da função-produção: $f_L > 0$ e $f_{LL} < 0$. A hipótese de que mais diversidade gera, *ceteris paribus*, mais produto se confirma caso $f_D > 0$ e $f_{DD} < 0$. Assume-se também complementaridade, com $f_{LD} > 0$. O produto é consumido pelas famílias ou usado para produção de bens públicos (g). Seja $U = u(c) + v(g, D)$, onde $u_c > 0$; $u_{cc} < 0$; $v_g > 0$; $v_{gg} < 0$. Assume-se, adicionalmente, que $v_D > 0$ e $v_{gD} < 0$, o que significa que a utilização dos bens públicos traz menos utilidade se tiver que ser compartilhada com grupos de outras origens. A restrição orçamentária, por sua vez, se expressa por $g = t * N * f(L; D)$, onde t é a alíquota do imposto de renda. Assim, o problema se resume a:

$$Max N * U = N * [u(c) + v(g, D)]$$

sujeito às restrições: (1) $N * c + g = N * f(L; D)$ e (2) $g = t * N * f(L; D)$. A condição de primeira ordem é dada por: $N * v_g(.) = u_c(.)$. Ou seja, tudo o mais constante (inclusive N , pois um crescimento da população poderia provocar variações na diversidade), o benefício marginal da taxação, em termos da produção dos bens públicos, deve se igual ao custo marginal da taxação, em termos de redução do consumo privado. Temos então que:

$$\text{sign} \{ dt/dD \} = \text{sign} \{ t * N^2 * v_{gg} * f_D + N * v_{gD} - (1 - t) * u_{cc} * f_D \}$$

Da equação acima, temos que $(dt/dD) < 0$, se v_{gD} for grande o suficiente. Assim, se o benefício marginal dos bens públicos cai com a diversidade, e então

mais diversidade pode levar a uma *menor* oferta de bens públicos (relativamente à de privados). Em outras palavras, locais com população mais diversa podem ser mais produtivos, mas o efeito nos gastos do governo (oferta de bens públicos) é ambíguo.

Em terceiro lugar, a diversidade pode ser benéfica para o desenvolvimento quando há diversidade ocupacional dos trabalhadores, que expande a potencial divisão do trabalho na sociedade, especialmente necessária para atividades mais complexas como a manufatura (Kim, 2007, p.17).

Por último, mas não menos importante, quando analisada em uma escala microeconômica, ao fato de que equipes de projeto mais diversificadas do ponto de vista étnico ou cultural teriam produção maior, quando se controla pela falta de comunicação (McLeod et al., 1996; O'Reilly, Williams e Barsade, 1997; Ben-Ner, Licht e Park, 2014). Em suma, trabalhadores de origens distintas podem ser encarados como fatores de produção diferentes, que interagem e, de alguma forma, se complementam e adicionam criatividade, ou outro componente qualquer não capturado pela função produção tradicional (Lazear, 1998). Há de se considerar, portanto, o grau de disjunção dos conhecimentos e informações trazidos por diferentes grupos. Novas ideias são criadas quando há trocas entre indivíduos com educação e acumulação de experiências distintas. Basta não haver sobreposição total do conjunto informacional e de conhecimento para que um trabalhador inevitavelmente aprenda algo com o outro (Lazear, 1998; Alesina, Spolaore and Wacziarg, 2000).¹

A teoria econômica, portanto, prevê que complementaridades nas distintas qualificações de uma mão-de-obra de origem heterogênea podem trazer benefícios para alguns processos produtivos. Por outro lado, a diversidade pode trazer descon-

¹Com frequência, ressalta-se a importância da comunicação. Se no grupo não se fala a mesma língua, as trocas são prejudicadas. Naturalmente, a barreira da comunicação se torna menos relevante em um contexto de longo prazo.

fiança mútua ou falta de coesão social, que pode reduzir a oferta de bens públicos.

2.2.2 Literatura Empírica

A literatura empírica sobre as consequências da diversidade vem ganhando corpo nas últimas duas décadas, mas tem dado ênfase à análise de curto e médio prazos. Menyhert (2018), por exemplo, utiliza a diversidade étnica como explicativa para o crescimento econômico em nível municipal entre 1880 e 1910 na Hungria. Eventos militares ocorridos entre os séculos XIV e XVIII são usados como fonte de variação exógena para a construção de instrumentos. O autor conclui que comunidades mais diversas cresceram de 20 a 60% mais rápido que municípios mais homogêneos. Os mecanismos teriam sido a diversidade setorial e ocupacional da força de trabalho. Ager e Brückner (2013), por sua vez, exploram o grande influxo de imigrantes para os EUA durante o período 1870-1920 para examinar os efeitos que fracionalização e a polarização tiveram sobre o crescimento do produto entre os condados. Uma vez que os novos imigrantes tendem a escolher locais onde já há imigrantes instalados, os autores utilizam variáveis instrumentais para estimar que os aumentos de 1 p.p. na fracionalização cultural elevam a renda per capita em 1,9%. O artigo, no entanto, não discute quais seriam os mecanismos pelos quais a fracionalização afeta o produto. No que tange aos efeitos de longo prazo, a literatura empírica acerca da diversidade é escassa.

Há uma série de artigos que utiliza um painel de países (Alesina, Harnoss e Rapoport, 2016; Alesina e La Ferrara, 2005; Bove e Elia, 2017), que traz consigo questões empíricas de variáveis omitidas, não-observáveis. Bove e Elia (2017), por exemplo, utilizam um painel com 230 países ou territórios e estimam os efeitos da diversidade de país de nascimento (e da polarização) no crescimento econômico entre 1960 e 2010. Os autores encontram coeficientes positivos estatisticamente significan-

tes para países em desenvolvimento e em períodos mais longos (40 ou 50 anos): 1 p.p. a mais de crescimento da fracionalização impulsiona em 0,1 p.p. na taxa de crescimento do produto. Para intervalos menores ou iguais a 30 anos (entre 1980-2010, por exemplo), os coeficientes não têm significância estatística. Utilizam variáveis instrumentais baseadas em um modelo gravitacional bilateral (origem-destino da imigração), porém, por ser *cross-countries*, o trabalho não discute os mecanismos responsáveis por propagar os efeitos da diversidade no produto ao longo do tempo.

Os artigos que se atêm à análise entre localidades mais homogêneas (unidades de um mesmo país ou municípios e condados de um mesmo estado) mensuram impactos para, novamente, períodos de tempo mais curtos, de no máximo 50 anos (Menyhert, 2018; Murard e Sakalli, 2018) e se concentram, basicamente, nos Estados Unidos (Ottaviano e Peri, 2005; Ager e Brückner, 2013). Além disso, nem todos discutem os mecanismos pelos quais a diversidade afeta o produto ao longo do tempo (Ager e Brückner, 2013; Bove e Elia, 2017). Com enfoque em economia regional, Kemeny (2013) analisa como a composição da força de trabalho afeta a produção de bens, serviços e ideias. Já no que se refere às áreas metropolitanas dos Estados Unidos, Ottaviano e Peri (2005) trazem evidência empírica da variação de fracionalização (uma das medidas mais comuns de diversidade) na variação da remuneração dos fatores. Segundo os autores, houve aumento significativo nos salários e no preço dos aluguéis nas cidades nas quais a fracionalização de nascidos no estrangeiro cresceu entre 1970 e 1990. O experimento natural do assentamento de refugiados turcos na Grécia ocorrido nos anos 1920 é usado por Murard e Sakalli (2018) para explicar diferenças em nível municipal de variáveis que denotam desenvolvimento de longo prazo (medidos pelos Censos de 1971 e 1991). Eles encontram relação positiva entre *proporção de estrangeiros* (e não *diversidade* de estrangeiros) na população total do município e os salários, riqueza e variáveis educacionais passados cerca de 50 anos. Documentam os mecanismos por meio dos quais a presença destes estrangeiros foi benéfica: especialização ocupacional, trazidas do país de origem (seda, lã, carpetes

e tabaco, por exemplo) e transformação estrutural (mais emprego em indústrias e serviços do que em agricultura).

Em suma, os efeitos da diversidade cultural, étnica, linguística, ou de local de nascimento no desempenho econômico (crescimento, renda per capita, produtividade dos fatores e oferta de bens públicos) já foram alvo de estudos empíricos recentes, mas sempre se procurou avaliar um horizonte de, no máximo, 50 anos. Além disso, a maioria dos trabalhos analisa painéis de países e, quando é o caso de se estudar regiões mais homogêneas, com raras exceções, o escopo estudado é o norte-americano.

2.2.3 Contribuição do Artigo

Neste contexto, o artigo examina a relação entre a diversidade de país de origem dos imigrantes e o desempenho econômico de longo prazo dos municípios de destino e contribui para a literatura já existente de três maneiras: em primeiro lugar, documenta o experimento natural da imigração em massa para o Brasil, um país essencialmente agrário ao receber o fluxo imigratório, mas que atualmente é de renda média, ao contrário de muitos artigos que se estudam os Estados Unidos (e.g., Alesina e La Ferrara, 2005; Ottaviano e Peri, 2006; Ager e Brückner, 2013), ou que concluem que os efeitos só são positivos e significativos a partir de determinado nível de renda (Alesina, Harnoss e Rapoport, 2015). Embora a literatura empírica sobre os impactos da diversidade esteja crescendo, é raro encontrar evidências para países em desenvolvimento ou territórios subnacionais. A restrição da análise aos municípios de São Paulo, uma unidade territorial relativamente uniforme, tem a vantagem adicional de eliminar eventuais diferenças institucionais e de dotação de recursos naturais que podem explicar caminhos divergentes entre regiões, como ocorre tipicamente em painéis de países (e.g., Alesina, Harnoss e Rapoport, 2015; Bove e

Elia, 2017). Em segundo lugar, estima os efeitos de longo prazo, dando mais tempo para a assimilação dos imigrantes e, ao menos, três gerações para que se observem os efeitos das complementaridades entre as diferentes origens. Este maior prazo de análise contrasta trabalhos empíricos recentes que estimam os efeitos da diversidade para prazos mais curtos, de no máximo 50 anos (e.g., Menyhert, 2018; Murard e Sakalli, 2018). Por último, o artigo não deixa de documentar os mecanismos através dos quais a diversidade de origem afetou o produto (transformação estrutural, diversidade ocupacional total e industrial, gastos do governo e métricas relacionadas à educação), como ocorre com alguns dos artigos recentemente publicados (e.g., Ager e Brückner, 2013; Bove e Elia, 2017).

2.3 Contexto Histórico

Estimativas do IBGE apontam para a entrada no Brasil de mais de 4,1 milhões de imigrantes entre 1872 e 1929, dos quais 3 milhões vindos dos cinco maiores emissores (Itália, Portugal, Espanha, Alemanha e Japão) e outros 600 mil de outras 70 nacionalidades (Bassanezi et al., 2008; Museu da Imigração, 2017). Apenas entre emigrantes europeus, as estimativas são de cerca 50 milhões com destino à América entre o início do século XIX e a Primeira Grande Guerra. Aproximadamente 11 milhões apenas para a América do Sul, dos quais 38% eram italianos, 28% espanhóis, 11% portugueses, 3% franceses e 3% alemães (Rosoli, 1992). Chegaram especificamente ao Brasil, em menor número, russos, belgas, suecos e suíços. Nas décadas de 1910 e 20, a imigração diversifica-se: são registradas entradas de romenos, lituanos, sírios, iugoslavos, húngaros, gregos, polacos, entre outros (Moura e Paiva, 2008). Para efeito de comparação, o Censo de 1872 registra população residente total de 9,930 milhões no país, dos quais 383 mil (3,8%) eram estrangeiros ou naturalizados. Em 1900, a população residente total passa a 17,4 milhões, porém os estrangeiros e naturalizados já somavam 1,279 milhão, ou 7,3% (IBGE, 2007).

A partir de meados do século XIX, a necessidade de mobilização de força de trabalho livre para a expansão da produção cafeeira impulsionou a imigração estrangeira no país. As primeiras experiências imigratórias precedem 1820, com a vinda de suíços e alemães para as colônias Leopoldina e Morro Queimado, no Rio de Janeiro e, em 1824, para a colônia Nova Friburgo. Anterior a essa experiência registrou-se apenas a da chegada de levas de açorianos à região hoje pertencente ao Estado de Santa Catarina, em meados do século XVII. No entanto, foi a partir da década de 1850 que se verificaram as iniciativas mais perenes de introdução de colonos estrangeiros.

Na província de São Paulo, onde o cultivo do café avançava em direção às terras centrais, a escassez de mão-de-obra era ainda mais crítica. Com a proibição do tráfico negreiro e a Lei do Ventre Livre, ficava cada vez mais claro que a escravidão estava com os dias contados. Manter escravos se tornou cada vez mais arriscado e custoso. Em 1881, São Paulo impõe tributação para a transferência de escravos vindos de outras províncias. Além disso, havia “as fugas, a resistência escrava, o movimento abolicionista, o temor da insurreição (...)” (Gonçalves, 2012, p.4). Neste contexto, o apoio à imigração subsidiada tornou-se uma posição pragmática dos fazendeiros paulistas. “Eles não eram nem abolicionistas nem necessariamente escravistas, mas reivindicadores de mão de obra.” (ibid, p.4).

Diante da crescente demanda internacional pelos grãos, a imigração estrangeira para o São Paulo deu-se, então, com forte intervenção estatal. Com o intuito de atrair mão-de-obra, foi promulgada a Lei provincial nº 28, de 29 de março de 1884, marco inicial da imigração subsidiada em São Paulo. Ela priorizava a vinda de famílias e homens adultos e, como instrumento de atração, estabeleceu quantias indenizatórias aos imigrantes. Criaram-se aparato estatal e organização administrativa voltados especificamente para o ordenamento e promoção da imigração. Em São Paulo, a Secretaria da Agricultura, Comércio e Obras Públicas (SACOP), criada no

mesmo ano da promulgação da primeira Constituição do Estado (1891), abrigava a “Inspetoria de Terras, Colonização e Imigração”. A Secretaria recebeu a atribuição de gerir a “Hospedaria dos Imigrantes”, localizada no bairro do Bom Retiro e, posteriormente, no Brás, em São Paulo, e que serviria de primeiro abrigo aos recém-chegados nos seus primeiros dias no Brasil. Em geral, os imigrantes se hospedavam por pouco tempo, até a regularização de seus contratos de trabalho.

Apenas entre 1880 e 1900, a província de São Paulo recebeu aproximadamente 940 mil imigrantes (Vasconcelos, 1994, p. 227). Para se ter uma ideia da dimensão deste fluxo, o Censo de 1872 aponta população total da província em cerca de 837 mil habitantes. A maioria dos imigrantes chegados a São Paulo tinha a viagem subsidiada² e já sabia o destino final no interior. Também era possível assinar contrato com fazendeiros que procuravam mão-de-obra na Hospedaria (Colistete e Lamounier, 2011; Gonçalves, 2012). O livro de matrícula da Hospedaria, cujo acervo digitalizado está disponível para consultas pela internet, registra a chegada de imigrantes de 76 nacionalidades diferentes. A próxima seção detalha este e os outros conjuntos de dados usados no artigo.

2.4 Dados

O artigo utiliza três conjuntos de dados. Primeiro, os registros coletados no momento da chegada dos imigrantes para matrícula na Hospedaria do Brás, em São Paulo. A partir deles, identificam-se as quantidades anuais de imigrantes por país de nascimento para cada município de destino. Em segundo lugar, os dados dos censos populacionais de 1872, 1920 e 2000 em nível municipal. O terceiro conjunto de dados consiste em variáveis geográficas e socioeconômicas do Ipeadata e Embrapa

²Segundo os dados do Museu da Imigração, 60% dos imigrantes chegados à Hospedaria entre 1889 e 1915 tiveram sua viagem subsidiada pelo governo brasileiro. A parcela restante é classificada como espontânea ou autofinanciada.

Solos, além de informações sobre infraestrutura de transportes.

O crescimento econômico e o adensamento populacional ao longo dos anos aumentaram a divisão administrativa no Estado de São Paulo. Enquanto em 1872 o território era dividido em 88 municípios, em 1920, eram 202 e, em 2000, havia 645. As análises aqui apresentadas utilizam a divisão administrativa de 202 municípios de 1920: os 645 municípios de 2000 são agregados conforme as fronteiras de 1920; e os dados originais dos 88 municípios de 1872 são conectados para os correspondentes territorialmente em 1920. Com este procedimento, um município de 1872 pode corresponder a diversos em 1920, mas os territórios passam a ser comparáveis no tempo, a partir de 1920 até 2000.

2.4.1 Variáveis de interesse

A variável de interesse no artigo é a diversidade de local de nascimento (*birth-place diversity*) e não, simplesmente, diversidade étnica ou cultural. A diversidade étnica e a diversidade de local de nascimento são empiricamente pouco correlacionadas (Alesina, Harnoss e Rapoport, 2016). Conceitualmente, isso ocorre, pois, pessoas que nascem em países diferentes, ainda que de uma mesma etnia, são educadas em sistemas nacionais de educação distintos. Aprendem, por exemplo, disciplinas de exatas e humanas com diferentes cargas horárias curriculares, e desenvolvem habilidades cognitivas também distintas. Além disso, uma vez reunidos em um único país, imigrantes de primeira geração formam um grupo mais diverso do que imigrantes de segunda e terceira gerações, que já cresceram falando o mesmo idioma e aprenderam uns com os outros dentro e fora da escola.

Mensurando diversidade

A maior parte da literatura empírica utiliza como medida de diversidade o índice de fracionalização (e.g., Alesina e La Ferrara, 2005; Ottaviano e Peri, 2006; Ager e Brückner, 2013; Beach e Jones, 2017):

$$Frac_{m,1920} = 1 - \sum_p (s_{m,p,1920})^2$$

Onde: $s_{m,p}$ é a proporção de imigrantes provenientes do país de nascimento ‘p’, no município ‘m’ em 1920; $s_{m,1}$, em particular, se refere aos habitantes nativos do município ‘m’. Inspirado no índice de concentração de Herfindahl-Hirschmann (IHH), a fracionalização mede a probabilidade de que dois habitantes de um município, sorteados aleatoriamente, tenham dois países de nascimento diferentes. A fracionalização zero denota diversidade zero, ou seja, ocorre quando todos os residentes do município são de um mesmo país de nascimento. Fracionalização igual à unidade, por sua vez, se dá quando cada habitante do município tem um país de nascimento diferente.

O índice de fracionalização, no entanto, tem duas dimensões e é importante separá-las (Alesina, Harnoss e Rapoport, 2016). A primeira é a proporção de estrangeiros $(1 - s_1)$,³ que independe do país de origem e capta a escala dos fluxos de imigrantes. A segunda é a variedade intragrupo de imigrantes. Em outras palavras, a proporção de estrangeiros mede a diversidade na dimensão intergrupos (entre nativos e não-nativos) e a variedade de estrangeiros mede a diversidade na dimensão intragrupo (apenas entre estrangeiros). Ao agruparmos todos os estrangeiros $(1 - s_1)$, independentemente de seu país de origem, resumindo o somatório em apenas dois termos, s_1 e $(1 - s_1)$, o índice de fracionalização pode ser simplificado da seguinte

³Aqui, s_1 é a proporção de nativos.

forma:

$$Frac = 1 - \sum_p (s_p)^2 = s_1 * (1 - s_1) + (1 - s_1) * s_1 = 2 * s_1 * (1 - s_1)$$

Temos, assim, que a fracionalização (ou diversidade intergrupos nativos e imigrantes) é função de $(1 - s_1)$. Utilizar a fracionalização pode nos levar a confundir o tamanho ou escala do fluxo com a variedade do intragrupo (apenas entre estrangeiros).

A diversidade intragrupo tem cálculo análogo ao da fracionalização, mas independente dos nativos. Define-se s_j como a proporção de imigrantes de origem j entre o total de estrangeiros: $s_j = (s_p)/(1 - s_1)$. A diversidade (ou variedade) de imigrantes, independente do tamanho da parcela de estrangeiros, então, fica definida conforme a equação 2.1, abaixo, e será a principal variável de interesse do artigo:

$$Div_{m,1920} = 1 - \sum_j (s_{m,j,1920})^2 \quad (2.1)$$

O Livro de Matrículas da Hospedaria dos Imigrantes

Grande parte do acervo de livros de matrícula da Hospedaria dos Imigrantes do Brás, em São Paulo, foi digitalizada pela Secretaria de Cultura do Governo do Estado nos anos 1990. Atualmente, o edifício da hospedaria abriga o Museu da Imigração de São Paulo, que disponibiliza seu acervo digitalizado em sua página na internet.⁴ O acervo digital contém informações sobre aproximadamente 1,6 milhão de imigrantes que chegaram ao Brasil a partir de 1882. Apesar das estimativas encontradas na historiografia apontarem para o recebimento de cerca de 2,5 milhões de imigrantes na Hospedaria (Holloway, 1984; Vangelista, 1991; Vasconcellos, 1994; Gonçalves, 2012), só há registro sistematizado de 65% das matrículas totais.⁵ Os

⁴<http://www.museudaimigracao.org.br/>.

⁵De acordo com o Centro de Preservação, Pesquisa e Referência (CPPR) do Museu da Imigração, nem todos os livros de matrícula foram digitalizados por questão de conservação.

registros trazem os nomes e sobrenomes dos recém-chegados, bem como seu gênero, estado civil, idade, nacionalidade, profissão, religião, relações de parentesco intrafamiliares, seu último país de residência, destino no Brasil (município, estação de trem ou fazenda) e informações sobre o financiamento de sua viagem (se vieram com subsídios dos governos receptores – central ou estaduais; se via sociedades promotoras de imigração; ou se a imigração foi espontânea ou autopatrocinada). Com base neste livro de matrículas, construiu-se o banco de dados que descreve as quantidades anuais de imigrantes por país de nascimento para cada município de destino.⁶ Quando o destino do imigrante era uma estação de trem ou fazenda, foi feita pesquisa sobre a sua localização para que fosse atribuído o município de destino corretamente. Os registros cujas estações ou fazendas não puderam ser localizadas foram descartados.⁷

Foram, então, geradas 76 dummies de nacionalidade que, colapsadas por município de destino e ano, permitiram a construção de 598.728 observações de quantidades de imigrantes por nacionalidade, ano e município de destino (76 nacionalidades x 39 anos x 202 municípios). A soma das quantidades de imigrantes por nacionalidade chegados no período 1882-1920 foi usada para se calcular a proporção de cada nacionalidade s_j em relação à população estrangeira total ($1 - s_1$, retirada do Censo de 1920), para cada um dos 202 municípios. Finalmente, estas proporções foram utilizadas para se computar os índices de diversidade por município em 1920.

⁶A Tabela A.1, no Apêndice A, mostra as quantidades anuais de imigrantes dos principais países de origem mais frequentes. A Tabela A.2 traz as quantidades anuais de imigrantes dos principais municípios de destino.

⁷O Livro de Matrículas da Hospedaria dos Imigrantes contém 1.264.204 registros até o ano de 1920, dos quais 689.161 são *non-missing* na variável destino. Destes, 2.863 (0,4%) são localidades fora do Estado de São Paulo e 1.303 (0,2%) não puderam ser localizados a partir do que constava no campo descritivo da base de dados digitalizada. A amostra final é de 684.995 observações.

2.4.2 Variáveis de resultado e mecanismos

A principal variável de resultado em análise no artigo é o logaritmo natural da renda per capita em nível municipal, calculada com base nos microdados do Censo Populacional de 2000, mas adaptada à malha administrativa de 202 municípios em 1920. Os 645 municípios de 2000 são agregados conforme as fronteiras de 1920. Desta forma, um município de 1920 pode corresponder a diversos em 2000, mas os territórios passam a ser comparáveis no tempo.

Na discussão dos mecanismos pelos quais a diversidade afeta o desenvolvimento de longo prazo das localidades, apresentada na Seção 2.8, fez-se uso de dados dos Censos de 1872, 1920 e de 2000. Em primeiro lugar, estimam-se os efeitos da diversidade na transformação estrutural da economia. Utilizam-se, portanto, as proporções de ocupados por setor da economia (agricultura, indústria e serviços/comércio). Na subseção seguinte, que investiga os efeitos da diversidade na diversidade ocupacional da mão-de-obra (total e industrial), utilizam-se as proporções dos ocupados por subsetor da economia geral e da indústria de transformação. Para o índice de diversidade ocupacional total, usaram-se as proporções de ocupados por subsetor disponíveis nos dados do Censo de 2000: judiciário, educação, saúde, defesa, administração pública, atividade religiosa, profissionais liberais, agricultura, indústria, serviços e varejo (não há dados com a mesma abertura setorial em 1920). Já para o cálculo do índice de diversidade ocupacional industrial em 1920, os dados dos subsetores disponíveis são: extrativista, edificações, vestuário, madeira e móveis, metalúrgica, química e alimentos. Para o índice de 2000, os dados setoriais disponíveis são os mesmos de 1920, mas acrescidos de: máquinas, automotiva e plásticos. A terceira subseção discute o mecanismo dos gastos dos governos municipais com educação. Para isso, utiliza-se a proporção do gasto com educação, conforme dados disponíveis para o ano de 1926. Por último, para as estimações dos efeitos da diversidade na educação e acumulação de capital humano, os seguintes dados dos Censos

de 1920 e 2000 são usados: (i) número de escolas e de (ii) professores por criança em idade escolar (entre 7 e 14 anos); (iii) proporção de crianças matriculadas (entre 7 a 14 anos); (iv) anos de escolaridade (acima de 5 anos de idade), disponível apenas no Censo de 2000.

2.4.3 Variáveis de controle

Entre as variáveis de controle que o artigo utiliza estão: (i) características geográficas dos municípios; (ii) os dados do Censo de 1872, para examinar as condições socioeconômicas prévias à exposição ao choque; (iii) dummy de presença de ferrovias no município, que assumem valor unitário para os anos a partir dos quais as fronteiras do município são cruzadas por alguma linha férrea; (iv) proporções de estrangeiros em 1872 (total de estrangeiros) e 1920 (total e por nacionalidade, para as cinco principais origens de imigrantes). Abaixo, detalham-se estes quatro blocos de variáveis:

(i) Os controles de características geográficas dos municípios têm como fontes o Ipeadata e a Embrapa Solos. Utilizaram-se dummies que indicam o tipo de solo predominante no município (dentre os quatro mais comuns no Estado: argissolo, latossolo, cambissolo e espondossolo). Elas foram incluídas nas regressões como controles, para lidar com potencial heterogeneidade na qualidade do solo ou produtividade agrícola. Entre as outras variáveis geográficas estão latitude, longitude, distância da capital e elevação. Previamente ao grande fluxo de imigrantes, regiões mais a noroeste do Estado tinham menor densidade populacional. Desta forma, é importante levar em conta o relativo isolamento geográfico do município, pois ele pode ter sido fator determinante no *timing* da expansão econômica.

(ii) Os dados do Censo de 1872 descrevem as características socioeconômicas

existentes previamente ao grande fluxo de imigrantes com destino ao Estado de São Paulo. O conjunto de variáveis inclui a proporção de escravos, de estrangeiros, de crianças matriculadas e de alfabetizados na população total, bem como a proporção de ocupados por setor (agricultura, indústria e serviços/comércio) e densidade populacional.

(iii) O outro conjunto de dados refere-se à infraestrutura ferroviária. O deslocamento da fronteira agrícola e da urbanização no Estado esteve intimamente ligada à proximidade de meios de transporte necessários para o escoamento da produção exportável (Dean, 1969). A expansão da rede de ferrovias muitas vezes foi financiada com excedente de capital do café, tanto que entre os acionistas das companhias ferroviárias era frequente figurarem as famílias dos cafeicultores (Gonçalves, 2008). Construído a partir de informações históricas sobre as companhias e datas de construção das linhas e estações de trem,⁸ foram utilizadas dummies de presença de ferrovias nos municípios em 1920, bem como efeitos fixos de linhas férreas, uma vez que cada conjunto de municípios era cortado por uma linha férrea diferente, com motivações, acionistas e datas de construção distintas. Para determinar se um município estava conectado à rede ferroviária, atribui-se variável indicadora igual à unidade se o limite de um município é cruzado por pelo menos uma linha ferroviária. Da malha de 202 municípios existentes no Estado de São Paulo em 1920, 52 (ou 26%) nunca tiveram seus limites cruzados por qualquer linha férrea. A proporção dos municípios conectados era de aproximadamente 20% (40 municípios) em 1880 e atingiu 74% (149 municípios) em 1920 (vide Tabela 1.3 para a proporção a cada quinquênio). Adicionalmente, para se capturar características das distintas regiões cortadas pelas diferentes linhas, criaram-se efeitos fixos de linhas férreas. O território de São Paulo era cortado por 8 diferentes linhas em 1920 (Central, São Paulo-Jundiaí, Paulista, Sorocabana, Mogiana, Noroeste, São Paulo-Minas, Araraquara), desta forma, para cada município atribuiu-se um valor entre 1 e 8, a depender de qual linha cruzou

⁸Dados disponíveis em <http://www.estacoesferroviarias.com.br/>.

pela primeira vez os limites do município.

(iv) as proporções de estrangeiros em 1872 e 1920 foram obtidas dos Censos de 1872 e 1920. Uma outra medida da proporção de estrangeiros em 1872 foi calculada a partir dos dados do Museu da Imigração. O fluxo anual de estrangeiros registrados na Hospedaria entre 1882 e 1920 foi somado para cada município de destino. Esta soma foi dividida pela população dos equivalentes municipais de 1872. Uma vez que o mesmo território, antes dividido em 88 municípios na época do Censo 1872, foi fragmentado politicamente em 202 unidades administrativas quando do recenseamento de 1920, replicou-se a proporção populacional de 1920 para os seus equivalentes territoriais de 1872.

2.4.4 Estatísticas Descritivas

A Tabela 2.1 apresenta um resumo das estatísticas descritivas baseadas na divisão administrativas de 1920, com 202 municípios. O Pannel A traz os dados sobre diversidade, a proporção de estrangeiros na população total e a presença de ferrovias no município. O índice de diversidade apresentado refere-se à diversidade intragrupo, análogo ao de fracionalização, comumente utilizada na literatura, entretanto, independente dos nativos. Ou seja, reflete a diversidade apenas entre os estrangeiros. Assim como a fracionalização, ela assume valores entre zero e um. Zero denota diversidade zero e valor unitário, para o caso de diversidade máxima, no caso limite em que cada estrangeiro do município tem uma origem distinta. As proporções de estrangeiros em 1872 e 1920 tem como fonte os respectivos censos. Nota-se que ela passa de 1% para 13% entre 1872 e 1920. É apresentada também a proporção de estrangeiros *observada* em 1872 (sua metodologia de construção é apresentada na seção anterior). Uma vez que a taxa de natalidade era alta e o fluxo de entrada de imigrantes era volumoso, em muitas localidades, a soma dos fluxos

de estrangeiros entre 1882 e 1920 supera a população rateada para o município em 1872, resultando em quocientes maiores que a unidade. Vale notar também, ainda no Painel A, que a rede ferroviária é ampliada e atinge 74% dos municípios já em 1920, enquanto, em 1872, apenas 2% dos municípios tinham seu território cortado por linhas férreas.

O Painel B traz o resumo descritivo das variáveis geográficas utilizadas como controle. Dado que todos os municípios da amostra estão no mesmo Estado, verifica-se pouca variância em latitude, longitude e elevação. Observa-se que o latossolo é preponderante em 55% dos municípios do Estado, enquanto o argissolo em 37%. Estes são os principais tipos de solos encontrados em São Paulo e ambos são adequados para o plantio do café, a principal cultura de grãos da região no final do século XIX. Os Painéis C, D e E, por sua vez, apresentam as características socioeconômicas da amostra em três registros temporais: 1872, 1920 e 2000. Vale notar que a população média municipal dobra entre 1872 e 1920 enquanto a proporção de alfabetizados cresce apenas de 20% para 30%. Já em 2000, verifica-se situação de universalização do ensino básico, com taxa de alfabetização e proporção de matriculados em idade escolar muito próximas dos 100%. A transformação estrutural da economia entre 1920 e 2000 também merece destaque, uma vez que o setor preponderante em termos de ocupação passa do setor primário para o terciário e a indústria passa a empregar mais de 25% da população ativa (10% em 1920).

[TABELA 2.1 AQUI]

2.5 Estratégia Empírica

No presente artigo, investiga-se a relação entre diversidade de país de nascimento dos imigrantes e a renda per capita contemporânea ao nível dos municípios.

Explora-se a forte onda de imigração para o Brasil no final do século XIX e início do XX como uma fonte de variação exógena da diversidade de origem entre os municípios paulistas. Além disso, em um segundo momento, documentam-se os mecanismos através dos quais a diversidade de origem afetou o produto.

A análise começa a partir da estimação por mínimos quadrados ordinários (MQO) e pode ser representada pela equação que segue:

$$y_{m,2000} = \alpha + \beta Div_{m,p,1920} + X'_m \gamma + e_m \quad (2.2)$$

Onde $y_{m,2000}$ é a renda per capita em 2000 do município m . O termo $Div_{m,p,1920}$ é o índice de diversidade dos imigrantes definido previamente pela equação 2.1. O termo X_m é um vetor de variáveis de controle e inclui características geográficas, tais como, latitude, longitude, distância da capital, elevação e indicadores de tipo de solo. Também inclui características socioeconômicas prévias ao grande fluxo de imigrantes, extraídas do Censo de 1872, como proporção de escravos, de estrangeiros, de crianças matriculadas e de alfabetizados na população total, a proporção de ocupados por setor (agricultura, indústria e serviços/comércio) e a densidade populacional. O vetor de controles também contém os efeitos fixos das oito linhas ferroviárias que cruzavam São Paulo em 1920. Os erros-padrão das regressões são robustos e clusterizados nas fronteiras municipais de 1920. Por último, nas especificações mais completas, controla-se também pelas proporções de estrangeiros em 1872 e 1920, para se assegurar que o efeito advém da diversidade e não do tamanho/escala do fluxo de imigrantes.

Por meio desta especificação, no entanto, não se pode garantir identificação causal. É possível que municípios que tenham apresentado crescimento econômico tenham também atraído uma diversidade maior de imigrantes, ou mesmo demandado uma força de trabalho mais diversificada. Além disso, ao longo de mais de cem anos, outras variáveis (omitidas, não-observáveis) correlacionadas com imigração e composição da mão-de-obra podem ter afetado o desempenho econômico de longo

prazo. Ademais, os imigrantes muitas vezes foram direcionados a locais mais distantes, sem acesso direto às ferrovias, com solo menos propício à agricultura, ou com menor potencial econômico, e onde o preço da terra era mais baixo. Há também potencial erro de medida na variável independente, uma vez que, em razão da má conservação, não se pode observar a totalidade dos registros dos imigrantes na Hospedaria. Estes fatores fazem com que a estimação por mínimos quadrados produza estimadores viesados. Propõe-se, assim, a utilização de variáveis instrumentais como estratégia de identificação.

A estratégia em dois estágios explora os diversos fatores conjunturais que impulsionaram a emigração dos vários países emissores de mão de obra para o Brasil. É natural supor que o número anual de imigrantes chegados ao Brasil seja endógeno ao ciclo econômico do café. O que não é necessariamente endógeno à atividade econômica, no entanto, é a composição de origem destes fluxos de trabalhadores. Como visto no primeiro capítulo, a composição de origem dos fluxos de imigrantes para o Brasil respondeu a diversos eventos idiossincráticos nos países emissores (política e alterações na legislação emigratória da Itália e Japão, conflitos e revoluções em Portugal, Alemanha, Rússia, além da Primeira Guerra, e as secas e fome na Espanha, apenas para ficar nos mais importantes).

Mais especificamente, a variável instrumental a ser utilizada explora, em primeiro lugar, fatos históricos externos ao ciclo econômico do café no Brasil. Aproveita-se, aqui, da grande variação no fluxo de imigrantes no período 1880-1920, não no quantitativo geral, mas em sua composição, decorrentes das inúmeras alterações nas políticas e condições gerais de emigração em cada país de origem (como, por exemplo, estiagem prolongada, pobreza, fome, alterações na legislação emigratória e conflitos bélicos). A Figura A.2, no Apêndice, compara barras com as quantidades anuais e com os índices anuais de diversidade dos imigrantes, de acordo com os dados das matrículas dos imigrantes na Hospedaria de São Paulo. Em particular, como se

pode observar, os períodos de fluxos mais diversos (em termos dos países de origem dos imigrantes) não coincidem com os de maior fluxo quantitativo de imigrantes.

Em segundo lugar, de modo a ganharmos variação seccional, explora-se o fato histórico de que a ocupação do território paulista se deu de leste para o oeste e acompanhou a construção das ferrovias. As primeiras linhas férreas, naturalmente, ligaram as principais regiões produtoras de café ao porto exportador de Santos. No entanto, à medida em que avançavam em direção aos rios Paraná e Grande, que fazem fronteira com os Estados vizinhos do Mato Grosso do Sul e Minas Gerais, respectivamente, passou a ocorrer o inverso, ou seja, cidades nasciam com a ferrovia (Saes, 1981, 1996; Lamounier, 2000; Stefani, 2007; Grandi, 2013). Portanto, o *timing* da conexão à rede ferroviária das localidades determina ou prediz a onda de imigrantes que ali vai se estabelecer, nos moldes do explorado por Nunn, Qian e Sequeira (2017), enquanto que fatores exógenos determinam a sua composição.

Em um estágio *zero*, portanto, utiliza-se um painel de dados ao nível da célula nacionalidade de origem-ano de chegada-município de destino, na qual regride-se o número de imigrantes $Imig_{p,t,m}$, de país de origem p , no ano $t \in \{1882, \dots, 1920\}$, com destino ao município m sobre (i) uma dummy $D_{n,t}$ indicando eventos exógenos de repulsão de imigrantes de país de origem n em um determinado período de tempo t ; (ii) uma dummy indicando presença de ferrovias $Ferr_{m,t}$ no município m no ano t , e; (iii) a interação destas duas dummies.⁹ Mais especificamente, a equação a ser estimada é definida por:

$$Imig_{p,t,m} = \alpha_p + \theta_t + \rho_m + \beta_1 D_{n,t} * Ferr_{m,t} + \beta_2 Ferr_{m,t} + \beta_3 D_{n,t} + u_{n,t,m} \quad (2.3)$$

Os termos α_p , θ_t e ρ_m são os efeitos fixos de país de origem, de ano de chegada e de município de destino, respectivamente. A partir dos valores preditos

⁹A construção do instrumento utilizado no presente artigo pode ser vista com mais detalhe no primeiro capítulo.

para $\hat{Imig}_{p,t,m}$, acumulou-se o número de imigrantes preditos com origem p e que chegaram no município m ao longo de todo o período de 1882 a 1920. A partir disso, calculou-se um índice de diversidade predito em 1920 para o município m , definido por $\hat{Div}_{m,1920}$. Desta forma, tem-se o instrumento para a diversidade de origem a ser utilizado em um primeiro estágio. Por fim, no segundo estágio podemos regressir então a variável dependente $y_{m,2000}$ sobre $Div_{m,1920}$ instrumentada por $\hat{Div}_{m,1920}$:

$$y_{m,2000} = \alpha + \beta Div_{m,1920} + X'_m \gamma + e_m \quad (2.4)$$

A vantagem de se utilizar a interação das duas fontes de variação (o *timing* da construção das ferrovias e a série de eventos exógenos à demanda por mão-de-obra nos destinos) é que a variação resultante não estará correlacionada com quaisquer outros canais que podem ter afetado a renda per capita atual, a não ser via a composição populacional. Esta é a hipótese de identificação a partir da qual os resultados da próxima seção são interpretados.

A Tabela 2.2 mostra os resultados do painel. As colunas reportam os resultados das regressões em MQO. Os efeitos fixos de linha férrea assumem valores de 1 a 8 a depender de qual linha cruzou pela primeira vez os limites do município. A coluna (1) não inclui tratamentos e traz apenas os coeficientes dos efeitos fixos de ano, município e país de origem do imigrante e serve apenas para referência. As colunas (2) e (3) incluem os tratamentos (eventos exógenos ao ciclo econômico brasileiro que afetaram a composição de origem dos fluxos anuais de imigrantes). Apenas a coluna (3) inclui as interações dos tratamentos com a presença de ferrovias.

[TABELA 2.2 AQUI]

Os resultados são analisados em detalhe no primeiro capítulo, mas, em resumo, os sinais dos coeficientes estimados estão de acordo com o esperado. Todos os coefici-

entes dos efeitos fixos de linhas férreas são positivos e, ao menos, significantes a 10%, sendo que a maioria a 1%. De fato, diante do documentado na literatura de histórica econômica, esperava-se impacto positivo da presença de ferrovia no município na chegada de imigrantes. Na coluna (2), com exceção do sinal do coeficiente do fator de repulsão da seca na Espanha, de 1893-7, os demais confirmaram a expectativa. A suspensão da imigração subsidiada na Itália, resultado do Decreto Prinetti, em 1902-7, por exemplo, tem efeito negativo de, em média, -27,096 imigrantes italianos por ano/município, nestes seis anos de vigência. Esta queda no fluxo de italianos equivale a 67,4% do fluxo anual médio (de 40,197) de italianos para São Paulo. Os conflitos em Portugal (1910-13) e Alemanha (1889-1893) provocaram forte onda emigratória e os coeficientes das dummies (+35,688 e +10,833, equivalentes a aumentos de 214% e 1124%, respectivamente) evidenciam o aumento destes fluxos. A alteração na legislação relativa à imigração japonesa, em 1912, também teve impacto positivo no número de imigrantes para o Brasil (coeficiente +10,825, equivalente a um aumento de 741% em relação ao fluxo anual médio).

A Tabela 2.3 traz os resultados do primeiro estágio, no qual a diversidade predita pelo estágio zero explica a diversidade observada em 1920. Os coeficientes são estatisticamente significantes a 1% e relativamente estáveis à medida que se incluem os controles geográficos (coluna 2), de características socioeconômicas prévias ao fluxo imigratório (3) e os efeitos fixos de linhas férreas (coluna 4). As inclusões das proporções de estrangeiros em 1920 (Censo) e 1872 (Observada) também alteram o coeficiente apenas marginalmente (de 1,07 para 1,2, nos dois casos), o que significa que, condicional à proporção de estrangeiros, a diversidade de origem predita é, de fato, fortemente correlacionada com a diversidade observada, o que significa um forte primeiro estágio. Dentre as seis especificações, a que obteve o menor valor no teste F sobre o primeiro estágio foi a que controla pela proporção de estrangeiros em 1920 (32,99), bem acima da regra de bolso para identificarmos instrumentos fracos na qual se exige $F > 10$ (Staiger e Stock, 1997).

[TABELA 2.3 AQUI]

2.6 Resultados

2.6.1 Mínimos Quadrados Ordinários (MQO)

A análise dos resultados começa com as estimações em Mínimos Quadrados Ordinários (MQO). A Tabela 2.4 traz os efeitos da diversidade de país de origem na renda per capita de longo prazo ao nível do município. A primeira coluna não inclui o vetor de controles e serve apenas como referência. A coluna (2) adiciona variáveis geográficas dos municípios como controles (distância para a capital, latitude, longitude, elevação e tipos de solo). Na coluna (3) agregam-se controles por características socioeconômicas em 1872, uma descrição prévia ao maior fluxo imigratório, a partir de 1880. Dentre as variáveis incluídas no vetor de controles estão: proporção de escravos, proporção de alfabetizados, densidade demográfica, proporção de trabalhadores ocupados na agricultura, na indústria e nos serviços e varejo. O coeficiente é bastante estável, em torno de 0,35, mostrando-se robusto à inclusão dos controles. A relação entre a diversidade e o desenvolvimento de longo prazo, portanto, não é resultado de características geográficas ou de condições socioeconômicas prévias ao fluxo imigratório.

A quarta coluna adiciona os efeitos fixos de linhas férreas aos controles. São Paulo era cortado por 8 diferentes linhas (Central, São Paulo-Jundiaí, Paulista, Sorocabana, Mogiana, Noroeste, São Paulo-Minas, Araraquara), construídas e geridas por companhias distintas (Grandi, 2013). A intenção é capturar características das diferentes regiões, cortadas pelas diferentes linhas, invariáveis no tempo. As colunas (5) e (6) têm a mesma especificação que a (4), exceto pela inclusão de um controle adicional: a proporção de estrangeiros em 1920 e em 1872, respectivamente. O obje-

tivo é verificar se o efeito da diversidade na renda não se dá, na verdade, pela escala (tamanho) do fluxo imigratório. Se não se controla pela proporção de estrangeiros, pode-se creditar à diversidade (variedade de estrangeiros) um resultado que, na verdade, seria efeito do tamanho relativo do fluxo de estrangeiros no município. Após a inclusão dos controles de proporção de estrangeiros, verifica-se que o coeficiente permanece estável em torno de 0,3 e estatisticamente significativa a 1%. Ou seja, a inclusão dos controles alterou apenas marginalmente o coeficiente da variável de interesse. Conclui-se que, condicional à proporção de estrangeiros, em média, os municípios com um desvio-padrão a mais de diversidade em 1920 (0,284) atingiram renda per capita aproximadamente 8,1% maior em 2000 ($0,285 \times 0,284 = 0,081$). Estes resultados devem ser interpretados com cuidado. No Censo de 1920, a identificação dos estrangeiros se dá pela nacionalidade e não pelo país de nascimento, como na maioria dos censos contemporâneos. É possível, portanto, que haja estrangeiros naturalizados brasileiros. Os problemas com a identificação destes indivíduos trazem viés de atenuação por erro de medida na estimação por mínimos quadrados ordinários.

[TABELA 2.4 AQUI]

2.6.2 Estimações com Instrumentos

A estimação pelo método de Mínimos Quadrados Ordinários (MQO), conforme discutido, pode nos levar a estimativas viesadas. Propõe-se, assim, a utilização de variáveis instrumentais como estratégia de identificação que contorna esses problemas e permite isolar o efeito causal que a diversidade de origem tem sobre o produto per capita municipal de longo prazo. Desta forma, na estimação em dois estágios, regressiu-se a variável dependente $y_{m,2000}$ sobre $Div_{m,1920}$ instrumentada por $\hat{Div}_{m,1920}$, conforme a equação 2.3.

A Tabela 2.5 traz os resultados da estimação em 2SLS. O Painel A resume o segundo estágio, no qual a variável dependente é o logaritmo natural da renda per capita em 2000. O impacto da diversidade varia pouco com a inclusão dos controles geográficos (de 0,723 para 0,644) e das características socioeconômicas de 1872 (para 0,573), sempre significativa a 1%. O resultado com a inclusão dos efeitos fixos de linha férrea está na quarta coluna e, na quinta e sexta colunas, são incluídas as proporções de estrangeiros em 1920 e 1872, respectivamente. As alterações do coeficiente são marginais, o que mostra a estabilidade dos resultados. Nas seis especificações os coeficientes são estatisticamente significantes a 1%. O coeficiente de 0,755 nos leva a concluir que municípios com um desvio-padrão a mais de diversidade em 1920 (0,284) tinham, em média, renda per capita de 21,4 a 24,5% maior em 2000. O efeito da diversidade, quando apurado por 2SLS, é mais forte que o estimado pelos mínimos quadrados ordinários (MQO). Isto se dá provavelmente em razão de um erro de medida idiossincrático na variável independente (diversidade observada), que gera viés de atenuação sobre os coeficientes estimados em MQO. O estimador de mínimos quadrados ordinários tende a subestimar o valor absoluto do parâmetro, se há erros de medida na variável explicativa. O Painel B sintetiza os resultados na forma reduzida. Os coeficientes do instrumento são, novamente, estáveis em torno de 0,9 e 1 e significantes a 1%. O Painel C repete os resultados do MQO apresentados na Tabela 2.4, apenas para efeito de comparação dos coeficientes.

[TABELA 2.5 AQUI]

2.7 Robutez

2.7.1 Estimação em 2SLS: diversidade predita apenas pelo tratamento

Os resultados apresentados até aqui evidenciam efeito positivo da diversidade de origem dos imigrantes na renda per capita de longo prazo. O artigo faz uso de uma estratégia empírica que pretende isolar a variação trazida pela diversidade com o uso de um instrumento construído a partir dos fluxos anuais de imigrantes observados ao longo de 40 anos. Eventos históricos ocorridos nos países de origem e, assim, independentes do ciclo econômico do café em São Paulo, afetaram a composição anual de origem dos imigrantes. Tais fluxos de distintas composições de origem a cada ano são distribuídas no território conforme o avanço das ferrovias. O instrumento, neste contexto, é a diversidade calculada com base no fluxo predito de imigrantes por um painel com efeitos fixos de município, ano e nacionalidade de origem. A exogeneidade do fluxo predito, no entanto, depende, fundamentalmente, da capacidade de ele retratar a distribuição territorial das distintas composições de origem dos fluxos. Esta distribuição geográfica é capturada pelo painel que contém também um termo de interação entre os fatores de repulsão (que variam por nacionalidade e tempo) e a expansão das linhas férreas (que varia por município e tempo). Ora, ao subtrairmos o fluxo predito resultante do painel que tem apenas os efeitos fixos (coluna 1 da Tabela 2.2) do fluxo predito resultante do painel com os efeitos fixos e as interações (coluna 3 da Tabela 2.2), teremos o fluxo predito somente pelas interações. Ou, como batizada nas tabelas, temos um instrumento calculado a partir de um fluxo predito com variação mais curta, mas supostamente ainda mais exógeno.

A Tabela 2.6 resume os resultados da estimação em dois estágios na qual o instrumento é justamente a diversidade predita apenas pelo tratamento (fatores de

repulsão exógenos $D_{n,t}$ x expansão ferroviária $Ferr_{m,t}$). Os Painéis A, B e C trazem, respectivamente, os resultados do primeiro, segundo estágio e forma reduzida. Embora os coeficientes das regressões tenham apresentado maior desvio-padrão, permanece a conclusão de que municípios que apresentavam um desvio-padrão a mais de diversidade em 1920 (0,284) tinham renda per capita de 12 a 14% maior em 2000, a depender da especificação (para efeito de comparação, a especificação que utiliza o primeiro instrumento chega a um efeito de 21 a 24%).

[TABELA 2.6 AQUI]

2.7.2 Estimação em 2SLS: erros padrão por *bootstrap*

Uma vez que o instrumento utilizado na estimação em dois estágios já é resultado de uma regressão (painel do estágio zero), o erro-padrão usual de dois estágios pode estar estimado com menos eficiência, pois parte da variação presente no estágio zero não está sendo aproveitada no 2SLS. Para efeito de comparação dos resultados, apresenta-se assim, nesta seção, o erro-padrão dos coeficientes de segundo estágio calculados por *bootstrap*. A Tabela 2.7 traz então os coeficientes estimados no segundo estágio, quando os resíduos são resultado da aplicação da técnica não-paramétrica de reamostragem introduzida por Efron (1979). Foram realizados 500 sorteios de amostras com reposição.

Os resultados são estáveis a realização da reamostragem. Os erros-padrão dos coeficientes da variável de interesse são marginalmente maiores, mas, para todas as especificações, quando controlado por variáveis geográficas, por características socioeconômicas prévias ao fluxo (Censo de 1872), por indicadores de presença de ferrovias e pela proporção de estrangeiros em 1920 permanecem significativos a 1%.

[TABELA 2.7 AQUI]

2.7.3 Cultura e Religião

O artigo até aqui documenta o efeito positivo da diversidade na renda de longo prazo dos municípios. Os imigrantes, no entanto, podem ter outras características importantes para o desenvolvimento econômico de longo prazo. Habilidades específicas, cultura ou normas sociais associadas com estes grupos podem também ter afetado a renda de longo prazo das localidades analisadas. Apesar de não termos como medir a maioria destas características, temos à disposição dados do Censo de 1920 acerca das proporções (na população total municipal) das cinco principais nacionalidades e dos católicos e protestantes, que podem ser testadas como controles, nos moldes do realizado por Rocha, Ferraz e Soares (2017).

Na Tabela 2.8, portanto, são reestimadas as regressões apresentadas na Tabela 2.4 (MQO) controlando pelas proporções de estrangeiros (alemães, portugueses, espanhóis, japoneses e italianos) e religiões (católicos em 1920 e protestantes em 1940).¹⁰ O coeficiente da variável Diversidade permaneceu praticamente inalterada, aproximadamente em 0,278, mesmo coeficiente encontrado nas regressões em Mínimos Quadrados Ordinários (MQO). Os resultados sugerem, portanto, que o efeito positivo da diversidade no desenvolvimento de longo prazo não foi liderado por fatores culturais associados com a presença de nacionalidades específicas ou religiões específicas. Habilidades específicas, cultura ou normas sociais associadas com estes grupos poderiam também ter afetado a renda de longo prazo das localidades.

[TABELA 2.8 AQUI]

¹⁰A proporção de protestantes não está disponível no Censo de 1920.

2.8 Canais de Transmissão

As Tabelas 2.2 a 2.8 mostram que a diversidade de origem dos imigrantes chegados entre 1882 e 1920 representou um choque positivo na renda atual das municipalidades de São Paulo. Nesta seção, documenta-se de que maneira este choque se propagou no tempo. Examinam-se os seguintes mecanismos ou canais de transmissão através dos quais a diversidade afetou a renda no longo prazo: transformação estrutural, diversidade ocupacional geral e na atividade industrial.

2.8.1 Transformação Estrutural

Na presente seção, o artigo traz evidências de que o choque de diversidade de origem dos imigrantes induziu uma transformação estrutural na atividade econômica dos municípios. As localidades com imigrantes de origens mais diversas em 1920 realizaram transferência mais rápida de empregos da agricultura para a indústria e, mais recentemente, para o setor de serviços. Essa diferença na estrutura do emprego não se verificava em 1872, antes do choque de diversidade ocorrida nas localidades a partir da década de 1880.

Uma explicação potencial para os benefícios econômicos de longo prazo da diversidade é a de que, durante os primeiros estágios do desenvolvimento industrial, a imigração forneceu uma oferta de mão-de-obra relativamente mais escolarizada, conforme documentado por Rocha, Ferraz e Soares (2017), e possivelmente, com diversidade de conhecimentos específicos, habilidades e ideias necessários para a decolagem da indústria. Este benefício seria especialmente capturado no contexto de economia regional e urbana, uma vez que as cidades são motores do crescimento econômico precisamente porque permitem pessoas com diferentes ideias a interagir (Jacobs, 1969; Glaeser et al., 1992). De fato, historiadores documentam que os imi-

grantes estavam desproporcionalmente representados na força de trabalho industrial, chegando a representar 57% da força de trabalho industrial nos Estados Unidos da década de 1880 (Hirschman e Mogford, 2009) e 81% entre os empreendedores industriais na Argentina de 1895 (Droller, 2015, p.24).

A Tabela 2.9 evidencia de que maneira a diversidade impulsionou a transformação estrutural da economia local no deslocamento da ocupação para segmentos de maior complexidade (indústria e serviços). O Painel A mostra que, já em 1920, municípios com maior diversidade de origem tinham menor proporção da mão de obra na agricultura e maior proporção em indústria, serviços e varejo. Antes disso, em 1872, essas diferenças eram não significativas. Em 1920, o café ainda era a atividade econômica mais lucrativa em São Paulo, correspondendo com aproximadamente 70% do total de exportações brasileiras (Abreu e Bevilaqua, 1996). Trata-se apenas, então, do resultado da composição de imigrantes que chegava a essas localidades. A indústria ainda era incipiente na economia local, basicamente provendo bens demandados pela indústria do café (Dean, 1969). O processo de transformação estrutural aprofundou-se e tornou-se vantagem econômica nas primeiras décadas do século XX, quando a industrialização decolou no território e o capital previamente acumulado pela produção do café foi investido na manufatura (Cardoso de Mello, 1984). O Painel A mostra que, quando se utiliza o índice de diversidade observado (MQO), em média, municípios com um desvio-padrão a mais de diversidade em 1920 ($0,284$), tinham 4,5 pontos percentuais a menos de ocupados no setor agrícola ($-0,158 \times 0,284 = -0,045$). Em 2000, 4,7 pontos percentuais ($-0,166 \times 0,284 = -0,047$). Ao utilizarmos a medida de diversidade instrumentada (VI), cujo desvio-padrão é de $0,104$, os efeitos estimados são parecidos: $-3,5$ p.p. em 1920 e $-3,7$ p.p. em 2000.

Ao mesmo tempo, de acordo com os resultados apresentados nos Paineis B e C, quando se utiliza o índice de diversidade observado (MQO), conclui-se que, em média, municípios com um desvio-padrão a mais de diversidade que a média tinham,

em 1920, 2,5 pontos percentuais a mais ($0,087 \times 0,284 = 0,025$) de ocupados no setor industrial e 2 pontos a mais no setor de serviços e varejo ($0,072 \times 0,284 = 0,020$). Repetindo-se o exercício desta vez com a diversidade instrumentada (VI), conclui-se que o efeito de um 1 d.p. no índice de diversidade era de 1,8 p.p. para o emprego industrial ($0,171 \times 0,104 = 0,178$) e para o emprego no setor terciário ($0,168 \times 0,104 = 0,175$). Finalmente, quanto às regressões para o ano 2000, novamente observa-se uma ampliação do efeito da diversidade (MQO) na composição setorial dos ocupados: em média, 1 p.p. a mais na indústria e 3,6 pontos a mais no setor terciário em municípios que apresentavam um desvio-padrão a mais de diversidade em 1920. A estimação utilizando a variável de interesse predita (VI) é inconclusiva quanto ao emprego industrial em 2000, mas denota efeito significativo da diversidade de 3,3 p.p. ($0,316 \times 0,104 = 0,033$) para o emprego no setor de serviços e varejo.

[TABELA 2.9 AQUI]

2.8.2 Diversidade Ocupacional

Nesta subseção, investiga-se o efeito da diversidade de origem dos imigrantes na diversidade ocupacional do setor industrial em 1920 e 2000, nos moldes do evidenciado por Murard e Sakalli (2018) e Menyhert (2018). Quanto mais diverso é o fluxo de imigrantes, maior é o potencial de complementaridade entre as culturas, competências, profissões e ofícios trazidos de seus locais de origem. Neste sentido, maior é a potencial complementaridade entre elas (Lazear, 1998; Alesina e La Ferrara, 2005; Kim, 2007).

O Painel A da Tabela 2.10 mostra que, quando utiliza-se o índice de diversidade observado (MQO), em média, municípios com um desvio-padrão a mais de diversidade em 1920 (0,284), já apresentavam 1,7 ponto percentual a mais em di-

versidade ocupacional industrial ($0,061 \times 0,284 = 0,017$). O efeito da diversidade é similar se utilizamos como explicativa a diversidade predita pelo instrumento: 1,9 p.p. a mais em diversidade ocupacional industrial ($0,183 \times 0,104 = 0,017$). Em 2000 (Painel B), a diversidade no emprego industrial era, em média, 2,1 pontos percentuais maior para os municípios com um desvio-padrão a mais de diversidade de origem (MQO) em 1920 ($0,075 \times 0,284 = 0,021$). Ao estimar esse mesmo efeito com a diversidade predita pela variável instrumental, tem-se que a diversidade no emprego industrial era 1,5 p.p. maior, em 2000, para os municípios mais diversos em 1920 ($0,145 \times 0,104 = 0,015$). Os setores industriais usados no cálculo do índice de diversidade ocupacional industrial em 1920 são: extrativista, edificações, vestuário, madeira e móveis, metalúrgica, química e alimentos. Para o índice de 2000, os dados setoriais disponíveis são os mesmos de 1920 acrescidos de: máquinas, automotiva e plásticos. O Painel C traz os resultados utilizando a variável dependente diversidade ocupacional total em 2000. Para o cálculo deste índice, levaram-se em consideração as proporções de ocupados no judiciário, magistério, medicina, defesa, administração pública, atividade religiosa, profissionais liberais, agricultura, indústria, serviços e varejo (não há dados com a mesma abertura setorial em 1920). Quando explicado pelo índice de diversidade observado (MQO), em média, municípios com um desvio-padrão a mais de diversidade em 1920 (0,284), têm diversidade ocupacional 1,1 ponto percentual maior ($0,037 \times 0,284 = 0,011$). Por outro lado, ao se utilizar a diversidade predita pela variável instrumental, o efeito na diversidade ocupacional é de 0,8 p.p. ($0,075 \times 0,104 = 0,008$).

[TABELA 2.10 AQUI]

2.8.3 Gastos do Governo Municipal

O artigo até aqui mostra que a diversidade de origem dos imigrantes chegados ao Estado de São Paulo na virada do século XIX para o XX representou um choque positivo na renda atual das municipalidades. Nesta subseção, examina-se se há diferença significativa entre municípios mais diversos versus mais homogêneos nos gastos dos governos municipais com educação já na década de 1920 (dados disponíveis para o ano de 1926). Beach e Jones (2013), por exemplo, concluem que o aumento da diversidade leva ao impasse na alocação dos recursos das prefeituras, provocando redução nas quantias gastas em bens públicos. Maior diversidade levaria a conflitos de opinião ou preferências e dificuldade para se alcançar consensos. Carvalho Filho e Colistete (2010), por outro lado, mostram que a presença de imigrantes elevou a oferta de educação pública nos municípios do Estado de São Paulo. Também mostram que, um século depois, os estudantes destas localidades obtêm melhor desempenho em testes padronizados, além de ter maior renda per capita. Alesina e La Ferrara (2005) sugerem por meio de um modelo teórico simples que os efeitos esperados da diversidade na produção de bens públicos não são óbvios.

Neste artigo, por meio de um modelo de mínimos quadrados ordinários, regrediram-se os gastos municipais com educação em 1926 no índice de diversidade observado e no predito pelo estágio zero. De acordo com os resultados apresentados na Tabela 2.11, municípios mais diversos em 1920 investiam maior parcela do orçamento em educação do que localidades mais homogêneas. Municípios com índice de diversidade observado 0,284 maior (um desvio-padrão), em média, gastavam 0,4% do orçamento a mais em educação em 1926 ($0,015 \times 0,284 = 0,004$). Já com o índice de diversidade predito (VI), a diferença média de gastos com educação em municípios mais heterogêneos é de 0,5% ($0,049 \times 0,104 = 0,005$).¹¹ Conclui-se que, no contexto estudado aqui, e para os gastos em educação, em particular, o artigo

¹¹Coefficientes estatisticamente significantes a 5%.

indica impacto positivo a respeito da controvérsia de Alesina e La Ferrara (2005) sobre os efeitos da diversidade da população na provisão de bens públicos.

[TABELA 2.11 AQUI]

2.8.4 Escolaridade

Outro mecanismo subjacente aos efeitos econômicos de longo prazo é o acúmulo de capital humano. Nesta subseção, examina-se o potencial canal de transmissão ao se estimar os efeitos da diversidade da imigração na persistência de melhores métricas educacionais ao longo tempo. Rocha, Ferraz e Soares (2017) documentam que a chegada dos imigrantes para São Paulo na virada do século XIX representou um choque positivo de capital humano nas localidades e que este choque inicial persistiu até os dias atuais por meio de maior uso de insumos educacionais (escolas e professores por criança em idade escolar). Seria o caso de que localidades com fluxo migratório mais diverso tiveram efeitos mais significativos do que as de fluxo mais homogêneo?

A Tabela 2.12 apresenta uma série de especificações em corte transversal para 1920 e 2000. Nelas, regrediram-se quatro diferentes medidas de escolaridade no índice de diversidade de origem em nível municipal observado (MQO) e predito (VI) para 1920. Como os dados coletados pelos Censos decenais variam no tempo, estas medidas não estão disponíveis para todos os anos. Todos os coeficientes são estatisticamente significantes a 1%, com exceção ao coeficiente de proporção de crianças matriculadas em 2000, com significância de 5%.

No Painel A da Tabela 2.12, é possível ver que municípios com maior índice de diversidade de origem tinham mais escolas por criança em idade escolar em

1920 Quando explicado pelo índice de diversidade observado (MQO), em média, municípios com um desvio padrão (1 d.p. = 0,284) a mais de diversidade em 1920 tinham 9,1 pontos percentuais mais escolas por criança em idade escolar em 1920. O efeito é similar quando se utiliza a diversidade predita pela variável instrumental: 9,2 p.p. ($0,882 \times 0,104 = 0,092$). Para o ano de 2000, quando aproximadamente 90% do Estado já estava alfabetizado, o coeficiente se torna negativo, o que pode indicar que o esforço das políticas públicas esteja concentrado na redução do déficit educacional das áreas mais atrasadas do Estado. Os demais painéis da Tabela 2.11, no entanto, mostram efeitos positivos da diversidade, por MQO ou VI, em outras métricas educacionais do Censo de 2000.

O Painel B da Tabela 2.12 mostra que os municípios de maior diversidade em 1920 tinham mais professores por criança em idade escolar e que esse efeito persistiu até 2000 (coeficiente 16,349 na coluna 2, painel B). Ao utilizar o índice de diversidade observado (MQO) como explicativa, conclui-se que os municípios com um desvio-padrão a mais de diversidade em 1920 tinham, em média, +1,7 professores por criança em idade escolar em 1920 ($6,149 \times 0,284 = 1,749$) e +4,7 professores/criança em 2000 ($16,349 \times 0,284 = 4,651$). Ao usar a diversidade predita (VI), o efeito é estimado em +2,1 ($19,927 \times 0,104 = 2,078$) e +6,3 ($60,537 \times 0,104 = 6,295$) professores por criança em idade escolar, em 1920 e 2000, respectivamente.

O Painel C, por sua vez, resume os efeitos da diversidade sobre a evolução da proporção de crianças matriculadas nos municípios. Consistentemente com o choque positivo do número de escolas e professores por criança em idade escolar, em municípios mais diversos, uma maior proporção das crianças estava matriculada em 1940. E, em 2000, o choque positivo ainda persiste, mas o coeficiente é menor, uma vez que a educação básica estava praticamente universal nessa época. Dados de proporção de matriculados não estão disponíveis para o Censo de 1920. Utilizando a diversidade de MQO, conclui-se que, municípios com um desvio-padrão a mais de

diversidade em 1920 tinham 4,4 p.p. mais crianças na escola em 1940 ($0,156 \times 0,284 = 0,044$) e 0,3 p.p. em 2000 ($0,009 \times 0,284 = 0,003$). Ao utilizar a diversidade VI, tem-se que municípios cuja parcela de estrangeiros era mais diversa em 1920 (+1 d.p) tinham, em 1940, 3,3 p.p. ($0,316 \times 0,104 = 0,044$) mais crianças matriculadas que municípios mais diversos. Em 2000, o efeito é idêntico ao estimado com a diversidade do MQO: 0,3 p.p. ($0,027 \times 0,104 = 0,003$).

Finalmente, o Painel D da Tabela 2.12 mostra que nos municípios (+1 d.p.) mais diversos, a população acima de 5 anos de idade tem, em média, 0,27 anos de escolaridade em 2000¹² a mais que os municípios de população estrangeira mais homogênea ($0,932 \times 0,284 = 0,265$). Ao utilizar o a diversidade instrumentada, encontra-se efeito similar, de 0,25 anos a mais de escolaridade média ($2,354 \times 0,104 = 0,246$).

[TABELA 2.12 AQUI]

2.9 Conclusões

Nas quatro décadas contadas a partir de 1880 o Brasil e, em especial, o Estado de São Paulo, receberam forte onda de imigração estrangeira de diversas origens. Há registros de imigrantes de mais de 70 nacionalidades diferentes e, em 1920, aproximadamente 13% da população total do Estado era estrangeira.

O artigo explora o episódio e mensura o impacto da diversidade trazida pela imigração no desenvolvimento de longo prazo. Com base em dados do Museu da Imigração utilizam-se variáveis instrumentais para tratar os vieses de variáveis omitidas e de causalidade reversa. O principal fator de atração dos imigrantes foi a forte

¹²Os dados de anos de escolaridade não estão disponíveis para os Censos de 1920 e 1940.

demanda por mão-de-obra, em razão da expansão da produção cafeeira e do fim do trabalho escravo. Foi, portanto, algo endógeno ao ciclo econômico, o que reforça a hipótese de causalidade reversa. Por outro lado, os fatores de repulsão dos emigrantes (secas, conflitos e revoluções, por exemplo) são independentes do ciclo do café em São Paulo e a distribuição territorial dos fluxos de imigrantes, cuja composição anual é variável em função do *timing* do desencadeamento dos fatores de repulsão, foi determinada pelo avanço das ferrovias. Desta forma, foi possível calcular índices de diversidade de origem baseados na composição predita para cada município que recebia a ferrovia ano a ano. A expansão das linhas (só 3% dos municípios eram cortado por ferrovias em 1872 e 75% deles estavam conectados à rede ferroviária em 1920) determinava os destinos dos imigrantes, pois ela levava os trabalhadores à fronteira de expansão das plantações de café.

Este contexto empírico permitiu, portanto, a estimação não-viesada do efeito da diversidade de origem no desenvolvimento de longo prazo. A depender do método de estimação e do cálculo dos erros-padrão (por *bootstrap*, por exemplo), conclui-se que, em média, os municípios que tinham um desvio-padrão a mais de diversidade de origem em seus imigrantes chegados até 1920 possuíam renda per capita de 8 a 23% maior que os municípios de população mais homogênea. A diferença entre os coeficientes estimados com o método dos mínimos quadrados ordinários (MQO) e de dois estágios (2SLS) pode ser atribuída a erros de medida idiossincráticos na variável explicativa que gera viés de atenuação sobre os coeficientes estimados em MQO. Quando há erros de medida na variável explicativa, o estimador de mínimos quadrados ordinários tende a subestimar o valor absoluto do parâmetro.

Dentre os mecanismos subjacentes a esses efeitos de longo prazo, o presente trabalho documenta quatro: na média, os municípios mais diversos em 1920 possuíam (i) maior proporção de ocupados na indústria e serviços do que no setor agrícola, em 1920 e em 2000; (ii) maior diversidade ocupacional (emprego indus-

trial ou total), em 1920 e em 2000; (iii) maior gasto com educação já na década de 1920; (iv) dentre as métricas sobre educação e acumulação de capital humano, mais escolas e professores por criança em idade escolar, maior proporção de crianças matriculadas, em 1920 e em 2000, e mais anos de escolaridade, em 2000.

A experiência dos diversos países receptores de diversidade de imigrantes no final do século XIX e início do XX deve jogar luz no redesenho de políticas públicas referentes à imigração nos dias de hoje. Os custos no curto prazo devem ser relativizados pelos benefícios de longo prazo, que começam a ser finalmente quantificados graças à disponibilidade dos microdados e do avanço do ferramental de análise na ciência econômica.

Tabela 2.1: Estatísticas Descritivas

VARIÁVEIS	(1) média	(2) desv-pad	(3) min	(4) max	(5) N
A. Diversidade e Proporção de Estrangeiros					
Índice de Diversidade (intragrupo)	0,431	0,284	0,000	0,821	202
Proporção de Estrangeiros Censo 1872	0,012	0,016	0,000	0,081	202
Proporção de Estrangeiros Observ 1872	1,879	5,066	0,000	53,628	202
Proporção de Estrangeiros Censo 1920	0,125	0,096	0,000	0,357	202
Dummy de presença de ferrovia 1872	0,024	0,156	0,000	1,000	202
Dummy de presença de ferrovia 1920	0,738	0,441	0,000	1,000	202
B. Variáveis Geográficas					
Distância para a capital em km (log)	5,139	0,693	2,078	6,243	202
Latitude	-22,594	0,989	-25,015	-20,131	202
Longitude	-47,607	1,404	-51,448	-44,386	202
Altitude (100m)	6,023	1,866	0,010	11,980	202
Dummy de latossolo	0,546	0,455	0,000	1,000	202
Dummy de argissolo	0,374	0,435	0,000	1,000	202
Dummy de cambissolo	0,114	0,300	0,000	1,000	202
Dummy de espondossolo	0,012	0,095	0,000	1,000	202
C. Variáveis do Censo de 1872					
Proporção de escravos	0,154	0,086	0,039	0,531	202
Proporção de alfabetizados (>6 anos)	0,202	0,110	0,048	0,452	202
Proporção de crianças matriculadas	0,145	0,101	0,027	0,764	202
População (em milhares)	11,147	7,198	1,566	41,751	202
Proporção dos ocupados na agricultura	0,594	0,099	0,351	0,908	202
Proporção dos ocupados na indústria	0,109	0,044	0,020	0,244	202
Proporção dos ocupados em serviços/varejo	0,296	0,091	0,058	0,569	202
D. Variáveis do Censo de 1920					
Proporção de alfabetizados (>6 anos)	0,296	0,099	0,103	0,699	202
Escolas/criança idade escolar (x 1.000)	0,387	0,377	0,000	1,863	202
Professores/criança idade escolar (x 1.000)	9,562	6,938	0,826	45,955	202
População (em milhares)	22,635	43,101	2,917	577,621	202
Proporção dos ocupados na agricultura	0,784	0,133	0,071	0,957	202
Proporção dos ocupados na indústria	0,091	0,071	0,005	0,495	202
Proporção dos ocupados em serviços/varejo	0,125	0,072	0,035	0,549	202
Proporção de Católicos	0,849	0,182	0,098	1,000	202
Proporção de Protestantes (1940)	0,019	0,015	0,000	0,076	202
E. Variáveis do Censo de 2000					
Proporção de alfabetizados (>6 anos)	0,952	0,016	0,854	0,989	202
Proporção de crianças matriculadas	0,962	0,018	0,863	0,991	202
Escolas/criança idade escolar (x 1.000)	0,387	0,377	0,000	1,863	202
Professores/criança idade escolar (x 1.000)	82,833	21,228	19,670	152,683	202
Anos de escolaridade (> 5 anos)	5,538	0,660	3,532	7,109	202
População (em milhares)	183,329	818,392	2,867	11.086,8	202
Renda per capita (log)	5,693	0,272	4,791	6,392	202
Proporção dos ocupados na agricultura	0,197	0,134	0,003	0,595	202
Proporção dos ocupados na indústria	0,252	0,087	0,097	0,545	202
Proporção dos ocupados em serviços/varejo	0,533	0,098	0,254	0,784	202

Note: em todos os painéis, a amostra consiste em 202 municípios considerando as fronteiras de 1920. Os dados geográficos são originalmente do Ipeadata (distância para a capital, latitude, longitude e altitude) e da Embrapa Solos (tipos de solo). O indicador de presença de ferrovia para cada município foi construído com base no conteúdo do site www.estacoesferroviarias.com.br. As demais variáveis correspondem a características socioeconômicas de cada municipalidade disponíveis nos Censos de 1872, 1920 e 2000.

Tabela 2.2: Paineis - Número de imigrantes por país de origem, por ano e por município de destino - Dados do Museu da Imigração (1882-1920)

VAR. DEP: Núm. Imigrantes	(1)	(2)	(3)
EF Linha 1 (Central)	7,518 (4,328)*	7,525 (4,326)*	7,992 (4,368)*
EF Linha 2 (SP Railway)	24,850 (9,730)**	24,856 (9,728)**	25,455 (9,685)***
EF Linha 3 (Paulista)	29,698 (4,323)***	29,706 (4,325)***	30,003 (4,460)***
EF Linha 4 (Sorocabana)	15,965 (3,871)***	15,907 (3,877)***	16,183 (3,964)***
EF Linha 5 (Mogiana)	19,936 (3,810)***	19,948 (3,812)***	20,287 (3,873)***
EF Linha 6 (Noroeste)	5,945 (2,257)***	5,948 (2,257)***	5,986 (2,543)**
EF Linha 7 (SP-Minas)	3,989 (1,661)**	3,991 (1,661)**	4,001 (2,102)*
EF Linha 8 (Araraquara)	26,047 (9,132)***	26,050 (9,133)***	26,144 (9,264)***
Dummy Prinetti (ITA)		-27,096 (5,110)***	-33,078 (5,412)***
Dummy Seca (ESP)		-9,519 (2,527)***	-9,652 (1,711)***
Dummy Revolução (POR)		35,688 (20,618)*	-9,286 (4,139)**
Dummy Legislação (JAP)		10,825 (2,488)***	25,054 (3,667)**
Dummy Conflito (ALE)		10,833 (1,983)***	22,925 (4,374)***
Ferr x Prinetti (ITA)			11,293 (4,617)**
Ferr x Seca (ESP)			0,192 (3,878)
Ferr x Revolução (POR)			67,073 (35,225)*
Ferr x Legislação (JAP)			-19,499 (3,134)***
Ferr x Conflito (ALE)			-31,944 (6,153)***
Observações	202	202	202
R-quadrado	0,175	0,178	0,180
Teste F-Parcial	n.a	10,12	7,28
EF Ano	SIM	SIM	SIM
EF Município	SIM	SIM	SIM
EF País de Origem	SIM	SIM	SIM
EF Linha Férrea	SIM	SIM	SIM

Nota: Erros padrão robustos em parênteses, clusterizados nas fronteiras municipais de 1920. Todas as colunas reportam os resultados das regressões em MQO, nas quais a variável dependente é o número de imigrantes do país de origem 'n', a chegar no município 'm', no ano 't'. As variáveis "Efeito Fixo de Linha Férrea" assumem valores de 1 a 8 a depender de qual linha férrea cruza o município. A coluna (1) não inclui tratamentos e traz apenas os coeficientes dos efeitos fixos de ano, município e país de origem do imigrante e serve apenas para referência. As colunas (2) e (3) incluem os tratamentos. Apenas a coluna (3) as interações dos tratamentos com a presença de ferrovias. Níveis de significância: *** p<0.01, ** p<0.05, * p<0.1.

Tabela 2.3: Primeiro Estágio

Var. Dep.: Índ. Diversidade 1920	(1)	(2)	(3)	(4)	(5)	(6)
Diversidade 1920 Preditada	1,701 (0,126)***	1,568 (0,147)***	1,632 (0,157)***	1,066 (0,165)***	1,197 (0,208)***	1,196 (0,183)***
Prop. Estrang. 1920					-0,363 (0,292)	
Prop. Estrang. 1872						-0,004 (0,002)***
Observações	202	202	202	202	202	202
R-quadrado	0,389	0,438	0,451	0,695	0,698	0,707
Estat-F 1o est.	181,72	113,81	107,64	41,96	32,99	42,47
Controles Geográficos	NÃO	SIM	SIM	SIM	SIM	SIM
Controles Censo 1872	NÃO	NÃO	SIM	SIM	SIM	SIM
Efeito Fixo Ferrovias	NÃO	NÃO	NÃO	SIM	SIM	SIM
Prop. Estrang. 1920	NÃO	NÃO	NÃO	NÃO	SIM	NÃO
Prop. Estrang. 1872	NÃO	NÃO	NÃO	NÃO	NÃO	SIM

Nota: Erros padrão robustos em parênteses, clusterizados nas fronteiras municipais de 1920. A variável Diversidade assume valores entre zero e um, onde zero denota diversidade zero, pois a totalidade da população municipal é nativa, e um para o caso de diversidade máxima, no caso limite em que cada habitante do município tem uma origem distinta. O modelo (1) não inclui controles e serve apenas para referência. Os modelos de (2) a (6) incluem controles geográficos incluem a distância para a capital, latitude, longitude, elevação e indicadores para os diferentes tipos de solo (latossolo, argissolo, cambissolo e espodossolo). Os modelos de (3) a (6) incluem características socioeconômicas de 1872 (proporção de escravos, proporção de alfabetizados, proporção de trabalhadores ocupados na agricultura, na indústria e nos serviços e varejo). Os modelos de (4) a (6) incluem uma dummy para a presença de ferrovias no território do município em 1920. O modelo (5) controla pela proporção de estrangeiros no município de acordo com os dados do Censo de 1920. O modelo (6) controla pela proporção de estrangeiros em 1872 construída a partir dos dados do Museu da Imigração. Todas as regressões são estimadas para os 202 municípios com base nas fronteiras municipais de 1920. Níveis de significância: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Tabela 2.4: Mínimos Quadrados Ordinários (MQO)

Var. Dep.: ln renda p.c. 2000	(1)	(2)	(3)	(4)	(5)	(6)
Índ. Diversidade 1920	0,485 (0,062)***	0,361 (0,047)***	0,342 (0,048)***	0,312 (0,080)***	0,285 (0,080)***	0,307 (0,082)***
Prop. Estrang. 1920					0,530 (0,317)*	
Prop. Estrang. 1872						0,002 (0,004)
Observações	202	202	202	202	202	202
R-quadrado	0,258	0,532	0,566	0,584	0,592	0,585
Controles Geográficos	NÃO	SIM	SIM	SIM	SIM	SIM
Controles Censo 1872	NÃO	NÃO	SIM	SIM	SIM	SIM
Efeito Fixo Ferrovias	NÃO	NÃO	NÃO	SIM	SIM	SIM
Prop. Estrang. 1920	NÃO	NÃO	NÃO	NÃO	SIM	NÃO
Prop. Estrang. 1872	NÃO	NÃO	NÃO	NÃO	NÃO	SIM

Nota: Erros padrão robustos em parênteses, clusterizados nas fronteiras municipais de 1920. Todas as colunas reportam os resultados das regressões em MQO nas quais a variável dependente é o logaritmo natural da renda per capita no nível do município em 2000. A variável "Diversidade" assume valores entre zero e um, onde zero denota diversidade zero, pois a totalidade da população municipal é nativa, e um para o caso de diversidade máxima, no caso limite em que cada habitante do município tem uma origem distinta. O modelo (1) não inclui controles e serve apenas para referência. Os modelos de (2) a (6) incluem controles geográficos incluem a distância para a capital, latitude, longitude, elevação e indicadores para os diferentes tipos de solo (latossolo, argissolo, cambissolo e espodossolo). Os modelos de (3) a (6) incluem características socioeconômicas de 1872 (proporção de escravos, proporção de alfabetizados, proporção de trabalhadores ocupados na agricultura, na indústria e nos serviços e varejo). Os modelos de (4) a (6) incluem uma dummy para a presença de ferrovias no território do município em 1920. O modelo (5) controla pela proporção de estrangeiros no município de acordo com os dados do Censo de 1920. O modelo (6) controla pela proporção de estrangeiros em 1872 construída a partir dos dados do Museu da Imigração. Todas as regressões são estimadas para os 202 municípios com base nas fronteiras municipais de 1920. Níveis de significância: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Tabela 2.5: Resultado da Estimação em dois estágios - 2SLS

VARIÁVEIS	(1)	(2)	(3)	(4)	(5)	(6)
<i>Painel A: Segundo estágio - Var. Dep.: Log da renda p. c. em 2000</i>						
Índ. Diversidade 1920	0,723 (0,095)***	0,644 (0,099)***	0,573 (0,090)***	0,824 (0,171)***	0,755 (0,160)***	0,863 (0,163)***
Prop. Estrang. 1920					0,203 (0,307)	
Prop. Estrang. 1872						-0,001 (0,003)
<i>Painel B: Forma Reduzida - Var. Dep.: Log da renda p. c. em 2000</i>						
Diversidade 1920 Predita	1,229 (0,160)***	1,010 (0,147)***	0,935 (0,155)***	0,878 (0,174)***	0,904 (0,182)***	1,032 (0,191)***
Prop. Estrang. 1920					-0,071 (0,342)	
Prop. Estrang. 1872						-0,005 (0,003)**
<i>Painel C: MQO - Var. Dep.: Log da renda p. c. em 2000</i>						
Índ. Diversidade 1920	0,485 (0,062)***	0,361 (0,047)***	0,342 (0,048)***	0,312 (0,080)***	0,285 (0,080)***	0,307 (0,082)***
Prop. Estrang. 1920					0,530 (0,317)*	
Prop. Estrang. 1872						0,002 (0,004)
Observações	202	202	202	202	202	202
R-quadrado Form. Red.	0,223	0,534	0,552	0,601	0,601	0,606
R-quadrado MQO	0,258	0,532	0,566	0,584	0,592	0,585
Wald Qui-quad 2o est.	181,72	113,81	107,64	41,96	32,99	42,47
Controles Geográficos	NÃO	SIM	SIM	SIM	SIM	SIM
Controles Censo 1872	NÃO	NÃO	SIM	SIM	SIM	SIM
Efeito Fixo Ferrovias	NÃO	NÃO	NÃO	SIM	SIM	SIM
Prop. Estrang. 1920	NÃO	NÃO	NÃO	NÃO	SIM	NÃO
Prop. Estrang. 1872	NÃO	NÃO	NÃO	NÃO	NÃO	SIM

Nota: Erros padrão robustos em parênteses, clusterizados nas fronteiras municipais de 1920. Os controles geográficos incluem: a distância para a capital, latitude, longitude, elevação, e indicadores para os diferentes tipos de solo (latossolo, argissolo, cambissolo e espodossolo). As características socioeconômicas de 1872 incluem: proporção de estrangeiros, proporção de escravos, proporção de alfabetizados, proporção de crianças frequentando escola, proporção de trabalhadores ocupados na agricultura, indústria e serviços e varejo. Todas as regressões são estimadas para os 202 municípios com base nas fronteiras municipais de 1920. Níveis de significância: *** p<0.01, ** p<0.05, * p<0.1.

Tabela 2.6: Resultado da Estimação em dois estágios - 2SLS - Instrumento: Diversidade Predita Apenas pelo Tratamento

VARIÁVEIS	(1)	(2)	(3)	(4)	(5)	(6)
<i>Painel A: Primeiro estágio - Var. Dep.: Diversidade em 1920</i>						
Div 1920 VI Trat	2,112 (0,528)***	1,845 (0,481)***	1,852 (0,507)***	0,751 (0,300)**	0,688 (0,332)**	0,741 (0,309)***
Prop. Estrang. 1920					0,605 (0,265)**	
Prop. Estrang. 1872						0,005 (0,003)*
<i>Painel B: Segundo estágio - Var. Dep.: Log da renda p. c. em 2000</i>						
Índ. Diversidade 1920	0,712 (0,127)***	0,505 (0,116)***	0,425 (0,105)***	0,487 (0,280)***	0,427 (0,290)	0,485 (0,283)*
Prop. Estrang. 1920					0,431 (0,353)	
Prop. Estrang. 1872						0,001 (0,004)
<i>Painel C: Forma Reduzida - Var. Dep.: Log da renda p. c. em 2000</i>						
Div 1920 VI Trat	1,510 (0,351)***	0,932 (0,211)***	0,787 (0,192)***	0,366 (0,182)**	0,294 (0,202)	0,360 (0,187)*
Prop. Estrang. 1920					0,690 (0,330)**	
Prop. Estrang. 1872						0,003 (0,005)
Observações	202	202	202	202	202	202
Estat-F 1o est.	15,58	14,69	13,36	6,27	4,31	5,76
Wald Qui-quad 2o est.	31,39	203,26	235,41	878,50	940,05	882,20
Controles Geográficos	NÃO	SIM	SIM	SIM	SIM	SIM
Controles Censo 1872	NÃO	NÃO	SIM	SIM	SIM	SIM
Efeito Fixo Ferrovias	NÃO	NÃO	NÃO	SIM	SIM	SIM
Prop. Estrang. 1920	NÃO	NÃO	NÃO	NÃO	SIM	NÃO
Prop. Estrang. 1872	NÃO	NÃO	NÃO	NÃO	NÃO	SIM

Nota: Erros padrão robustos em parênteses, clusterizados nas fronteiras municipais de 1920. Os controles geográficos incluem: a distância para a capital, latitude, longitude, elevação, e indicadores para os diferentes tipos de solo (latossolo, argissolo, cambissolo e espodossolo). As características socioeconômicas de 1872 incluem: proporção de estrangeiros, proporção de escravos, proporção de alfabetizados, proporção de crianças frequentando escola, proporção de trabalhadores ocupados na agricultura, indústria e serviços e varejo. Todas as regressões são estimadas para os 202 municípios com base nas fronteiras municipais de 1920. Níveis de significância: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Tabela 2.7: Resultado da Estimação em dois estágios - 2SLS - Erros Padrão dos coeficientes calculados por bootstrap

VARIÁVEIS	(1)	(2)	(3)	(4)	(5)	(6)
<i>Segundo estágio - Var. Dep.: Log da renda p. c. em 2000</i>						
Índ. Diversidade 1920	0,723 (0,093)***	0,644 (0,102)***	0,573 (0,095)***	0,824 (0,192)***	0,755 (0,183)***	0,863 (0,201)***
Prop. Estrang. 1920					0,203 (0,361)	
Prop. Estrang. 1872						-0,001 (0,005)
Observações	202	202	202	202	202	202
Wald Qui-quad 2o est.	61,37	188,39	198,27	637,79	607,15	543,50
Controles Geográficos	NÃO	SIM	SIM	SIM	SIM	SIM
Controles Censo 1872	NÃO	NÃO	SIM	SIM	SIM	SIM
Efeito Fixo Ferrovias	NÃO	NÃO	NÃO	SIM	SIM	SIM
Prop. Estrang. 1920	NÃO	NÃO	NÃO	NÃO	SIM	NÃO
Prop. Estrang. 1872	NÃO	NÃO	NÃO	NÃO	NÃO	SIM

Nota: Erros-padrão dos coeficientes calculados por bootstrap em parênteses, clusterizados nas fronteiras municipais de 1920. Os controles geográficos incluem: a distância para a capital, latitude, longitude, elevação, e indicadores para os diferentes tipos de solo (latossolo, argissolo, cambissolo e espodossolo). As características socioeconômicas de 1872 incluem: proporção de estrangeiros, proporção de escravos, proporção de alfabetizados, proporção de crianças frequentando escola, proporção de trabalhadores ocupados na agricultura, indústria e serviços e varejo. Todas as regressões são estimadas para os 202 municípios com base nas fronteiras municipais de 1920. Níveis de significância: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Tabela 2.8: Efeito da Diversidade na Renda per capita, condicional nas identidades nacionais e religiões

VARIÁVEL DEP: Renda per capita 2000	(1)	(2)
Diversidade 1920 MQO	0,278 (0,056)***	
Diversidade 1920 VI		0,826 (0,186)***
Proporção Alemães	-4,834 (8,812)	-1,966 (8,183)
Proporção Espanhóis	0,267 (0,588)	-0,066 (0,604)
Proporção Italianos	0,642 (0,416)	0,206 (0,449)
Proporção Portugueses	1,489 (1,485)	1,219 (1,442)
Proporção Japoneses	0,733 (0,935)	0,766 (0,891)
Proporção Católicos	-0,054 (0,082)	-0,054 (0,092)
Proporção Protestantes	1,665 (1,339)	2,047 (1,304)
Observações	202	202
R-quadrado	0,583	0,566
Controles Geográficos	SIM	SIM
Controles Censo 1872	SIM	SIM

Nota: Erros padrão robustos em parênteses, clusterizados nas fronteiras municipais de 1920. Os controles geográficos incluem: a distância para a capital, latitude, longitude, elevação, e indicadores para os diferentes tipos de solo (latossolo, argissolo, cambissolo e espodossolo). As características socioeconômicas de 1872 incluem: proporção de estrangeiros, proporção de escravos, proporção de alfabetizados, proporção de crianças frequentando escola, proporção de trabalhadores ocupados na agricultura, indústria e serviços e varejo. Todas as regressões são estimadas para os 202 municípios com base nas fronteiras municipais de 1920. Níveis de significância: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Tabela 2.9: Os Efeitos de Longo Prazo da Diversidade de País de Origem no Emprego e na Transformação Estrutural

	(1)	(2)	(3)	(4)	(5)	(6)
Painel A: Ocup. Agric.	1872	1920	2000	1872	1920	2000
Diversidade 1920 MQO	-0,039 (0,247)	-0,158 (0,025)***	-0,166 (0,024)***			
Diversidade 1920 VI				-0,117 (0,696)*	-0,338 (0,084)***	-0,352 (0,081)***
Painel B: Ocup. Ind.	1872	1920	2000	1872	1920	2000
Diversidade 1920 MQO	0,022 (0,104)**	0,087 (0,014)***	0,036 (0,019)*			
Diversidade 1920 VI				0,039 (0,027)	0,171 (0,044)***	0,031 (0,058)
Painel C: Ocup. Serv.	1872	1920	2000	1872	1920	2000
Diversidade 1920 MQO	0,017 (0,023)	0,072 (0,014)***	0,125 (0,018)***			
Diversidade 1920 VI				0,168 (0,062)	0,316 (0,044)***	0,076 (0,064)***
Observações	202	202	202	202	202	202
R-quadrado (Painel A)	0,013	0,528	0,486	0,015	0,526	0,480
R-quadrado (Painel B)	0,021	0,455	0,403	0,009	0,380	0,370
R-quadrado (Painel C)	0,003	0,572	0,549	0,007	0,425	0,397
Controles Geográficos	SIM	SIM	SIM	SIM	SIM	SIM
Controles Censo 1872	SIM	SIM	SIM	SIM	SIM	SIM

Nota: Erros padrão robustos em parênteses, clusterizados nas fronteiras municipais de 1920. Todas as colunas reportam os resultados das regressões em MQO. No Painel A, a variável dependente é a proporção dos ocupados na agricultura em 1920 e em 2000. No Painel B, a variável dependente é a proporção dos ocupados na indústria e serviços em 1920 e em 2000. Para todos os painéis, as variáveis explicativas são a (i) diversidade de país de nascimento observada (MQO) apenas entre os estrangeiros para o ano de 1920 ou (ii) a diversidade de país de nascimento predita (VI) calculada apenas entre os estrangeiros para o ano de 1920. Elas assumem valores entre zero e um, onde zero denota diversidade zero, e um para o caso de diversidade máxima, caso limite em que cada estrangeiro do município tem uma origem distinta. Para todas as colunas há controles para fatores geográficos (distância para a capital, latitude, longitude, elevação, e indicadores para os diferentes tipos de solo) e para as características socioeconômicas de 1872 (proporção de escravos, proporção de alfabetizados, proporção de trabalhadores ocupados na agricultura, indústria e serviços/varejo). Níveis de significância: *** p<0.01, ** p<0.05, * p<0.1.

Tabela 2.10: Os Efeitos de Longo Prazo da Diversidade de País de Origem na Diversidade Ocupacional Industrial

Painel A: Div. Ocupacional Industrial 1920	(1)	(2)
Diversidade 1920 MQO	0,061 (0,026)**	
Diversidade 1920 VI		0,183 (0,057)***
Painel B: Div. Ocupacional Industrial 2000	(1)	(2)
Diversidade 1920 MQO	0,075 (0,022)***	
Diversidade 1920 VI		0,145 (0,052)***
Painel C: Div. Ocupacional Total 2000	(1)	(2)
Diversidade 1920 MQO	0,037 (0,009)***	
Diversidade 1920 VI		0,075 (0,023)***
Observações	202	202
R-quadrado (Painel A)	0,178	0,179
R-quadrado (Painel B)	0,503	0,485
R-quadrado (Painel C)	0,371	0,337
Controles Geográficos	SIM	SIM
Controles Censo 1872	SIM	SIM

Nota: Erros padrão robustos em parênteses, clusterizados nas fronteiras municipais de 1920. Todas as colunas reportam os resultados das regressões em MQO. No Painel A, a variável dependente é a diversidade ocupacional industrial em 1920. A medida assume valores entre zero e um, onde zero denota diversidade ocupacional zero, na qual a totalidade da população ocupada do município trabalha num só setor da indústria, e um para o caso de diversidade máxima, no caso limite em que cada ocupado do município atue em um setor distinto. No Painel B, a variável dependente é diversidade ocupacional industrial em 2000. Os dados de emprego setorial industrial disponíveis para o cálculo do índice em 1920 são: extrativista, edificações, vestuário, madeira e móveis, metalúrgica, química e alimentos. Para o índice de 2000, os dados setoriais disponíveis são os mesmos de 1920 além de: máquinas, automotiva e plásticos. No Painel C, os dados setoriais disponíveis são: judiciário, magistério, medicina, defesa, administração pública, atividades religiosas, profissionais liberais, agricultura, indústria, serviços e varejo (não há dados disponíveis para estes setores em 1920). Para os três painéis, as variáveis explicativas são a (i) diversidade de país de nascimento calculada (MQO) apenas entre os estrangeiros para o ano de 1920 ou (ii) a diversidade de país de nascimento predita (VI) calculada apenas entre os estrangeiros para o ano de 1920. Elas assumem valores entre zero e um, onde zero denota diversidade zero, e um para o caso de diversidade máxima, no caso limite em que cada estrangeiro do município tem uma origem distinta. Para todas as colunas há controles para fatores geográficos (distância para a capital, latitude, longitude, elevação, e indicadores para os diferentes tipos de solo) e para as características socioeconômicas de 1872 (proporção de escravos, proporção de alfabetizados, proporção de trabalhadores ocupados na agricultura, indústria e serviços/varejo). Níveis de significância: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Tabela 2.11: Os Efeitos de Longo Prazo da Diversidade de País de Origem nos Gastos Municipais com Educação

Painel A: Gastos Municipais com Educação	(1)	(2)
Índ. Diversidade 1920 MQO	0,015 (0,006)**	
Índ. Diversidade 1920 VI		0,049 (0,021)**
Observações	186	186
R-quadrado	0,164	0,172
Controles Geográficos	SIM	SIM
Controles Censo 1872	SIM	SIM

Nota: Erros padrão robustos em parênteses, clusterizados nas fronteiras municipais de 1920. Todas as colunas reportam os resultados das regressões em MQO. As variáveis dependentes são os gastos dos governos municipais com educação em 1926. As variáveis explicativas são a (i) diversidade de país de nascimento calculada apenas entre os estrangeiros para o ano de 1920 ou (ii) a diversidade de país de nascimento predita calculada apenas entre os estrangeiros para o ano de 1920. Elas assumem valores entre zero e um, onde zero denota diversidade zero, pois a totalidade da população municipal é nativa, e um para o caso de diversidade máxima, no caso limite em que cada estrangeiro do município tem uma origem distinta. Para todas as colunas há controles para fatores geográficos (distância para a capital, latitude, longitude, elevação, e indicadores para os diferentes tipos de solo) e para as características socioeconômicas de 1872 (proporção de escravos, proporção de alfabetizados, proporção de trabalhadores ocupados na agricultura, indústria e serviços/varejo). Todas as regressões são estimadas para os 186 municípios com base nas fronteiras municipais de 1920 para os quais havia dados disponíveis. Níveis de significância: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Tabela 2.12: Os Efeitos de Longo Prazo da Diversidade de País de Origem em Variáveis Educacionais

	(1)	(2)	(3)	(4)
Painel A: Escolas/criança (x1000)	1920	2000	1920	2000
Diversidade 1920 MQO	0,321 (0,112)***	-3,370 (0,771)***		
Diversidade 1920 VI			0,882 (0,284)***	-6,895 (1,719)***
Painel B: Professores/criança (x1000)	(1) 1920	(2) 2000	(3) 1920	(4) 2000
Diversidade 1920 MQO	6,149 (1,501)***	16,349 (4,770)***		
Diversidade 1920 VI			19,927 (4,918)***	60,537 (16,707)***
Painel C: Proporção crianças matriculadas	(1) 1940	(2) 2000	(3) 1940	(4) 2000
Diversidade 1920 MQO	0,156 (0,023)***	0,009 (0,004)**		
Diversidade 1920 VI			0,316 (0,105)***	0,027 (0,010)***
Painel D: Anos de Escolaridade		(1) 2000		(2) 2000
Diversidade 1920 MQO		0,932 (0,130)***		
Diversidade 1920 VI				2,354 (0,396)***
Observações	202	202	202	202
R-quadrado (Painel A)	0,183	0,614	0,177	0,593
R-quadrado (Painel B)	0,433	0,138	0,444	0,161
R-quadrado (Painel C)	0,455	0,411	0,399	0,412
R-quadrado (Painel D)	n.a.	0,497	n.a.	0,464
Controles Geográficos	SIM	SIM	SIM	SIM
Controles Censo 1872	SIM	SIM	SIM	SIM

Nota: Erros padrão robustos em parênteses, clusterizados nas fronteiras municipais de 1920. Resultados das regressões em MQO nas quais as variáveis dependentes são, respectivamente, nos Painéis A, B e C: o número de escolas por criança em idade escolar (entre 7 e 14 anos), o número de escolas por criança em idade escolar e a proporção de crianças matriculadas. Os dados foram retirados dos Censo populacionais de 1920 e de 2000. A variável "Diversidade de Origem" assume valores entre zero e um, onde zero denota diversidade zero, pois a totalidade da população municipal é nativa, e valor unitário para o caso de diversidade máxima, no caso limite em que cada estrangeiro do município tem uma origem distinta. Os controles geográficos incluem: a distância para a capital, latitude, longitude, elevação, e indicadores para os diferentes tipos de solo (latossolo, argissolo, cambissolo e espodossolo). As características socioeconômicas de 1872 incluem: presença de ferrovias, proporção de estrangeiros, proporção de escravos, proporção de alfabetizados, densidade populacional, proporção de trabalhadores ocupados na agricultura, indústria e serviços e varejo. Todas as regressões são estimadas para os 202 municípios com base nas fronteiras municipais de 1920.

3 IMIGRAÇÃO E PRODUTIVIDADE: CEM ANOS E UM OCEANO DE DISTÂNCIA

3.1 Introdução

O presente artigo tem como objetivo documentar a persistência de produtividade passado ao menos um século da chegada de imigrantes a um novo continente. Mais especificamente, o artigo investiga se as diferenças regionais do país de origem dos imigrantes influenciam os resultados no mercado de trabalho dos descendentes no país de destino. Associar registros de indivíduos de gerações distantes é empiricamente desafiador. Explora-se, aqui, um conjunto de dados dos imigrantes italianos com destino ao Brasil durante o período de imigração em massa (1880-1920). Por meio de um processo de *fuzzy matching*, os sobrenomes dos descendentes foram identificados nos microdados de mercado de trabalho no Brasil em 2016, portanto, ao menos três gerações de distância.

A literatura sobre o papel das transferências inter-geracionais na determinação dos salários é extensa (e.g., Chiswick, 1988; Borjas, 1992, 1995b; Angrist e Krueger, 1991). Para Borjas (1992, 1995b), há externalidades (ou *spillovers*) étnicas no processo. Variáveis étnicas agregadas ajudam a explicar resultados individuais em educação e nos rendimentos, por exemplo. Em particular, os ganhos dos filhos de imigrantes são afetados não apenas pelas habilidades médias dos pais, como nos modelos usuais de mobilidade de renda intergeracional, mas também pelas habilidades médias da etnia na geração dos pais. Estes resultados mostram a importância do ambiente em que os indivíduos estão inseridos. Fernandez e Fogli (2009) investigam, adicionalmente, o papel da composição do bairro de residência na transmissão

cultural e mostram que quanto maior a densidade de um grupo étnico, maior o impacto nas novas gerações. Giuliano e Nunn (2017) corroboram os resultados dos autores anteriores e concluem que a variabilidade do ambiente de uma geração para a seguinte importa. Ambientes com mais estabilidade de uma geração para outra dão maior importância à ancestralidade. Borjas (1993) revela uma relação entre os ganhos dos imigrantes e os ganhos de seus filhos nascidos nos EUA. Os ganhos de americanos de segunda geração são fortemente afetados por variáveis que descrevem as condições econômicas nos países de origem de seus pais.

O artigo se relaciona também com a linhas de pesquisa empírica mais ampla, sobre assimilação dos imigrantes, que estuda a adaptação dos recém-chegados ao destino. Apesar de o enfoque ser de curto prazo, esta linha de pesquisa tem encontrado evidências ambíguas, que corroboram tanto o argumento da adaptação quanto o da persistência, a depender do contexto estudado. Antecol (2000), por um lado, documenta correlação positiva entre a participação feminina no mercado de trabalho do país de origem e das imigrantes nos Estados Unidos. Com relação às gerações seguintes, a correlação segue positiva, mas mais fraca. Por outro lado, Borjas (2007) e Blau e Kahn (2007) analisam a experiência da assimilação dos imigrantes mexicanos nos Estados Unidos e concluem que, mesmo na segunda geração, as diferenças entre nativos e descendentes se mantêm para uma série de variáveis, como por exemplo, fecundidade e oferta de trabalho feminino. No contexto europeu, as evidências são de maior assimilação aos padrões dos nativos: Van Ours e Vennman (2003) mostram que, apesar de a primeira geração de imigrantes ter escolaridade e salários mais baixos que a mesma coorte de nativos, quando se analisa a segunda geração, a diferença em frequência escolar vai a zero; resultados similares foram encontrados por Bauer e Riphan (2007) para a Suíça; Gang e Zimmermann (2000) comparam a escolaridade dos imigrantes na segunda geração com os dos nativos da mesma faixa etária utilizando dados da Alemanha. Concluem que há padrão claro entre países de origem e escolaridade, mesmo na segunda geração. O tamanho da rede étnica

também tem efeito positivo nos anos de escolaridade. Para os filhos de estrangeiros, a escolaridade dos pais não desempenha nenhum papel nas escolhas educacionais.

Os modelos tradicionais de mobilidade intergeracional presumem que as correlações através das gerações declinam geometricamente. Em outras palavras, a correlação entre avós e filho é o quadrado da correlação pai-filho, e assim por diante. Alguns artigos recentes questionaram a suposição de que o processo de transmissão intergeracional do capital humano tem uma memória de apenas um período. O principal elemento novo no presente trabalho é o fato de se vincular pessoas através de mais de duas gerações. Estudos recentes têm empregado métodos empíricos inovadores para relacionar mais de uma geração. Chan e Boliver (2013) mostraram associação entre as posições de classe dos avós e dos netos, mesmo após a posição de classe dos pais ser levada em conta. Lindahl et al. (2015) vinculam ganhos individuais (e educação) por três gerações na Suécia e encontram persistência mais forte ao longo de três gerações do que o previsto modelos simples para duas gerações. Em exercícios que se assemelham mais ao presente trabalho, Collado et al. (2013) e Clark e Cummins (2014) exploram a distribuição de sobrenomes para estimar a mobilidade social de longo prazo. As variáveis de interesse, no entanto, são status socioeconômico e riqueza familiar, respectivamente.

O artigo explora o episódio de grandes proporções que foi a emigração italiana em massa para as Américas, mais especificamente para a região de São Paulo, no Brasil. No período 1850-1914, a Itália foi a maior exportadora de mão de obra do mundo. Rosoli (1992) estima que 18 milhões tenham deixado a península entre 1870 e 1930. Sori (1979) e Benedini e Arquilla (2007) trabalham com estimativas de 20 milhões de emigrantes para períodos semelhantes. Trata-se do maior êxodo populacional da história moderna, equivalente a aproximadamente um terço da população atual do país. As dificuldades na economia rural ampliaram a duração do expatrio para os países vizinhos de sazonal para plurianual. A falta de trabalho no

campo, a fome e a ocorrência de algumas epidemias, como o cólera, a varíola e a pelagra, impulsionaram a emigração para América.

O caso italiano é emblemático, pois tem traços culturais distintos em suas regiões, notadamente, conforme documentado pela literatura das ciências sociais, entre o norte e o sul. Banfield (1958) e Putnam (1993) argumentam que as diferenças no comportamento cívico, social e econômico entre as regiões norte e sul tem origem na histórias distante e nas tradições medievais, que remontam a ocupação dos mercenários normandos e que essas diferentes dotações de “capital social” contribuem para explicar o atraso econômico do sul da Itália. Nesta mesma linha, Guiso, Sapienza e Zingales (2016) mostram que as cidades italianas que alcançaram o autogoverno na Idade Média têm hoje um nível mais alto de capital cívico do que cidades semelhantes na mesma área que não o fizeram. A experiência de autogoverno da Idade Média promoveu crenças na própria capacidade de completar tarefas e atingir metas - e essa atitude positiva, transmitida através das gerações, aumenta o capital cívico hoje.

Diante deste contexto, o artigo documenta a persistência de produtividade passadas ao menos três gerações desde a chegada dos primeiros emigrantes italianos ao Brasil. Em específico, o artigo investiga se as diferenças regionais do país de origem dos imigrantes influenciam os salários no mercado de trabalho dos descendentes no país de destino. As elasticidades encontradas variam de 0,023 a 0,05, a depender dos controles incluídos em cada especificação, sempre estatisticamente significantes a 1%. Este resultado confirma a persistência de longo prazo entre região de origem dos descententes e seus rendimentos do trabalho contemporâneos e são consistentes com os de outros autores que examinaram a mobilidade socioeconômica de longo prazo (e.g., Collado et al., 2013; Clark e Cummins, 2014). O principal elemento novo no presente trabalho é o fato de se vincular pessoas através de mais de duas gerações, o que é empiricamente desafiador e só é possível com o cruzamento dos

sobrenomes, nos moldes do realizado por Barone e Mocetti (2016).

O restante do artigo se desenvolve em mais quatro seções. A Seção 3.2 resume o contexto histórico da emigração e a dispersão dos italianos em São Paulo. A Seção 3.3 descreve o conjunto de dados utilizados e estratégia empírica adotada. A Seção 3.4, por sua vez, traz os resultados das estimações. Finalmente, o artigo traz as considerações finais apontando próximos passos para a agenda de pesquisa.

3.2 Contexto Histórico da Emigração na Itália e Dispersão Territorial dos Italianos em São Paulo

A Europa do século XVIII e XIX apresentou crescimento demográfico jamais visto. Os ganhos de produtividade resultantes da revolução industrial quebraram o equilíbrio malthusiano e a população cresceu de 144 para 486 milhões entre 1800 e 1920 (Arruda, 2007). A terra deixou de ser fonte de subsistência para servir à produção em larga escala, com especialização e concorrência entre países. Para as regiões essencialmente agrárias e sem vantagens comparativas, a consequência foi falta de trabalho no campo e a fome. Houve também a ocorrência de algumas epidemias, como o cólera, a varíola e a pelagra, que, em conjunto com os outros fatores de insatisfação, se tornaram combustível para revoltas e levantes que colocavam a ordem social em risco. Deste modo, os governos viam a emigração como solução parcial redutora de pressão interna. O êxodo foi, em razão disso, incentivado pelas nações emissoras (Gonçalves, 2008).

Este era o contexto geral dos principais países de onde partiram os imigrantes. Na Itália recém-unificada, em específico, a migração sazonal de trabalhadores para os países vizinhos do norte, notadamente França, Alemanha e Império Austro-Húngaro, passou de sazonal para plurianual. O campo sofria com o avanço do mundo industrial

e com a concorrência dos grãos norte-americanos e argentinos. Diante da queda na demanda por sua mão-de-obra, os excedentes populacionais rurais optam pela emigração para a América. Contribuíram, ainda, para o êxodo: (i) redução do custo dos transportes (e do tempo de viagem, de três meses para duas semanas), e; (ii) estabelecimento de redes, viabilizadas pela popularização dos serviços postais e do telégrafo. À medida que as notícias relativas ao sucesso dos primeiros paisanos na América iam chegando, mais e mais emigrantes se encorajavam.

No período 1850-1914, a Itália foi o país que mais exportou mão-de-obra. Não só para o Brasil. Rosoli (1992) estima que 18 milhões tenham deixado a península entre 1870 e 1930. Sori (1979) trabalha com estimativas de 20 milhões de emigrantes para o período entre 1861 e 1940. Segundo Benedini e Arquilla (2007), nos cem anos contados a partir de 1870, registrou-se a partir da península itálica o maior êxodo populacional da história moderna: cerca de 20 milhões de pessoas, aproximadamente um terço da população atual do país. O Brasil foi o terceiro a receber mais italianos, atrás de Estados Unidos e Argentina.

Ao mesmo tempo, o Brasil vivia o oposto. Na segunda metade do século XIX, o Brasil e, em maior grau, São Paulo, enfrentavam falta de mão-de-obra necessária para o atendimento da forte demanda por exportações de café. Em 1850, o tráfico de escravos fora banido e, em 1871 e 1888, as Leis do Ventre Livre e Áurea, respectivamente, restringiram ainda mais a oferta de trabalho dos africanos. Neste contexto, houve grande esforço no governo imperial para o recrutamento de imigrantes europeus, neste primeiro momento, inclusive com a intenção nem sempre velada de se branquear a população brasileira (Schwarcz, 1993; Skidmore, 1976; Seyferth, 1985). A imigração estrangeira para o São Paulo deu-se, então, com forte intervenção estatal, em resposta às necessidades dos produtores do café, produto com 70% de participação na pauta de exportações do país em 1920 (Abreu e Bevilaqua, 1996). Estimativas do IBGE apontam para a entrada no Brasil de mais de 4,1 milhões de

imigrantes entre 1872 e 1929. Para efeito de comparação, o Censo de 1872 registra população residente total de 9,9 milhões no país, dos quais 383 mil (3,8%) eram estrangeiros ou naturalizados. Já em 1900, a população residente total atingira 17,4 milhões, porém os estrangeiros e naturalizados já somavam 1,279 milhão, ou 7,3% (IBGE, 2007).

Do total de imigrantes que vieram ao Brasil, aproximadamente 58% foram acolhidos por São Paulo (Almeida, 2014). A preferência por este local pode estar vinculada às facilidades concedidas pelo governo local, tais como passagem e alojamento, somadas às oportunidades de trabalho de uma economia em expansão (Fausto, 2009). Apenas entre 1880 e 1900, a província de São Paulo recebeu aproximadamente 940 mil imigrantes (Vasconcelos, 1994, p. 227). Para se ter uma ideia da dimensão deste fluxo, o Censo de 1872 aponta população total da província em cerca de 837 mil habitantes. Segundo os dados do Museu da Imigração, 60% dos imigrantes matriculados na Hospedaria entre 1889 e 1915 tiveram sua viagem subsidiada pelo governo brasileiro. A parcela restante é classificada como espontânea ou autofinanciada. Os imigrantes com viagem subsidiada já tinham, na maioria dos casos, destino final no interior. O transporte dos colonos para as fazendas foi realizada, em grande parte, pela rede ferroviária, através de oito linhas que foram construídas de leste a oeste. A construção das ferrovias iniciou-se em 1867, ligando o porto de Santos, principal responsável pelo escoamento das exportações de café do país, à cidade de Jundiaí, no coração da área produtora dos grãos à época. A expansão da rede de ferrovias muitas vezes era financiada com excedente de capital do café, tanto que, entre os acionistas das companhias ferroviárias, era frequente figurarem as famílias dos cafeicultores (Gonçalves, 2008). Em 1920, as linhas já cortavam 74% dos 202 municípios. Desta forma, como pode-se ver na Figura 3.2, os imigrantes já povoavam a quase totalidade do território paulista em 1920. Fica evidente também que a dispersão dos italianos (mapa em azul), a nacionalidade de origem modal, se espelha na dispersão da totalidade dos imigrantes (mapa em vermelho).

3.3 Dados e Estratégia Empírica

3.3.1 Dados

O artigo utiliza dados de seis fontes de dados distintas. Em primeiro lugar, utilizam-se os registros do Museu da Imigração de São Paulo. Deles foram extraídos os sobrenomes dos imigrantes italianos chegados em São Paulo entre 1882 e 1920. Em segundo lugar, para atribuir a cada sobrenome uma região de origem na Itália, utilizou-se o site *cognomix.it*, cuja base de dados são as páginas brancas italianas (listas telefônicas residenciais). A investigação das condições dos descendentes italianos no mercado de trabalho atual, em São Paulo, fez uso dos dados da versão identificada da RAIS (Relatório Anual de Informações Sociais), do Ministério do Trabalho, de 2016. Ele contém 19,8 milhões de observações sobre todos os trabalhadores no mercado de trabalho formal desta unidade da federação. O artigo também faz uso dos dados regionais do ISTAT (Instituto Nacional de Estatística italiano) referentes ao ano de 2015: produto per capita, escolaridade e proporção de ocupados na agricultura, indústria e serviços, e no seus respectivos subsetores. Por último, para as variáveis de controle, o artigo utiliza dados extraídos dos Censos decenais de 1872 e 1920, além de informações sobre a infraestrutura ferroviária e de características geográficas em nível municipal, cujas fontes são o Ipeadata e a Embrapa Solos. A quinta fonte de dados é o Censo de 1872, registro prévio ao grande fluxo imigratório e de onde foram retiradas características socioeconômicas em nível municipal. Por último, utilizou-se do cronograma de inaugurações das estações ferroviárias que cortavam o território paulista em 1920 retiradas do site *estacoesferroviaria.com.br*.

A montagem do banco de dados final utilizado nas estimações começa a partir da extração dos sobrenomes italianos do livro de matrículas do Museu da Imigração de São Paulo. Ele registra, entre 1882 e 1920, 92.077 sobrenomes distintos

cuja nacionalidade declarada pelo imigrante era italiana. Em um segundo passo, identificou-se a região de origem na Itália para 20.910 (22,7%) deles, por meio do *webscrap* do *cognomix.it*. Na maioria dos casos, os sobrenomes não foram encontrados nas listas telefônicas italianas por erros de grafia, tanto no momento do registro, na chegada do imigrante, quanto ao longo do tempo, nos registros cartoriais através das gerações.¹ Desta forma, ao realizar o *fuzzy matching* destes sobrenomes e suas respectivas regiões de origem italianas com os sobrenomes constantes na versão identificada da Relação Anual de Informações Sociais (RAIS-Migra), do Ministério de Trabalho, de 2016, construiu-se a base de dados com aproximadamente 1,16 milhão de registros de trabalhadores formais para os quais se sabe a região de origem na Itália.

A Tabela 3.1 traz um resumo das estatísticas descritivas das variáveis utilizadas nas estimações. No Painel A, dados sobre os descendentes de italianos no mercado de trabalho formal de São Paulo, em 2016. O salário médio dos descendentes de imigrantes italianos identificados pelo sobrenome empregados formalmente em São Paulo é de R\$3,5 mil, com desvio-padrão de 1,4 vezes a média. Já a sua escolaridade média é de 7,1 anos, com desvio-padrão de 1,7 anos. A idade média era de 38,5 em 2016 e a proporção dos que se declaram não-brancos era de 33,6%. A proporção de mulheres era 41,8%. O Painel B traz os dados referentes às regiões italianas. A renda per capita média regional era de 28,5 mil euros (euros correntes de 2015), mas com desvio-padrão de 7,2 mil. Para se ter uma ideia da desigualdade regional de renda na Itália, a região mais rica tem PIB per capita 132% mais alta que a mais pobre. As estatísticas de proporção de ocupados na agricultura e de diplomados nas universidades corroboram esta desigualdade de renda. O Painel C resume dados socioeconômicos das localidades de São Paulo em 1872, ou seja, período anterior aos grandes fluxos imigratórios. Por fim, no Painel D tem-se as

¹É frequente a ocorrência de sobrenomes com grafias quase idênticas. Por exemplo, os sobrenomes "Abondante" e "Abbondante" estão nos registros da Hospedaria. Apenas um deles (Abbondante) foi encontrado no banco de dados do *cognomix.it*.

proporções de municípios cujos limites eram cortados por ao menos uma ferrovia em 1872 e 1920.

3.3.2 Estratégia Empírica

Com o objetivo de se estimar a importância do desenvolvimento das regiões italianas de origem nos rendimentos do trabalho dos descendentes de italianos no Brasil atual, utiliza-se uma estimação por mínimos quadrados ordinários (MQO) que pode ser representada pela equação que segue:

$$y_{i,2016} = \alpha_m + \rho_s + \beta y_{r,2015} + X_i' \gamma + e_{i,m,r} \quad (3.1)$$

Onde $y_{i,2016}$ é o log do salário médio do indivíduo i , participante do mercado de trabalho formal de São Paulo em 2016. Os termos α_m e ρ_s são os efeitos fixos de município e de setor CNAE da empresa empregadora, respectivamente. O termo $y_{r,2015}$ é a renda per capita da Região r italiana em 2015 e, por fim, o termo X_i é um vetor de variáveis de controle mincerianas tais como, idade, idade ao quadrado, indicador de raça não-branca e indicador de gênero.

Em um segundo momento, o artigo investiga se as características socioeconômicas iniciais das localidades de destino dos imigrantes atenuam ou fortalecem o efeito estimado inicialmente pela equação 3.1. Para isso, foram incluídos termos de interação da renda regional italiana com variáveis do Censo de 1872 em nível municipal:

$$y_{i,2016} = \alpha_m + \rho_s + \beta_1 y_{r,2015} + \beta_2 y_{r,2015} * S_{m,1872} + X_i' \gamma + e_{i,m,r} \quad (3.2)$$

Onde $S_{m,1872}$ são características socioeconômicas dos municípios em 1872: frequência escolar, proporção de escravos, proporção de alfabetizados (>6 anos), proporção dos ocupados na agricultura, indústria e serviços, produção de café em hectares e a dummy de presença de ferrovias em 1920.

3.4 Resultados

A Tabela 3.3 traz os resultados das estimações nas quais se confirma a persistência da relação entre região de origem dos descendentes e seus rendimentos do trabalho contemporâneos. Todas as colunas reportam os resultados das regressões em MQO nas quais a variável dependente é o logaritmo natural do salário médio do trabalhador formal no Estado de São Paulo reportado pela RAIS de 2016. A coluna (1) não inclui explicativas além do logaritmo natural da renda per capita das regiões da Itália, medidas em euros correntes de 2015. Os modelos de (2) a (5) incluem as explicativas mincerianas tradicionais (idade, idade ao quadrado, dummy de não-brancos e dummy indicadora do gênero feminino). O modelo (3) inclui efeitos fixos de município baseados na divisão administrativa de 1920 (202 municípios), enquanto as especificações (4) e (5) utilizam a divisão administrativa atual (de 645 municípios). Apenas o modelo (6) inclui os efeitos fixos de setor, de acordo com a classificação CNAE reportada na RAIS. Todas as regressões são estimadas a partir dos dados de 1,155 milhão de descendentes italianos encontrados na RAIS por meio de processo de *fuzzy matching* dos sobrenomes. As elasticidades encontradas variam de 0,023 a 0,05, a depender dos controles incluídos em cada especificação, sempre estatisticamente significantes a 1%.

As estimativas do artigo podem ser viesadas em razão de erros de medida. Ainda que a literatura específica para a Itália afirme que há enorme heterogeneidade e “localismo” de sobrenomes italianos, o que reforça ainda mais a qualidade das associação dos sobrenomes a sua região (Barone e Mocetti, 2016), a ocorrência de migração sul-norte na Itália, principalmente nas décadas de 1950-1970, é uma ameaça à estratégia de identificação. A região italiana de prevalência do sobrenome pode ter mudado desde 1880. Para se ter uma ideia do impacto da migração sul-norte ocorrida no intervalo entre as coletas dos dados aqui analisado 1920 x 2015, de acordo com IBGE (2007), as oito regiões pertencentes genericamente ao norte

italiano² são responsáveis por aproximadamente 47% do fluxo imigratório para o Brasil entre 1876 e 1920. Ao mesmo tempo, o cruzamento dos sobrenomes italianos registrados pela Hospedaria entre 1882-1920, praticamente a mesma janela temporal do IBGE, com as listas telefônicas de 2010 aponta para participação das oito regiões do norte de aproximadamente 59%. Ao mesmo tempo, o fato de as linhas telefônicas serem residenciais pode introduzir viés de seleção, pois os mais jovens, em geral, não utilizam mais telefone fixo, por exemplo. Além disso, é importante ressaltar que, aqui, analisa-se apenas a linha paterna, pois os sobrenomes maternos, em geral, se perdem na sucessão geracional.

A Tabela 3.4 traz os resultados da Tabela 3.3 acrescidas de termos que integram a renda per capita regional italiana com características socioeconômicas dos municípios de destino extraídas do Censo de 1872, ou seja, que retratam condições destas localidades prévias ao grande fluxo de imigrantes. A coluna (1) apenas repete os resultados reportados na coluna (6) da Tabela 3.3, ou seja, inclui as explicativas mincerianas tradicionais (idade, idade ao quadrado, dummy de não-brancos e dummy indicadora do gênero feminino). As especificações a partir da coluna (2) incluem, a cada coluna, um novo termo de interação da renda per capita regional com uma característica dos municípios em 1872. Apenas a inclusão da interação da renda regional italiana com a proporção municipal de escravos em 1872 e com a dummy de presença de ferrovias em 1920 resultou em coeficientes com significância estatística. Em ambos os casos, este termo de interação atenua o efeito da região de origem nos rendimentos do descendente.

²Lombardia, Vêneto, Piemonte, Emilia-Romagna, Ligúria, Friuli-Venezia Giulia, Trentino Alto-Ágide e Valle d'Aosta.

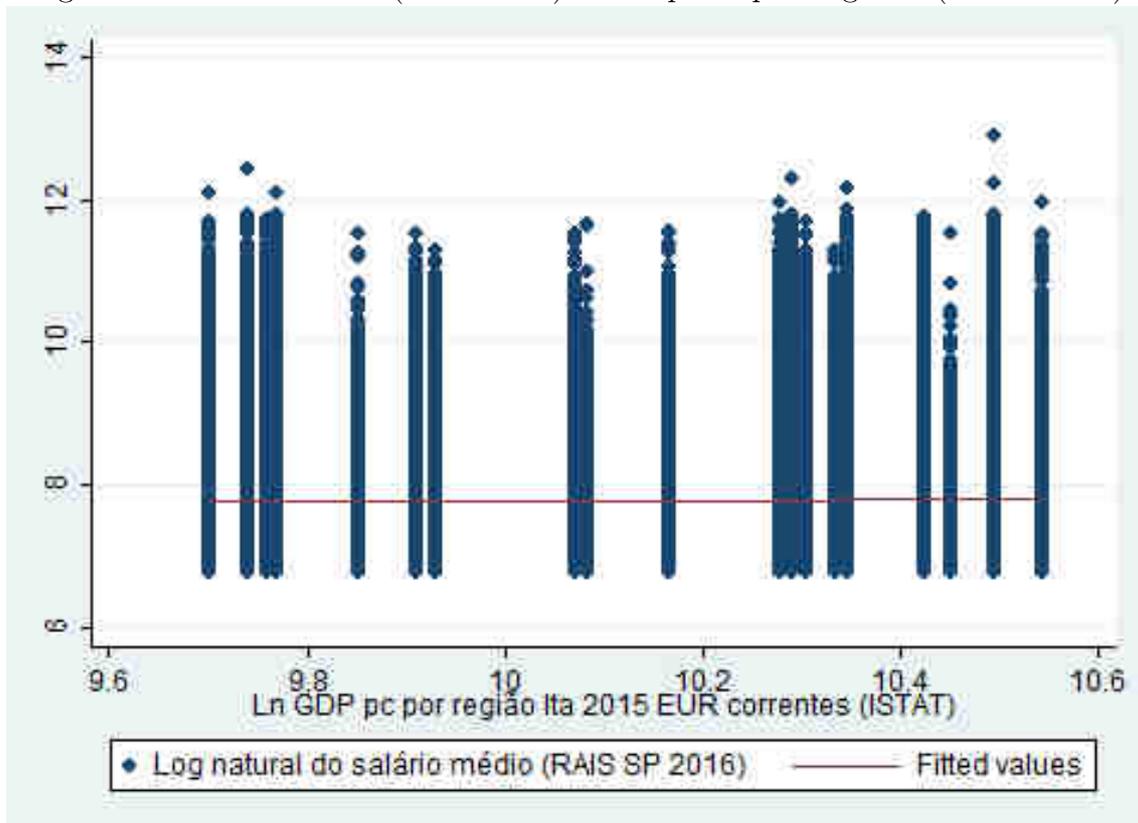
3.5 Considerações Finais e Próximos Passos

O artigo investiga se as diferenças regionais do país de origem dos imigrantes influenciam os resultados no mercado de trabalho dos descendentes no país de destino. Para isso, utiliza dados dos emigrantes italianos com destino ao Brasil durante o período de imigração em massa (1880-1920). Por meio de um processo de *fuzzy matching*, os sobrenomes dos emigrantes foram identificados nos microdados de mercado de trabalho no Brasil. Estimou-se, aqui, a importância do desenvolvimento das regiões italianas de origem nos rendimentos do trabalho dos descendentes de italianos no Brasil atual.

O artigo encontra elasticidade-ponto dos ganhos dos descendentes em relação às regiões de origem dos ancestrais de 0,023, estatisticamente significativa a 1%. Os resultados encontrados são consistentes com os de outros autores que examinaram a mobilidade socioeconômica de longo prazo (Barone e Mocetti, 2016).

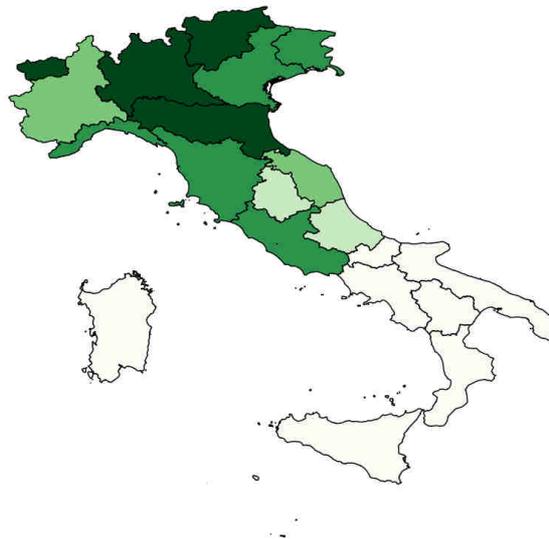
Se os resultados socioeconômicos da maioria dos indivíduos são fortemente relacionados aos de seus ancestrais, as crianças de origem pobre provavelmente serão relativamente mais pobres como adultos com a desigualdade irá se perpetuar. Procurar as mesmas evidências em diferentes regiões ou nações de origem e investigar melhor os mecanismos (e.g., capital humano, acesso a certas profissões) por trás da persistência socioeconômica de longo prazo são agendas promissoras para futuras pesquisas.

Figura 3.1: Salário Médio (RAIS 2016) x PIB per capita regional (ISTAT 2015)



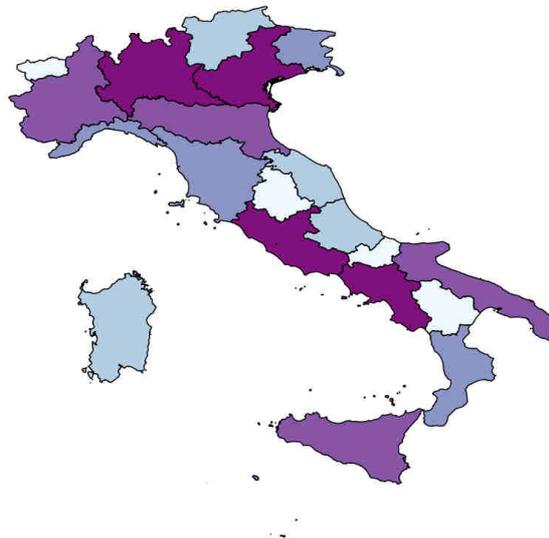
Fonte: RAIS/MTE, ISTAT/ITALIA e Museu da Imigração de São Paulo. Nota: Elaboração do autor. Variáveis em logarítimo natural.

Figura 3.2: PIB per capita por região na Itália (Euros correntes 2015)



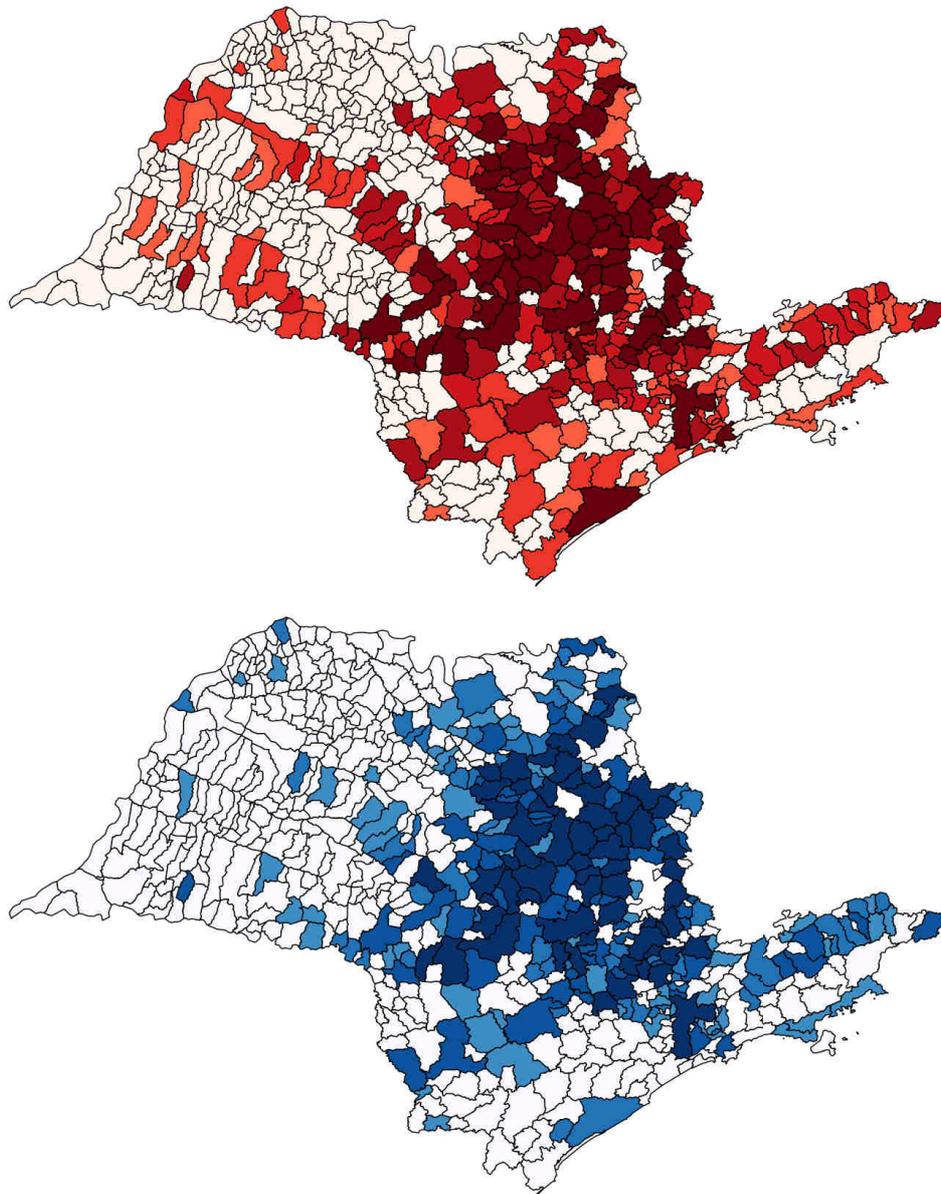
Fonte: ISTAT (*Istituto Nazionale di Statistica*). Elaboração do autor.

Figura 3.3: Distribuição dos Sobrenomes por região de origem na Itália



Fonte: Museu da Imigração de São Paulo e site <http://www.cognomix.it>.
Elaboração do autor.

Figura 3.4: Distribuição dos Imigrantes no Estado de São Paulo (1882-1920): Imigrantes totais em vermelho; Imigrantes italianos em azul



Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

Tabela 3.1: Estatísticas Descritivas

VARIÁVEIS	(1) média	(2) desv-pad	(3) min	(4) max	(5) N
A. Variáveis do Mercado de Trabalho - RAIS 2016 SP					
Salário médio mensal (R\$)	3,481	4,916	880	393.739	1.155.139
Salário médio mensal (log)	7,785	0,742	6,780	12,883	1.155.139
Anos de Escolaridade	7,079	1,665	1	11	1.155.139
Idade	38,842	12,117	14	99	1.155.139
Dummy não-branco	0,336	0,472	0	1	1.155.139
Dummy mulher	0,417	0,493	0,000	1,000	1.155.139
B. Variáveis das Regiões na Itália - ISTAT 2015					
Renda per capita (EUR correntes)	28,450	7,208	16,326	37,892	20
Renda per capita (log)	10,140	0,279	9,701	10,542	20
Proporção Secundária	0,358	0,034	0,292	0,422	20
Proporção Universitária	0,173	0,026	0,132	0,233	20
Proporção Agricultura	0,030	0,016	0,011	0,057	20
Proporção Indústria	0,228	0,064	0,125	0,342	20
Proporção Serviços	0,742	0,062	0,605	0,851	20
C. Variáveis do Censo de 1872					
Frequência escolar	0,145	0,101	0,027	0,764	202
Proporção de escravos	0,154	0,086	0,039	0,531	202
Proporção de alfabetizados (>6 anos)	0,202	0,110	0,048	0,452	202
População (em milhares)	11,147	7,198	1,566	41,751	202
Proporção dos ocupados na agricultura	0,594	0,099	0,351	0,908	202
Proporção dos ocupados na indústria	0,109	0,044	0,020	0,244	202
Proporção dos ocupados em serviços/varejo	0,296	0,091	0,058	0,569	202
Produção de café (ton)	0,032	0,041	0,000	0,240	202
D. Variáveis de Infraestrutura Ferroviária					
Dummy de presença de ferrovia 1872	0,024	0,156	0,000	1,000	202
Dummy de presença de ferrovia 1920	0,738	0,441	0,000	1,000	202

Nota: O indicador de presença de ferrovia para cada município foi construído com base no conteúdo do site www.estacoesferroviarias.com.br. As demais variáveis correspondem a características socioeconômicas de cada municipalidade disponíveis no Censo de 1872.

Tabela 3.2: Frequência das Regiões de Prevalência dos Sobrenomes dos Imigrantes Italianos (ordem decrescente de número de descendentes)

Região	Núm. descendentes	Proporção	Proporção acum.
Lombardia	265.440	23,0%	23,0%
Vêneto	220.229	19,1%	42,0%
Campania	104.678	9,1%	51,1%
Lázio	89.454	7,7%	58,9%
Sicília	85.142	7,4%	66,2%
Emília-Romagna	74.983	6,5%	72,7%
Piemonte	67.625	5,9%	78,6%
Púglia	63.719	5,5%	84,1%
Toscana	50.585	4,4%	88,5%
Calábria	38.342	3,3%	91,8%
Friuli-Venezia Giulia	29.291	2,5%	94,3%
Ligúria	13.744	1,2%	95,5%
Abruzzo	10.286	0,9%	96,4%
Outras	41.621	3,6%	100,0%
Total	1.155.139	100,0%	100,0%

Note: Associou-se a região na qual se encontrou o maior número de ocorrências de cada sobrenome italiano constante nas matrículas da Hospedaria de São Paulo (1882-1920).

Tabela 3.3: Salários em São Paulo x Renda per capita regiões italianas

	(1)	(2)	(3)	(4)	(5)
Var. Dep.: Log Salário Médio (RAIS 2016 SP)					
Log Renda per capita por Região (ISTAT 2015)	0,038 (0,007)***	0,025 (0,007)***	0,050 (0,003)***	0,051 (0,003)***	0,023 (0,004)***
Idade (RAIS 2016 SP)		0,074 (0,004)***	0,072 (0,004)***	0,071 (0,004)***	0,054 (0,002)***
Idade ao Quadrado (RAIS 2016 SP)		-0,001 (0,000)***	-0,001 (0,000)***	-0,001 (0,000)***	-0,001 (0,000)***
Dummy Não-Brancos (RAIS 2016 SP)		-0,038 (0,026)***	-0,094 (0,031)***	-0,094 (0,031)***	-0,134 (0,024)***
Dummy Mulher (RAIS 2016 SP)		-0,160 (0,026)***	-0,177 (0,022)***	-0,174 (0,021)***	-0,163 (0,004)***
Observações	1.155.139	1.155.139	1.155.139	1.155.139	1.155.139
R-quadrado	0,000	0,098	0,163	0,170	0,406
Efeito Fixo Município	NÃO	NÃO	Mun 1920	Mun 2000	Mun 2000
Efeito Fixo Setor CNAE	NÃO	NÃO	NÃO	NÃO	SIM

Nota: Erros padrão robustos em parênteses, clusterizados nas fronteiras municipais de 1920. Todas as colunas reportam os resultados das regressões em MQO nas quais a variável dependente é o logaritmo natural do salário médio do trabalhador formal no Estado de São Paulo reportado pela RAIS de 2016. O modelo (1) não inclui explicativas além do logaritmo natural da renda per capita das regiões da Itália, medidas em euros correntes de 2015. Os modelos de (2) a (5) incluem as explicativas mincerianas tradicionais (idade, idade ao quadrado, dummy de não-brancos e dummy indicadora do gênero feminino). O modelo (3) inclui efeitos fixos de município baseados na divisão administrativa de 1920 (202 municípios), enquanto as especificações (4) e (5) utilizam a divisão administrativa atual (de 645 municípios). Apenas o modelo (6) inclui os efeitos fixos de setor, de acordo com a classificação CNAE reportada na RAIS. Todas as regressões são estimadas a partir dos dados de 1,155 milhão de descendentes italianos encontrados na RAIS por meio de processo de *fuzzy matching* dos sobrenomes. Níveis de significância: *** p<0.01, ** p<0.05, * p<0.1.

Tabela 3.4: Salários em São Paulo x Renda per capita regiões italianas e Interações

Var. Dep.: Log Salário Médio	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log Renda por Região	0,023 (0,004)***	0,019 (0,006)***	0,017 (0,004)***	0,021 (0,008)***	0,038 (0,016)***	0,023 (0,007)***	0,004 (0,010)***	0,023 (0,004)***	0,006 (0,007)	1,799 (1,157)
Idade (RAIS)	0,054 (0,002)***									
Idade ao Quadrado (RAIS)	-0,001 (0,000)***									
Dummy Não-Brancos (RAIS)	-0,134 (0,024)***									
Dummy Mulher (RAIS)	-0,163 (0,004)***									
Renda Região X Freq. Escolar		0,022 (0,026)								-0,006 (0,034)
Renda Região X Prop. Escravos			0,039 (0,012)***							0,073 (0,020)***
Renda Região X Taxa Alfabet.				0,008 (0,029)						0,022 (0,027)
Renda Região X Prop. Agrícola					-0,029 (0,024)					-1,832 (1,164)
Renda Região X Prop. Indústria						-0,002 (0,049)				-1,879 (1,163)
Renda Região X Prop. Serviços							0,054 (0,033)			-1,731 (1,158)
Renda Região X Prod. Café (ha)								-0,004 (0,068)		-0,074 (0,059)
Renda Região X Dummy Ferrovia									0,017 (0,007)**	0,013 (0,007)*
Observações	1,15mi									
R-quadrado	0,406	0,406	0,406	0,406	0,406	0,406	0,406	0,406	0,406	0,406
Efeito Fixo Município 2000	SIM									
Efeito Fixo Setor CNAE	SIM									

Nota: Erros padrão robustos em parênteses, clusterizados nas fronteiras municipais de 1920. Todas as colunas reportam os resultados das regressões em MQO nas quais a variável dependente é o logaritmo natural do salário médio do trabalhador formal no Estado de São Paulo, reportado pela RAIS de 2016, e a variável explicativa é o logaritmo natural da renda per capita das regiões da Itália, medidas em euros correntes de 2015. O modelo (1) equivale ao (6) da Tabela 3.3, ou seja, inclui as explicativas mincerianas tradicionais (idade, idade ao quadrado, dummy de não-brancos e dummy indicadora do gênero feminino). As especificações a partir da coluna (2) incluem a cada coluna um novo termo de interação da renda per capita regional e uma característica dos municípios em 1872 (baseadas na divisão administrativa de 1920). Todas as regressões são estimadas a partir dos dados de 1,155 milhão de descendentes italianos encontrados na RAIS por meio de processo de *fuzzy matching* dos sobrenomes. Níveis de significância: *** p<0.01, ** p<0.05, * p<0.1.

4 CONCLUSÃO

A presente tese traz três ensaios complementares sobre diversidade, imigração e desenvolvimento de longo prazo. Os resultados têm implicações diretas para as políticas públicas nacionais e leis de imigração. Se maior diversidade de origens gera externalidades positivas no longo prazo e as características destas origens persistem no tempo, mesmo passadas três ou quatro gerações, é racional que os países incorram em certos custos de curto prazo para a atração e assimilação dos estrangeiros no presente.

A experiência dos diversos países receptores de imigrantes no final do século XIX e início do XX deve jogar luz no debate acerca do necessário redesenho de políticas públicas referentes à imigração nos dias de hoje. Os custos no curto prazo devem ser relativizados pelos benefícios de longo prazo, que começam a ser finalmente quantificados graças à disponibilidade dos microdados e do avanço do ferramental de análise na ciência econômica.

REFERÊNCIAS

- ABRAMITZKY, R.; BOUSTAN, L.; ERIKSSON, K. Europe's tired, poor, huddled masses: self-selection and economic outcomes in the age of mass migration. **American Economic Review**, Pittsburgh, v.102(5), p.1832—1856, 2012. page.5757
- ABRAMITZKY, R.; BOUSTAN, L.; ERIKSSON, K. Have the Poor always been less likely to Migrate? Evidence from inheritance practices during the Age of Mass Migration. **Journal of Development Economics**, Amsterdam, v.102, p.2–14, 2013. page.5757
- ABRAMITZKY, R.; BOUSTAN, L.; ERIKSSON, K. A Nation of Immigrants: assimilation and economic outcomes in the age of mass migration. **Journal of Political Economy**, Chicago, v.122(3), p.467–506, 2014. page.5757
- ABRAMOVITZ, M.; DAVID, P. A. American Macroeconomic Growth in the Era of Knowledge-Based Progress: the long run perspective. In: ENGERMAN, S. E.; GALLMAN, R. E. (Ed.). **Cambridge Economic History of the United States**. New York, NY: Cambridge University Press, 2000. v.3, p.1–92. page.5757
- ABREU, M. P.; BEVILAQUA, A. S. Brazil as an Export Economy: 1880-1930. **Texto para Discussão n. 363, PUC-Rio**, Rio de Janeiro, 1996. page.5757
- ACEMOGLU, D.; GALLEGO, F.; ROBINSON, J. A. Institutions, Human Capital, and Development. **Annual Review of Economics**, Palo Alto, v.6(1), p.875–912, 2014. page.5757
- ACEMOGLU, D.; JOHNSON, S.; ROBINSON, J. A. Colonial Origins of Comparative Development: an empirical investigation. **American Economic Review**, Pittsburgh, v.91(5), p.1369–1401, 2001. page.5757
- AGER, P.; BRUCKNER, M. Cultural Diversity and Economic Growth: evidence from the united states during the age of mass migration. **European Economic Review**, Amsterdam, v.64, p.76–97, 2013. page.5757
- ALESINA, A.; BAQIR, R.; W., E. Public Goods and Ethnic Divisions. **Quarterly Journal of Economics**, Oxford, v.114(4), p.1243–1284, 1999. page.5757

- ALESINA, A.; HARNOSS, J.; RAPOPORT, H. Birthplace Diversity and Economic Prosperity. **Journal of Economic Growth**, Berlin, v.21, n.2, p.101–138, 2016. page.5757
- ALESINA, A.; LA FERRARA, E. Ethnic Diversity and Economic Performance. **Journal of Economic Literature**, Pittsburgh, v.43, n.3, p.762–800, 2005. page.5757
- ALESINA, A.; SPOLAORE, E.; WACZIARG, R. Economic Integration and Political Disintegration. **American Economic Review**, Pittsburgh, v.90(5), p.1276–1296, 2000. page.5757
- ALEXANDER, J. G. **Daily Life in Immigrant America, 1870–1920**. Chicago: Reed Elsevier Inc., 2007. 352p. page.5757
- ALMEIDA, R. G. C. Trajetória de um patricio: a imigração sirio-libanesa em são paulo. **Revista Outras Fronteiras**, Cuiabá, v.1(2), p.302–323, 2014. page.5757
- ANGRIST, J. D.; PISCHKE, J. S. **Mostly harmless econometrics: an empiricist's companion**. Princeton: Princeton University Press, 2008. 373p. page.5757
- ANTECOL, H. An Examination of Cross-Country Differences in the Gender Gap in Labor Force Participation Rates. **Labour Economicss**, London, v.7(4), p.409–426, 2000. page.5656
- ARAUJO, M. L. V. **Os caminhos da riqueza dos paulistanos na primeira metade dos oitocentos**. São Paulo: Hucitec, 2006. 223p. page.5757
- ARRUDA, J. J. A. A Expansão Europeia Oitocentista: emigração e colonização. In: SOUZA, F.; MARTINS, I. (Ed.). **A Emigração Portuguesa para o Brasil**. Porto: Afrontamento, 2007. p.13–40. page.5757
- ASHRAF, Q.; GALOR, O. The Out of Africa Hypothesis, Human Genetic Diversity, and Comparative Economic Development. **American Economic Review**, Pittsburgh, v.103(1), p.1–46, 2013. page.5757
- BAGANHA, M. I. B. Uma Imagem Desfocada: a emigração portuguesa e as fontes sobre a emigração. **Análise Social**, Lisboa, v.XXVI (112-113), p.723–739, 1991. page.5757
- BANDIERA, O.; RASUL, I.; VIARENGO, M. The making of modern America: migratory flows in the age of mass migration. **Journal of Development Economics**, Amsterdam, v.102, p.23–47, 2013. page.5757

- BANFIELD, E. C. **The Moral Basis of a Backward Society**. New York: Free Press, 2005. 188p. page.5656
- BARONE, G.; MOCETTI, S. Intergenerational Mobility in the Very Long Run: florence 1427-2011. **Bank of Italy Working Paper No. 1060**, Roma, v.125, p.1–42, 2016. page.5656
- BASSANEZI, M. S. C. B. et al. **Atlas da Imigração Internacional em São Paulo, 1850–1950**. São Paulo: Editora UNESP, 2008. 138p. page.5757
- BAUER, P.; RIPHAHN, R. T. Heterogeneity in the intergenerational transmission of educational attainment: evidence from switzerland on natives and second-generation immigrants. **Journal of Population Economics**, Maastricht, v.20(1), p.121–148, 2007. page.5656
- BEACH, B.; JONES, D. B. Gridlock: ethnic diversity in government and public good provision. **American Economic Journal: Economic Policy**, Pittsburgh, v.9, n.1, p.112–136, 2017. page.5757
- BELLOTTO, M. A Imigração Espanhola no Brasil. Estado do fluxo migratório para o Estado de São Paulo (1931-1936). **Revista de Estudios Interdisciplinarios de América Latina y El Caribe**, Tel Aviv, v.3(2), 1992. page.5757
- BERGHahn, V. R. **Imperial Germany, 1871–1914: economy, society, culture, and politics**. 2.ed. New York: Berghahn Books, 2005. 400p. page.5757
- BLACKBOURN, D.; ELEY, G. **The Peculiarities of German History: bourgeois society and politics in nineteenth-century germany**. New York: Oxford University Press, 1984. 300p. page.5757
- BLAY, F.; KHAN, L. Gender and assimilation among Mexican Americans. In: BORJAS, G. (Ed.). **Mexican immigration to the United States**. Chicago: University of Chicago Press, 2007. p.57–106. page.5555, page.5757
- BODVARSSON, B.; BERG, H. Van der. **The Economics of Immigration – Theory and Policy**. Berlin: Springer, 2009. 492p. page.5757
- BORIS, F. **Fazer a América: a imigração em massa para a américa**. São Paulo: Editora da Universidade de São Paulo, 1999. 577p. page.5757
- BORIS, F. **História Concisa do Brasil**. São Paulo: Editora da Universidade de São Paulo, 2009. 156p. page.5757

- BORJAS, G. Ethnic capital and intergenerational mobility. **Quarterly Journal of Economics**, London, v.107(1), p.123–150, 1992. page.5555
- BORJAS, G. The intergenerational mobility of immigrants. **Journal of Labor Economics**, Chicago, v.11(1), p.113–135, 1993. page.5555
- BORJAS, G. Ethnicity, neighborhoods, and human capital externalities. **American Economic Review**, Pittsburgh, v.85(3), p.365–390, 1995b. page.5555
- BORJAS, G. Introduction. In: BORJAS, G. (Ed.). **Mexican immigration to the United States**. Chicago: University of Chicago Press, 2007. p.1–56. page.5555, page.5757
- BORJAS, G.; FREEMAN, R.; KATZ, L. How much do immigration and trade affect labor market outcomes? **Brookings Papers on Economic Activity**, Washington, v.1, p.1–90, 1997. page.5757
- BORJAS, G. J. The Economics of Immigration. **Journal of Economic Literature**, Pittsburgh, v.XXXII, n.1, p.1667–1717, 1994. page.5757
- BORJAS, G. J. The economic benefits of Immigration. **Journal of Economic Perspectives**, Pittsburgh, n.9, p.3–22, 1995. page.5757
- BORJAS, G. J. The labor demand curve is downward sloping: re-examining the impact of immigration on the labor market. **Quarterly Journal of Economics**, Oxford, n.118, p.1335–1374, 2003. page.5757
- BRASIL. **IBGE. Brasil 500 anos de povoamento**. Rio de Janeiro: Centro de Documentação e Disseminação de Informações, 2007. 232p. page.5757
- CARD, D. The impact of the Mariel Boatlift on the Miami labor market. **Industrial and Labor Relations Review**, Ithaca, v.43, p.245–257, 1990. page.5757
- CARD, D. Immigration Inows, Native Outows and the Local Labor Market Impacts of Higher Immigration. **Journal of Labor Economics**, Chicago, v.19, p.22–64, 2001. page.5757
- CARD, D. Immigration and Inequality. **American Economic Review**, Pittsburgh, v.99, p.1–21, 2009. page.5757
- CARD, D.; DI NARDO, J. Do immigrant inows lead to native outows? **American Economic Review**, Pittsburgh, v.90, p.360–367, 2000. page.5757

CARNEIRO, M. L. T.; TAKEUCHI, M. Y. **Imigrantes Japoneses no Brasil. Trajetória, Imaginário e Memória.** São Paulo: Editora da Universidade de São Paulo, 2010. 712p. page.5757

CARNEIRO, P.; LEE, S.; REIS, H. Please Call Me John: name choice and the assimilation of immigrants in the united states, 1900-1930. **IZA DP No. 9792**, Bonn, p.1-43, 2016. page.5656

CENNI, F. **Italianos no Brasil: andiamo in'merica.** São Paulo: Editora da Universidade de São Paulo, 2009. page.5757

CHAN, T. W.; BOLIVER, V. The grandparents effect in social mobility: evidence from british birth cohort studies. **American Sociological Review**, Chicago, v.78(4), p.662-678, 2013. page.5656

CLARK, G.; CUMMINS, N. Intergenerational wealth mobility in England, 1858-2012: surnames and social mobility. **Economic Journal**, Wisconsin, v.125, p.61-85, 2014. page.5656

COLISTETE, R. P.; LAMOUNIER, M. L. The End of Plantation? Coffee and Land Inequality in Early Twentieth Century. **Annals of Statistics**, São Paulo, v.7, p.1-30, 2011. page.5757

COLLADO, M.; ORTUÑO-ORTIN, I.; ROMEU, A. Long-run intergenerational social mobility and the distribution of surnames. **Universidad de Alicante**, Alicante, p.1-44, 2013. page.5656

DEAN, W. **The Industrialization of Sao Paulo, 1880-1945.** Austin: University of Texas Press, 1969. 263p. page.5757

DEAN, W. **Rio Claro. Um sistema brasileiro de grande lavoura, 1820-1920.** 1.ed. Rio de Janeiro: Paz e Terra, 1977. 206p. page.5757

DI FRANCESCO, N. **A hospedagem dos imigrantes na imperial cidade de São Paulo.** São Paulo: Editora Brasiliense, 1994. page.5757

DROLLER, F. Migration, Population Composition and Long-run Economic Development: evidence from settlements in the pampas. **The Economic Journal**, London, p.2321-2352, 2017. page.5757

DUSTMANN, C.; FABBRI, F.; PRESTON, I. The Impact of Immigration of British Labour Market. **The Economic Journal**, London, v.115, p.324-341, 2005. page.5757

EASTERLY, W.; LEVINE, R. The European Origins of Economic Development. **Journal of Economic Growth**, New York, v.21(3), p.225–257, 2016. page.5757

EFRON, B. Bootstrap methods: another look at the jackknife. **Annals of Statistics**, Bethesda, v.7, p.1–26, 1979. page.5757

ESPAÑA. **Catálogo y Publicación de Sequías Históricas**: fichas de eventos de sequía. Madrid: Ministério da Agricultura, 2013. 262p. page.5757

FENG, S.; KRUEGER, A. B.; OPPENHEIMER, M. Linkages among Climate Change, Crop Yields and Mexico-US Cross-Border Migration. **Proceedings of the National Academy of Sciences**, Washington, v.107(32), p.14257–14262, 2010. page.5757

FERNÁNDEZ, R.; FOGLI, A. Fertility: the role of culture and family experience. **Journal of the European Economic Association**, Viena, v.4(2-3), p.552–561, 2006. page.5555

FERNÁNDEZ, R.; FOGLI, A. Culture: an empirical investigation of beliefs, work, and fertility. **American Economic Journal: Macroeconomics**, Pittsburgh, v.1(1), p.146–177, 2009. page.5555

FISHER, D. H. **Albion's Seed**: a cultural history. New York: Oxford University Press, 1989. 946p. page.5757

FOGED, M.; PERI, G. Immigrants Effect on Native Workers: new analysis on longitudinal data. **IZA Discussion Paper n.8961**, Bonn, p.1–47, 2015. page.5757

GIULIANO, P.; NUNN, N. Understanding cultural persistence and change. **forthcoming**, Harvard University, 2017. page.5656

GLAESER, E. L. et al. Growth in cities. **The Journal of Political Economy**, Chicago, v.100(6), p.1126–1152, 1992. page.5757

GONÇALVES, P. C. **Mercadores de braços**: riqueza e acumulação na organização da emigração europeia para o novo mundo. São Paulo: Tese (Doutorado em História Econômica) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, 2008. 519p. page.5757

GONÇALVES, P. C. A cidade de São Paulo: um entreposto de braços para a lavoura cafeeira. **Revista Cordis de História Social da Cidade**, São Paulo, v.2, 2009. page.5757

GOUNARIS, B. The Macedonian Struggle (1903-1912): paving the way for liberation. In: MALCOLM, R. (Ed.). **A History of Macedonia. Books and Research Papers on the History of Macedonia and the Greek Heritage Collection**. Berkeley: University of California Press, 1990. v.II, p.508–527. page.5757

GRANDI, G. **Estado e capital ferroviário em São Paulo: a companhia paulista de estradas de ferro entre 1930 e 1961**. São Paulo: Faculdade de Filosofia, Letras e Ciências Humanas, 2013. 248p. page.5757

GUIISO, L.; SAPIENZA, P.; ZINGALES, L. Long term persistence. **Journal of the European Economic Association**, Viena, 2016. page.5656

HANIOGLU, M. S. **A Brief History of the Late Ottoman Empire**. Princeton: Princeton University Press, 2010. 134p. page.5757

HASHIMOTO, F. **Cem Anos de Imigração Japonesa**. São Paulo: Editora Unesp, 2008. 376p. page.5757

HATTON, T. J.; WILLIAMSON, J. G. **The Age of Mass Migration: causes and economic impact**. Oxford: Oxford University Press, 1998. 320p. page.5757

HIRSCHMAN, C.; MOGFORD, E. Immigration and the American Industrial Revolution from 1880 to 1920. **Social Science Research**, Amsterdam, v.38, p.897–920, 2009. page.5757

HOBSBAWN, E. **A Era do Capital, 1848-1875**. 5.ed. Rio de Janeiro: Paz e Terra, 1996. 518p. page.5757

HOLLOWAY, T. H. **Imigrantes para o café: café e sociedade em são paulo, 1886-1934**. Rio de Janeiro: Paz e Terra, 1984. 297p. page.5757

IMIGRAÇÃO. Museu da Imigração: banco de dados digitalizado. **Secretaria de Estado da Cultura**, São Paulo, 2016. page.5757

JACOBS, J. **The economy of cities**. 1.ed. New York: Random House, 1969. 288p. page.5757

KEMENY, T. Immigrant Diversity and Economic Development in Cities: a critical review. **SERC Discussion Paper n.149**, London, p.1–47, 2013. page.5757

KIM, S. Immigration, Industrial Revolution and Urban Growth in the United States, 1820-1920: factor endowments, technology and geography. **NBER Working Paper n.12900**, Cambridge, 2007. page.5757

KNOWLTON, C. S. **Sírios e libaneses: mobilidade social e espacial**. São Paulo: Ed. Anhembi, 1961. 61p. page.5757

LAMOUNIER, M. L. The “labour question” in nineteenth-century Brazil: railways, export agriculture, and labour scarcity. **Working papers in Economic History**, London, v.59, p.1–48, 2000. page.5757

LAZEAR, E. Diversity and Immigration. **NBER Working Paper n.6535**, Cambridge, 1998. page.5757

LAZEAR, E. Culture and language. **Journal of Political Economy**, Chicago, v.107(6), p.95–26, 1999. page.5757

LEITE, J. C. Emigração portuguesa: a lei e os números (1855-1914). **Análise Social**, Lisboa, v.XXIII (97), p.463–480, 1987. page.5757

LESSER, J. **Immigration, Ethnicity, and National Identity in Brazil, 1808 to the Present**. Cambridge: Cambridge University Press, 2013. 224p. page.5757

LINDAHL, M. et al. Long-term intergenerational persistence of human capital: an empirical analysis of four generations. **Journal of Human Resources**, Wisconsin, v.50(1), p.1–33, 2015. page.5656

LUTERBACHER, J. et al. European Seasonal and Annual Temperature Variability, Trends, and Extremes Since 1500. **Science**, Washington, DC, v.303(5663), p.1499–1503, 2004. page.5757

MARCÉN, M. C.; LIRIA, A. A. Períodos secos y sequías en la depresión central del Ebro. **Geographicalia**, Madrid, v.11-12, p.55–72, 1981. page.5757

MAUCH, C.; VASCONCELLOS, N. **Os Alemães no Sul do Brasil: cultura, etnicidade e história**. Canoas: Editora Ulbra, 1994. 224p. page.5757

MAURO, P. Corruption and growth. **Quarterly Journal of Economics**, London, v.110, p.681–712, 1995. page.5757

MELLO, J. M. Cardoso de. **O Capitalismo Tardio**. São Paulo: Editora Brasiliense, 1984. 152p. page.5757

MENYHERT, B. Economic Growth Spurred by Diversity: evidence from the austro-hungarian monarchy. page.5757
, Budapest, p.1–81, 2018.

- MEYER, J. Network approach versus brain drain: lessons from the diaspora. **International Migration Review**, New York, v.39, p.91–110, 2001. page.5757
- MIZRAHI, R. **Imigrantes no Brasil - Judeus do Descobrimento aos Dias Atuais**. São Paulo: Editora Companhia Nacional, 2005. 120p. page.5757
- MONASTERIO, L. Sobrenomes e Ancestralidade no Brasil. **Texto para Discussão IPEA n.2229**, Brasília, v.115, p.324–341, 2013. page.5757
- MONASTERIO, L. Surnames and ancestry in Brazil. **PLoS ONE**, San Francisco, v.12(5), p.1–15, 2017. page.5656, page.5757
- MONTALVO, J.; REYNAL-QUEROL, M. Ethnic diversity and economic development. **Journal of Development Economics**, Amsterdam, v.76(2), p.293–323, 2005. page.5757
- MOURA, S.; PAIVA, O. d. C. **Hospedaria de Imigrantes de São Paulo**. São Paulo: Editora Paz e Terra, 2008. page.5757
- MUNSHI, K. Networks in the modern economy: mexican migrants in the us labor market. **Quarterly Journal of Economics**, London, v.118, p.549–599, 2003. page.5757
- MURARD E.; SAKALLI, S. O. Mass Refugee Inflow and Long-Run Prosperity: lessons from the greek population resettlement. **IZA Discussion Paper n.11613**, Bonn, 2018. page.5757
- NETO, J. M. S. **A Chegada dos Letos (Prefácio de Nova Odessa: 100 anos da terra fértil, os frutos do progresso)**. Campinas: Editora Komedi, 2006. 111p. page.5757
- NETTO, A. D. **O problema do café no Brasil**. São Paulo: IPE/USP, 1981. 288p. page.5757
- NUNN, N.; QIAN, N.; SEQUEIRA, S. Migrants and the Making of America: the short and long run effects of immigration during the age of mass migration. **CEPR Discussion paper**, London, n.11899, 2017. page.5757
- OLIVEIRA, E. V. **Republicanos y Rebeldes más allá de las fronteras: brasil y la guerra civil española**. Valencia, 2015. 381p. page.5757
- O'REILLY E.; WILLIAMS, K. B. S. Culture and language. **Journal of Political Economy**, Chicago, v.107(6), p.S95–S126, 1997. page.5757

O'ROURKE, K. H.; WILLIAMSON, J. G. Around the European periphery 1870-1913: globalization, schooling and growth. **European Review of Economic History**, Cambridge, v.1, p.153–190, 1997. page.5757

OTTAVIANO, C.; PERI, G. Cities and Cultures. **Journal of Urban Economic**, Amsterdam, v.58, n.1, p.304–337, 2005. page.5757

OTTAVIANO, C.; PERI, G. The economic value of cultural diversity: evidence from united states cities. **Journal of Economic Geography**, Oxford, v.6, n.1, p.9–44, 2006. page.5757

PAULING, A. et al. Five Hundred Years of Gridded High-Resolution Precipitation Reconstructions over Europe and the Connection to Large-Scale Circulation. **Climate Dynamics**, Amsterdam, v.26, p.387–405, 2006. page.5757

PORTES, A. **The economic sociology of immigration**: essays on networks, ethnicity and entrepreneurship. New York: Russell Sage Foundation, 1995. 326p. page.5757

PUTNAM, R. D. **Making Democracy Work. Civic Traditions in Modern Italy**. Princeton: Princeton University Press, 1993. 247p. page.5656

PUTNAM, R. D. E Pluribus Unum: diversity and community in the twenty-first century. **Scandinavian Political Studies**, Oslo, n.30, p.137–174, 1997. page.5757

PUTTERMAN, L.; WEIL, D. Post-1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality. **Quarterly Journal of Economics**, Oxford, v.125(4), p.1627–1682, 2010. page.5757

REIS, J. A industrialização num país de desenvolvimento lento e tardio: portugal, 1870-1913. **Análise Social**, Lisboa, v.XXIII(96), p.207–227, 1987. page.5757

REIS-VIEIRA, J. E. **Imigração Italiana**: economia cafeeira e industrialização no brasil (1890-1930). Rio de Janeiro: Monografia, Universidade Cândido Mendes, 2012. page.5757

ROCHA, R.; FERRAZ, C.; SOARES, R. S. Human Capital Persistence and and Development. **American Economic Journal: Applied Economics**, Pittsburgh, v.9, n.4, p.105–136, 2017. page.5757

ROSOLI, G. Un Quadro Globale della Diaspora Italiana nelle Americhe. **Altreitelli**, Turim, v.8, p.1–15, 1992. page.5757

- SAES, F. A. M. **As ferrovias de São Paulo 1870-1940**. São Paulo: Hucitec, 1981. 200p. page.5757
- SAES, F. A. M. Estrada de Ferro e Diversificação da Atividade Econômica na Expansão Cafeeira em São Paulo, 1870-1900. In: SZMRECSANIY, T. (Ed.). **História Econômica da Independência e do Império**. São Paulo: Hucitec, 1996. p.177–196. page.5757
- SAIZ, A. Room in the Kitchen for the Melting Pot: immigration and rental prices. **Review of Economics and Statistics**, Cambridge, v.85, p.502–521, 2003. page.5757
- SCHWARCZ, L. M. **O Espetáculo das Raças**. São Paulo: Companhia das Letras, 1993. 296p. page.5757
- SCHÄFFER, N. O. Os alemães no Rio Grande do Sul: dos números iniciais aos censos demográficos. In: **Os alemães no sul do Brasil: cultura, etnicidade, história**. Canoas: Editora da Ulbra, 1984. p.163–184. page.5757
- SENA, R. A. D. **As Reformas Militares do Século XX Português: da primeira república ao fim do estado novo**. Lisboa: Universidade Nova de Lisboa, 2010. 109p. page.5757
- SEYFERTH, G. A Antropologia e a Teoria do Branqueamento da Raça no Brasil. **Revista do Museu Paulista**, São Paulo, v.30, p.81–98, 1985. page.5757
- SHLEIFER, A.; VISHNY, R. W. Corruption and growth. **Quarterly Journal of Economics**, London, v.108, p.599–617, 1993. page.5757
- SKIDMORE, T. E. **Preto no Branco: raça e nacionalidade no pensamento brasileiro**. Rio de Janeiro: Paz e Terra, 1976. 328p. page.5757
- SOLOMOU, S.; WU, W. Weather Effects on European Agricultural Output, 1850– 1913. **European Review of Economic History**, Oxford, v.3, p.351–373, 1999. page.5757
- SORI, E. **L’Emigrazione Italiana dalla’Unità alla Seconda Guerra Mondiale**. Bologna: Il Mulino, 1979. 512p. page.5757
- SOUZA, I. I. **Espanhóis: história e engajamento**. São Paulo: Companhia Editora Nacional, 2006. 104p. page.5757
- STAIGER, D.; STOCK, J. Instrumental Variables Regression with Weak Instruments. **Econometrica**, Medford, v.65(3), p.557–586, 1997. page.5757

- STEFANI, C. R. B. **O sistema ferroviário paulista: um estudo sobre a evolução do transporte de passageiros sobre trilhos.** São Paulo: Dissertação de Mestrado, Universidade de São Paulo, 2007. page.5757
- SÃO-PAULO(ESTADO). **Boletim do Departamento Estadual do Trabalho do Estado de São Paulo. Dados para a História da Imigração e da Colonização em São Paulo.** São Paulo, 1916. v.5, n.19. page.5757
- TAKEUCHI, M. Y. **Japoneses: a saga do povo do sol nascente.** São Paulo: Companhia Editora Nacional, 2007. 152p. page.5757
- TELAROLLI, R. J. Immigration and epidemics in the State of São Paulo. **História, Ciências, Saúde, Manguinhos**, v.III(2), p.265–283, 1996. page.5757
- TRUZZI, O. **Sírios e libaneses: narrativas de história e cultura.** São Paulo: Companhia Editora Nacional, 2005. 96p. page.5757
- UI. THE URBAN INSTITUTE. Children of Immigrants Data Tool. Dados do Integrated Public Use Microdata Series (IPUMS). **American Community Surveys**, Washington, DC, 2014. page.5757
- VAN OURS, J.; VEENMAN, J. The educational attainment of second-generation immigrants in The Netherlands. **Journal of Population Economics**, Maastricht, v.16(4), p.739–753, 2003. page.5656
- VANGELISTA, C. **Os Braços da Lavoura: imigrantes e caipiras na formação do mercado de trabalho paulista (1850-1930).** São Paulo: Hucitec, 1991. 293p. page.5757
- VASCONCELLOS, H. D. Oscilações do movimento imigratório no Brasil. **Revista de Imigração e Colonização**, Rio de Janeiro, v.1(2), p.23–47, 1994. page.5757
- VASCONCELOS, A. M. N.; BOTEAGA, T. **Política migratória e o paradoxo da globalização.** Porto Alegre: Editora da PUC-RS, 2015. 234p. page.5757
- VENDRAME, M. Em busca da República de Deus: revoltas camponesas e agentes da emigração no norte italiano. **Programa de Pós-Graduação em História da Universidade do Vale do Rio dos Sinos**, São Leopoldo, RS, p.22–42, 2016. page.5757
- VILELA, E. M. Sírios e libaneses: redes sociais, coesão e posição de status. **Revista Brasileira de Ciências Sociais**, São Paulo, v.26(76), p.157–225, 2011. page.5757

WEITZMAN, M. On Diversity. **Quarterly Journal of Economics**, Oxford, v.107(2), p.363–405, 1992. page.5757

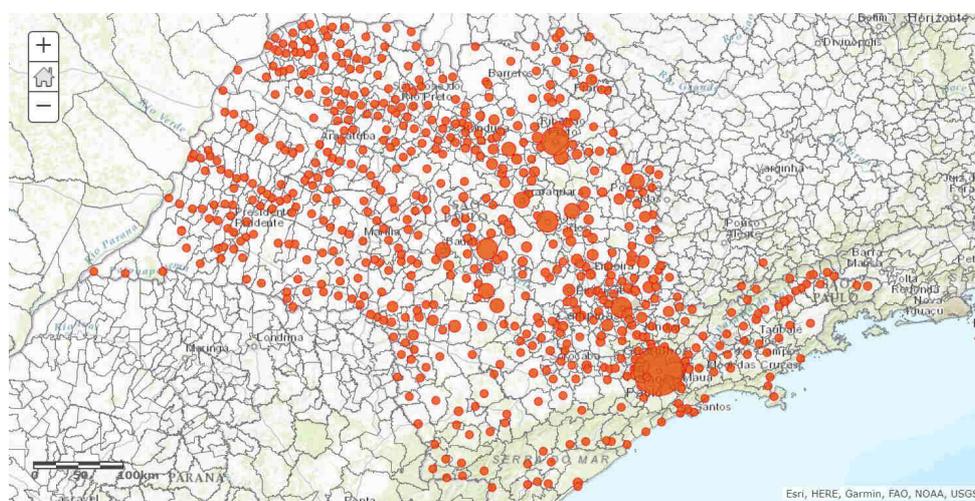
WINTERS, P.; JANVRY, A.; SADOULET, E. Family and community networks on Mexico-US migration. **Journal of Human Resources**, Wisconsin, v.6, p.159–184, 2001. page.5757

WRIGHT, G. The Origins of American Industrial Success: 1879–1940. **American Economic Review**, Pittsburgh, v.80, p.651–678, 1990. page.5757

ZAHRA, T. **The Great Departure**: mass migration from europe and the making of the free world. 1a.ed. New York: W. W. Norton Company, 2015. 416p. page.5757

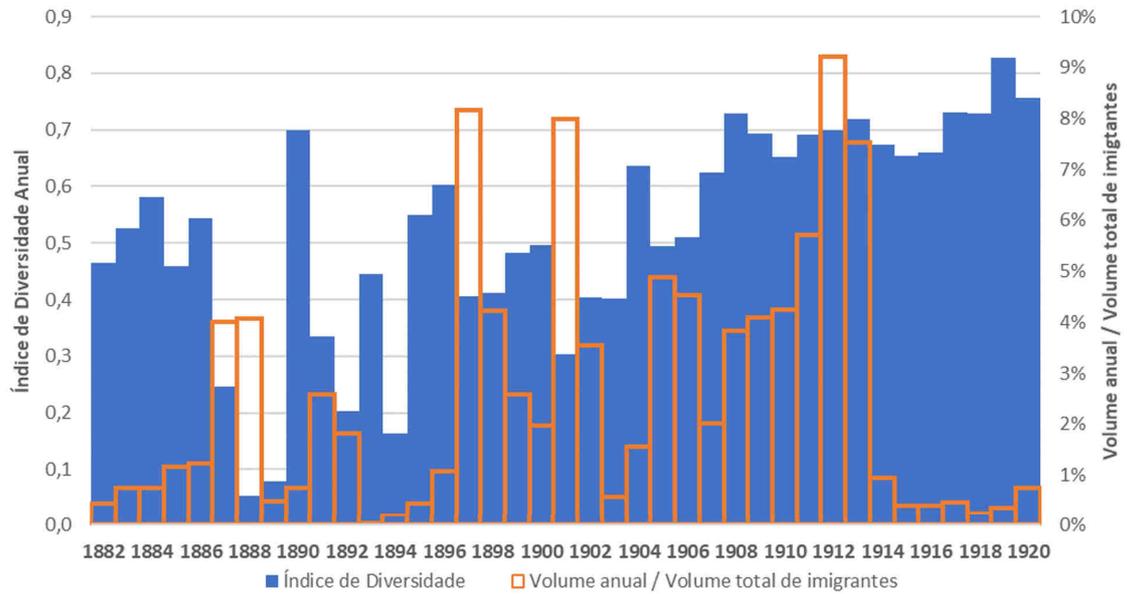
APÊNDICE A

Figura A.1: Distribuição dos imigrantes no Estado de São Paulo (1882-1920)



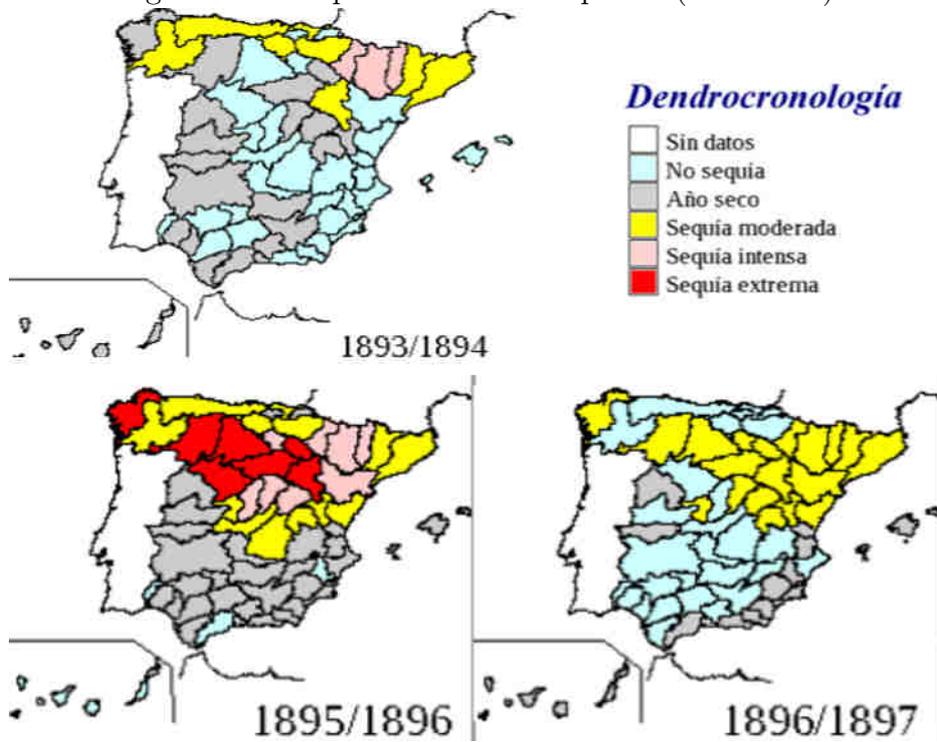
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

Figura A.2: Quantidades de Imigrantes e Índices Anuais de Diversidade de Origem (1882-1920).



Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

Figura A.3: Mapas da Seca na Espanha (1893-1897)



Fonte: ESPANHA (2013).

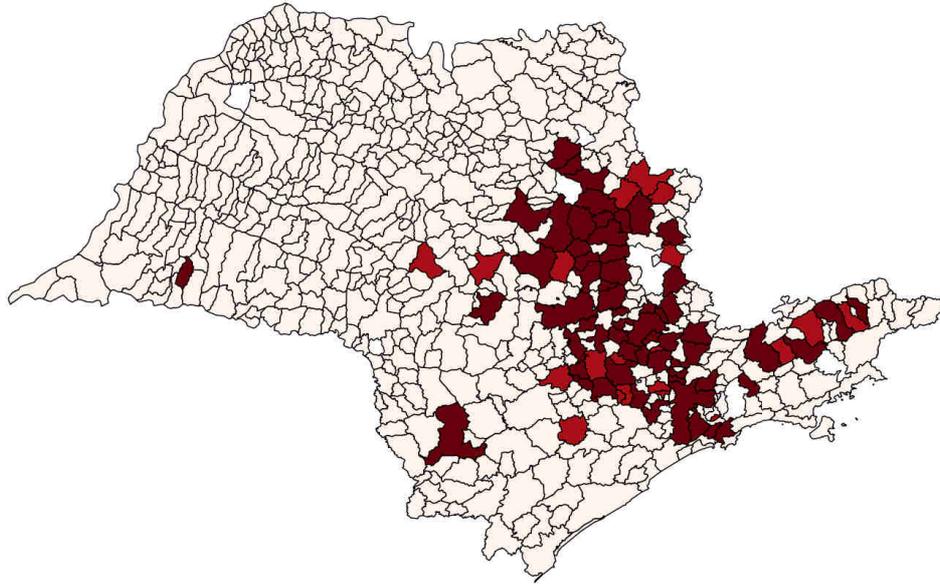


Figura A.4: Dispersão dos imigrantes - São Paulo, fluxo acumulado até 1885
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

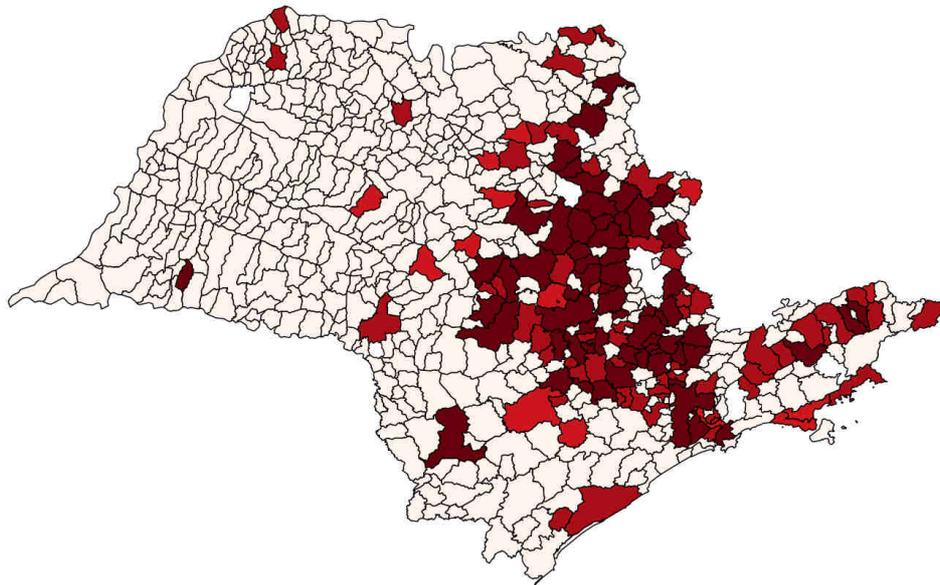


Figura A.5: Dispersão dos imigrantes - São Paulo, fluxo acumulado até 1890
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

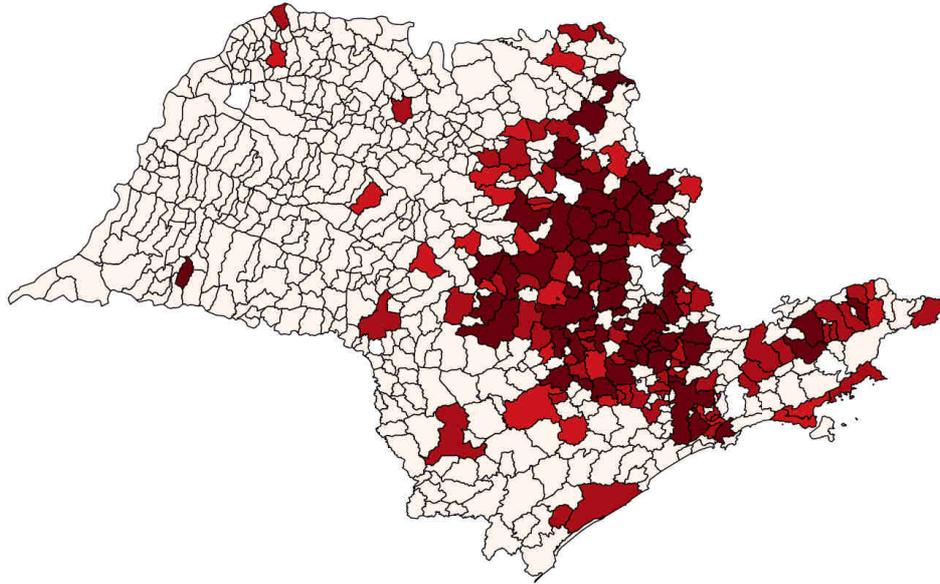


Figura A.6: Dispersão dos imigrantes - São Paulo, fluxo acumulado até 1895
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

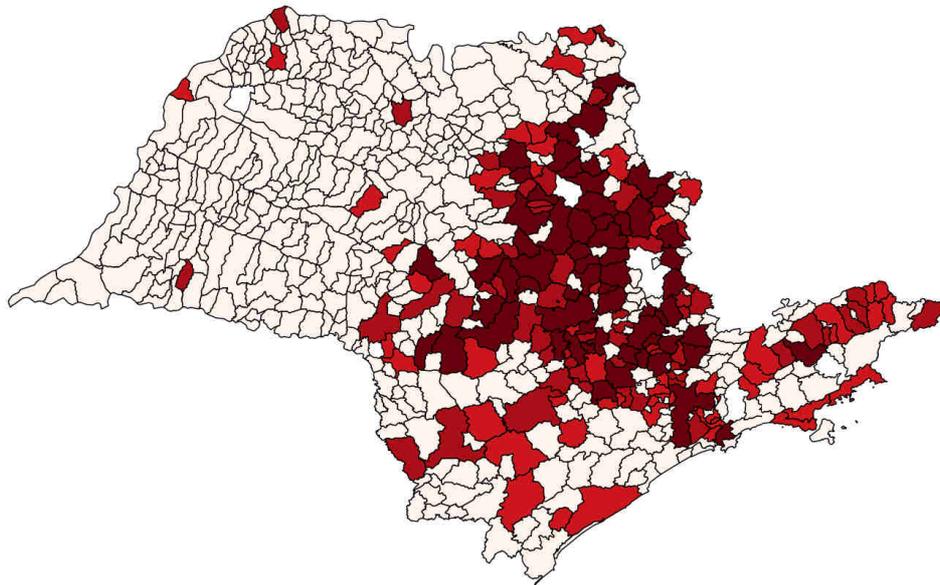


Figura A.7: Dispersão dos imigrantes - São Paulo, fluxo acumulado até 1900
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

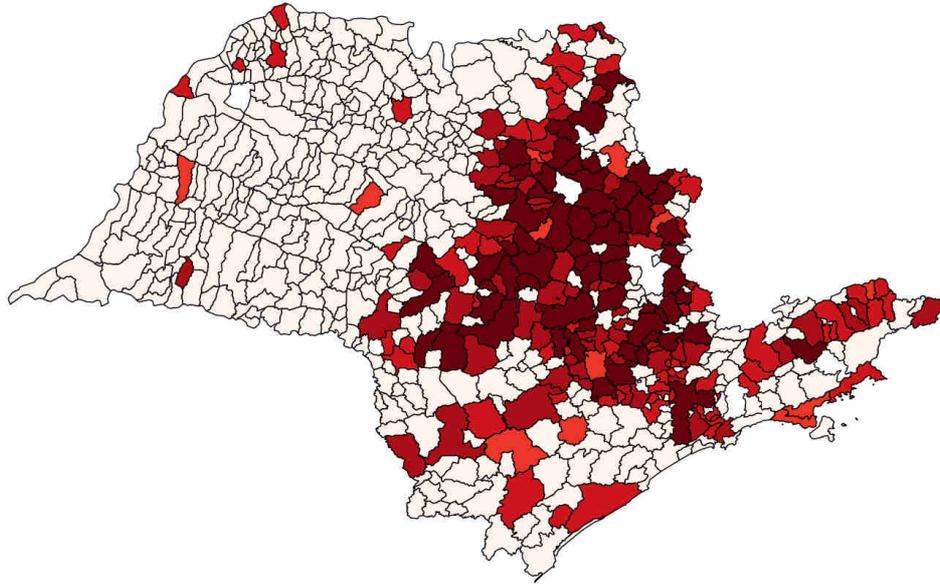


Figura A.8: Dispersão dos imigrantes - São Paulo, fluxo acumulado até 1905
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

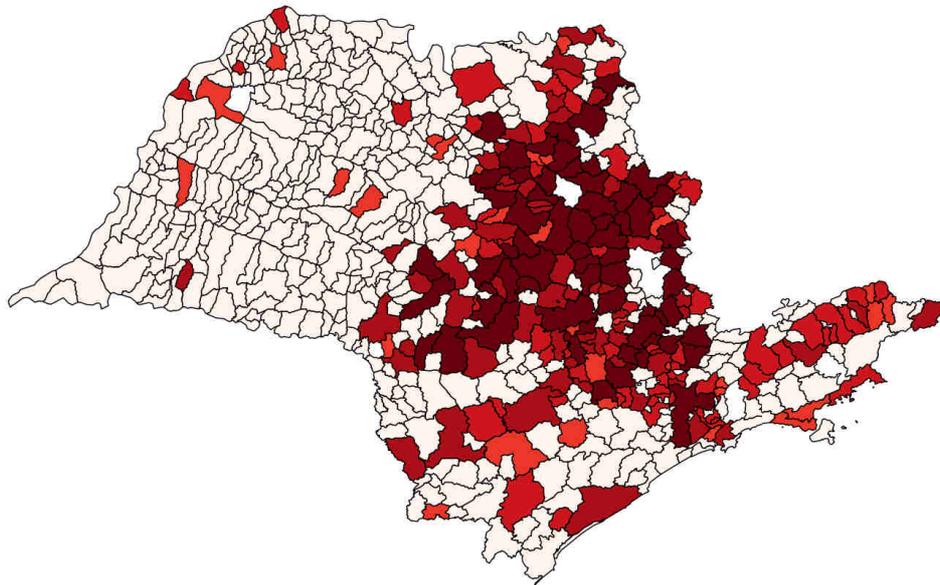


Figura A.9: Dispersão dos imigrantes - São Paulo, fluxo acumulado até 1910
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

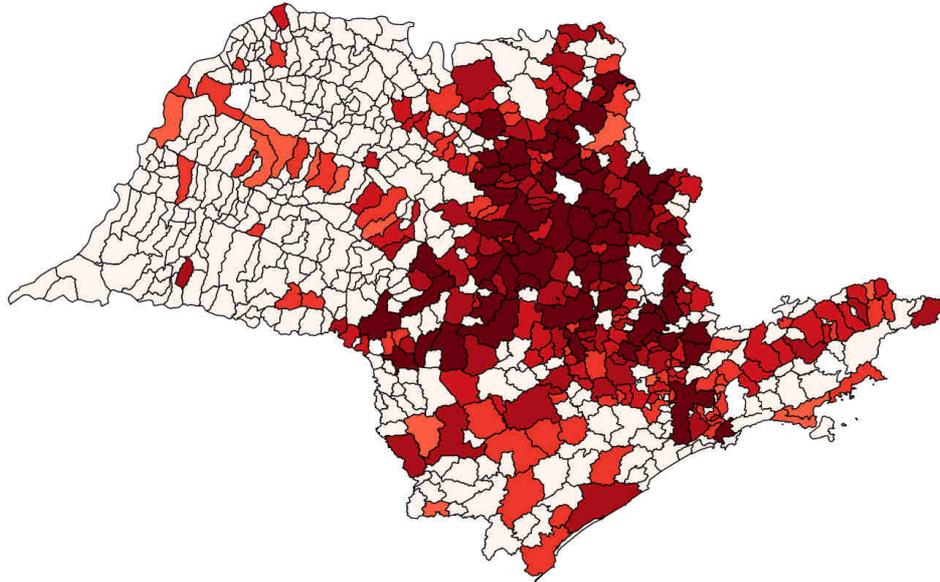


Figura A.10: Dispersão dos imigrantes - São Paulo, fluxo acumulado até 1915
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

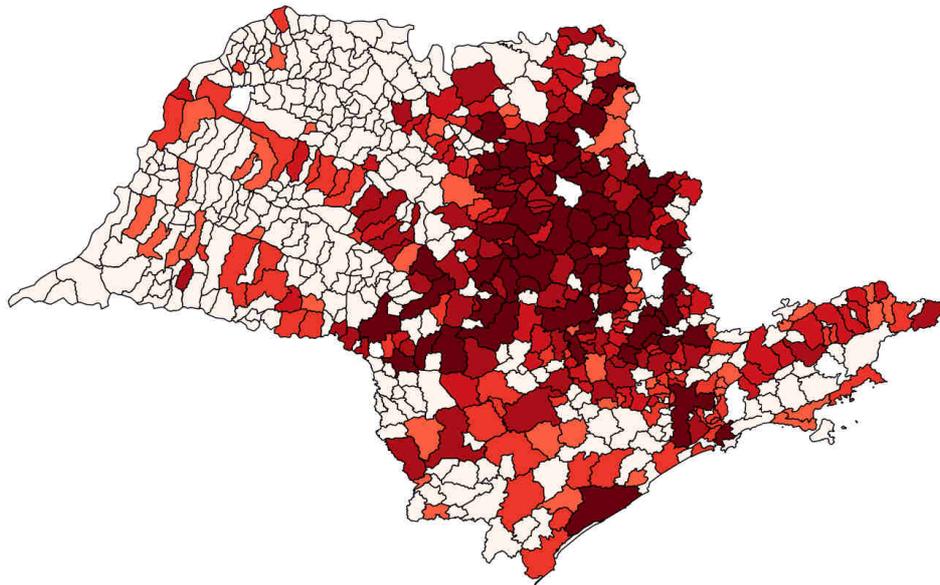


Figura A.11: Dispersão dos imigrantes - São Paulo, fluxo acumulado até 1920
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

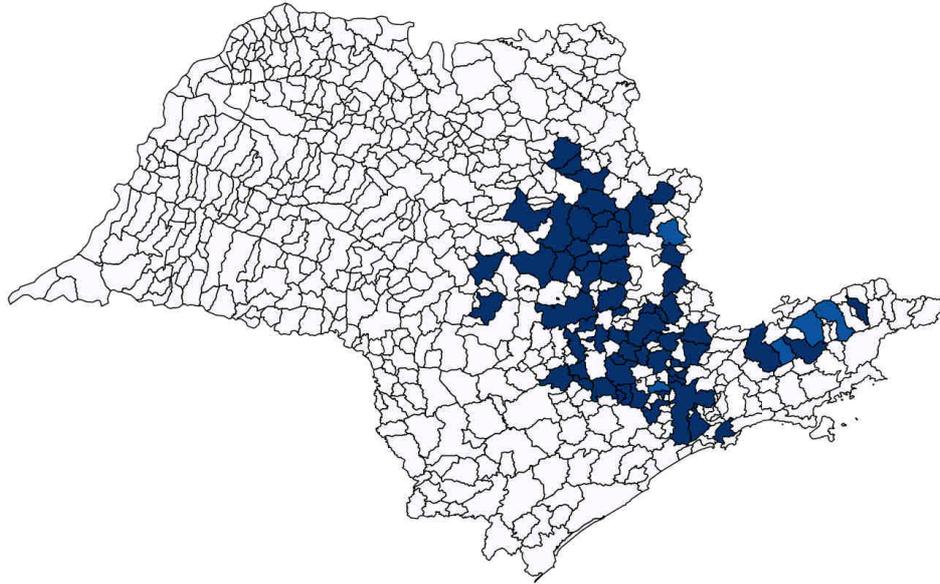


Figura A.12: Dispersão dos italianos - São Paulo, fluxo acumulado até 1885
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

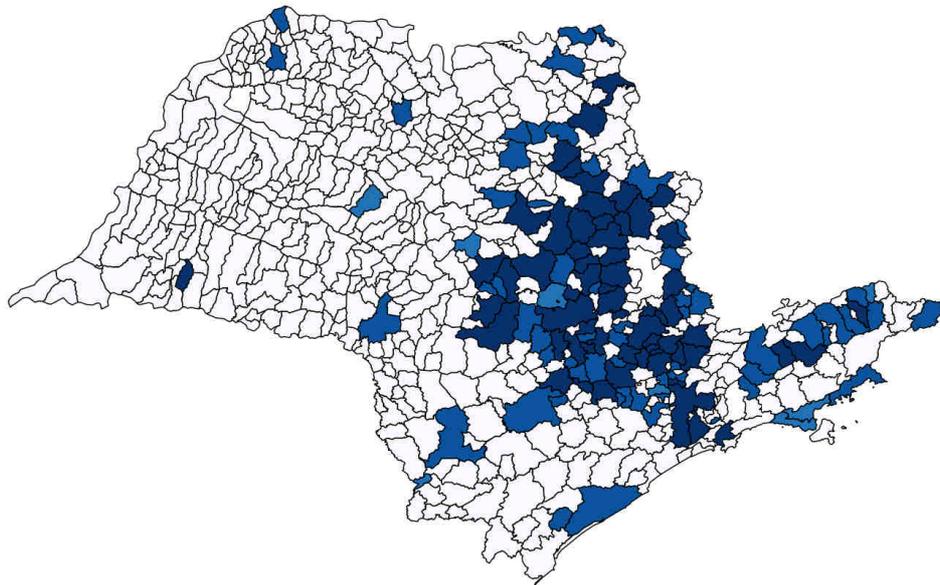


Figura A.13: Dispersão dos italianos - São Paulo, fluxo acumulado até 1890
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

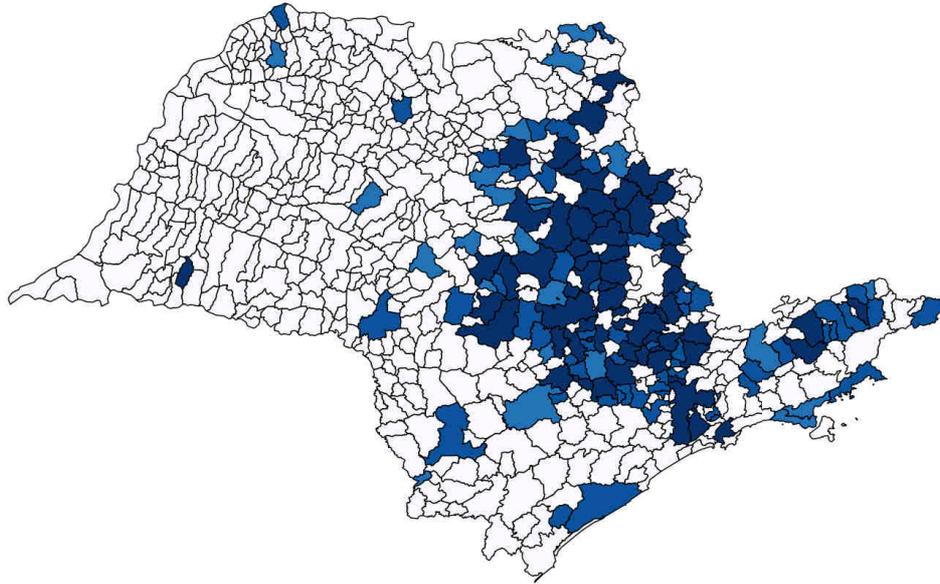


Figura A.14: Dispersão dos italianos - São Paulo, fluxo acumulado até 1895
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

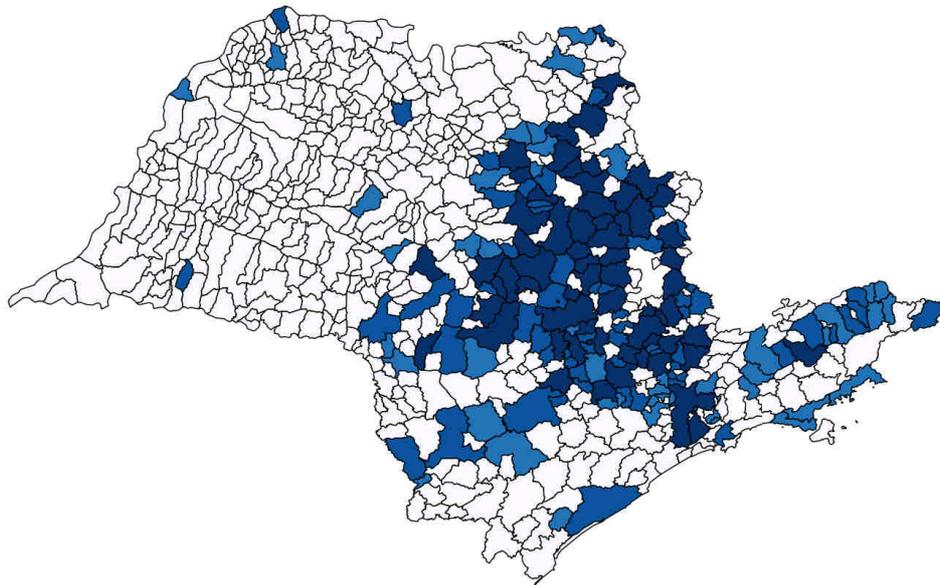


Figura A.15: Dispersão dos italianos - São Paulo, fluxo acumulado até 1900
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

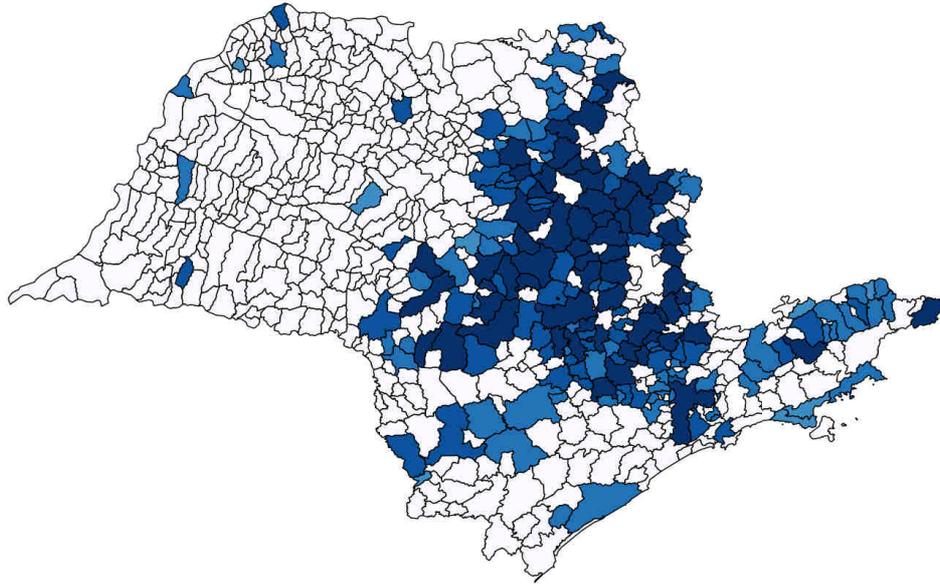


Figura A.16: Dispersão dos italianos - São Paulo, fluxo acumulado até 1905
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

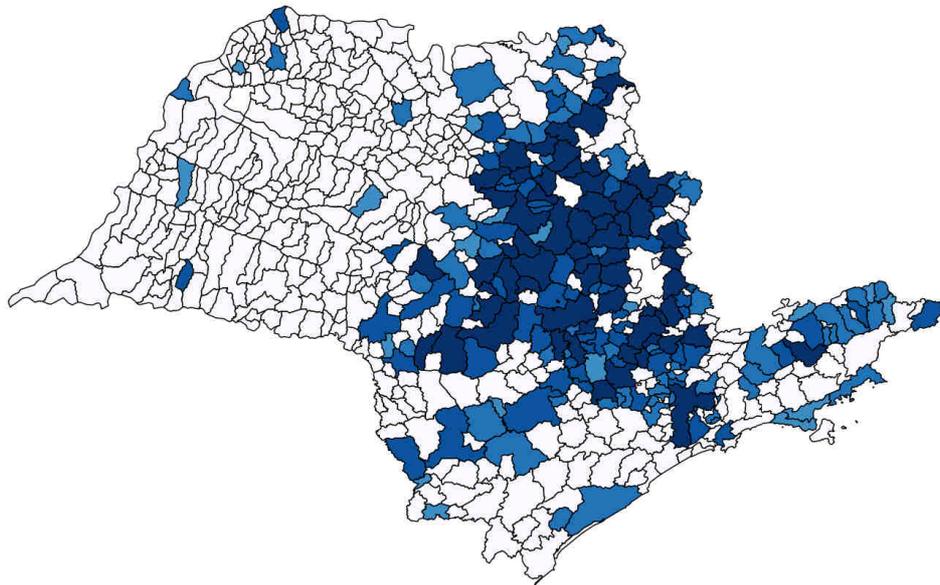


Figura A.17: Dispersão dos italianos - São Paulo, fluxo acumulado até 1910
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

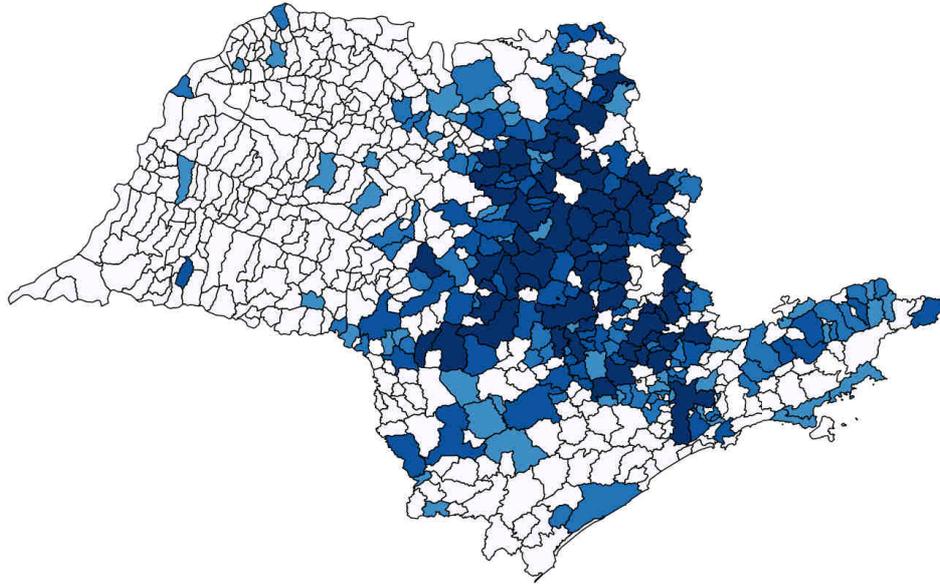


Figura A.18: Dispersão dos italianos - São Paulo, fluxo acumulado até 1915
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

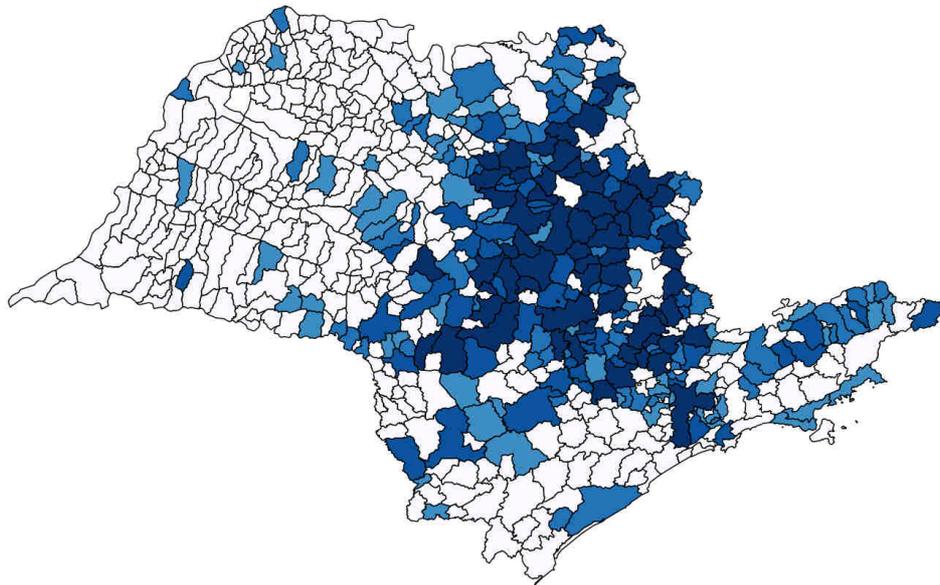


Figura A.19: Dispersão dos italianos - São Paulo, fluxo acumulado até 1920
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

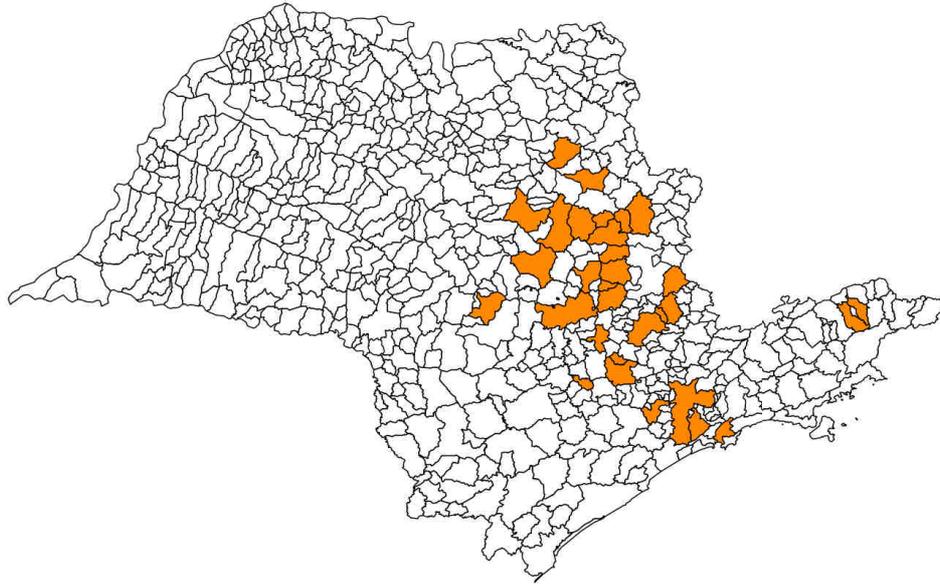


Figura A.20: Dispersão dos espanhóis - São Paulo, fluxo acumulado até 1885
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

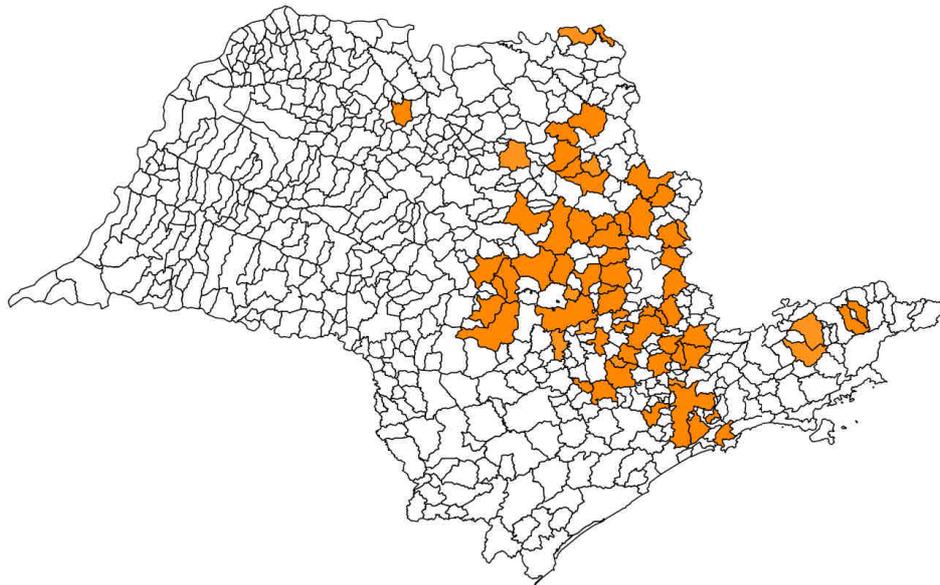


Figura A.21: Dispersão dos espanhóis - São Paulo, fluxo acumulado até 1890
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

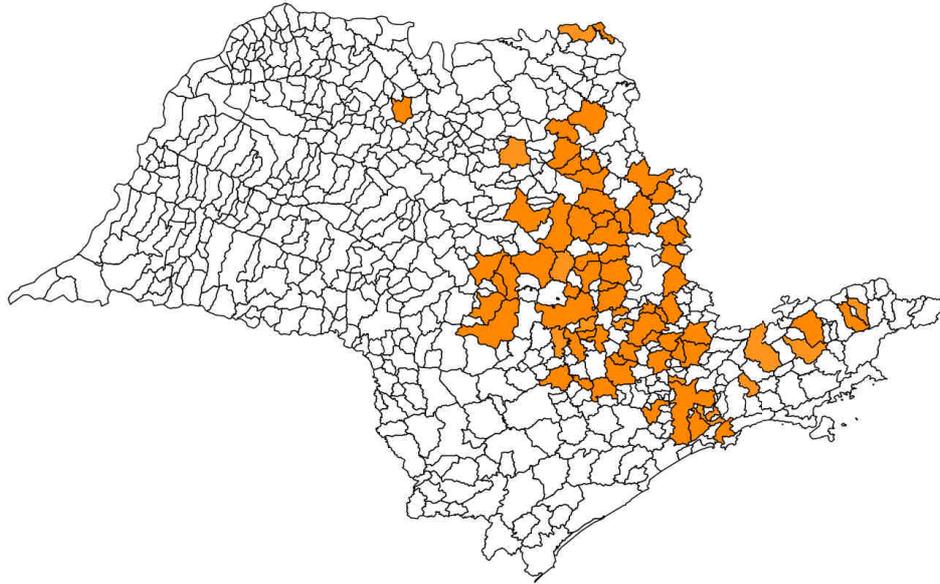


Figura A.22: Dispersão dos espanhóis - São Paulo, fluxo acumulado até 1895
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

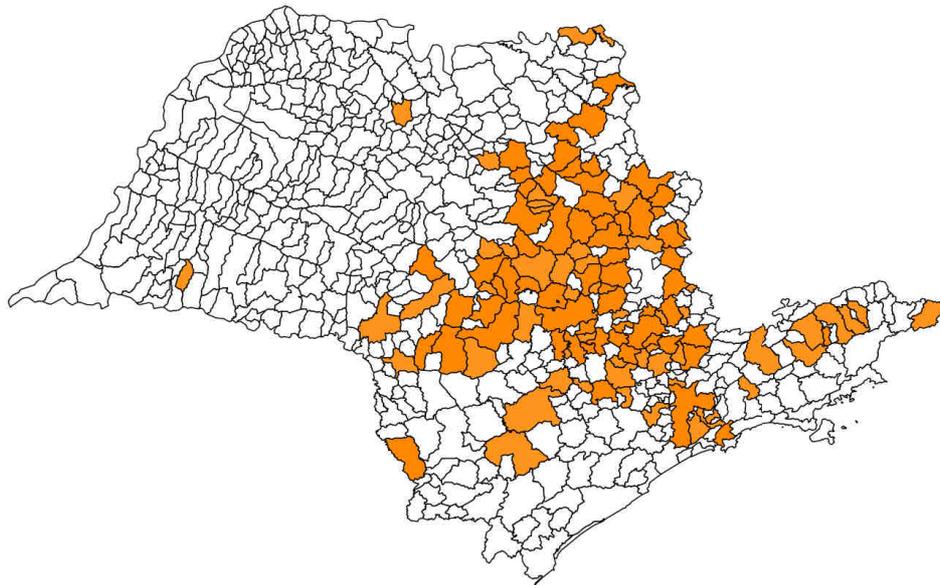


Figura A.23: Dispersão dos espanhóis - São Paulo, fluxo acumulado até 1900
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

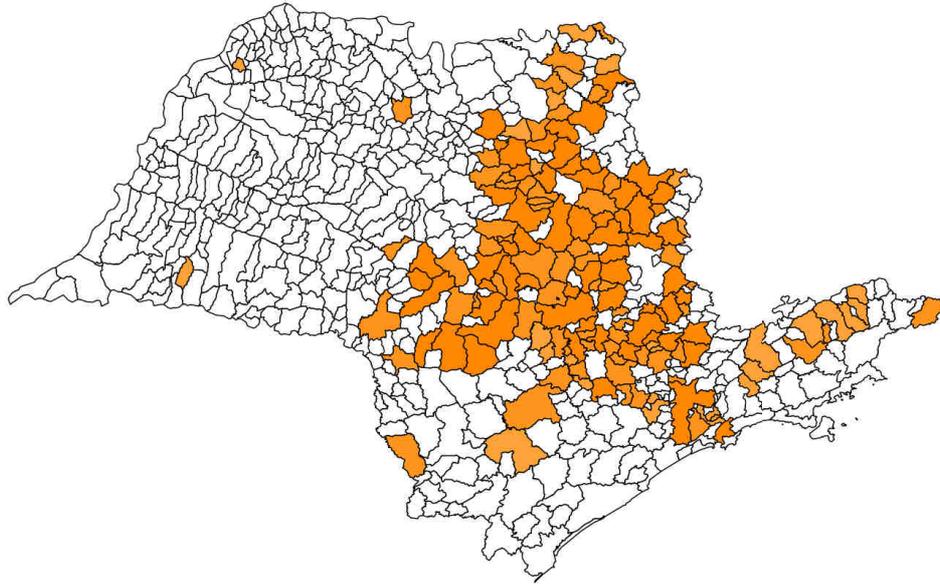


Figura A.24: Dispersão dos espanhóis - São Paulo, fluxo acumulado até 1905
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

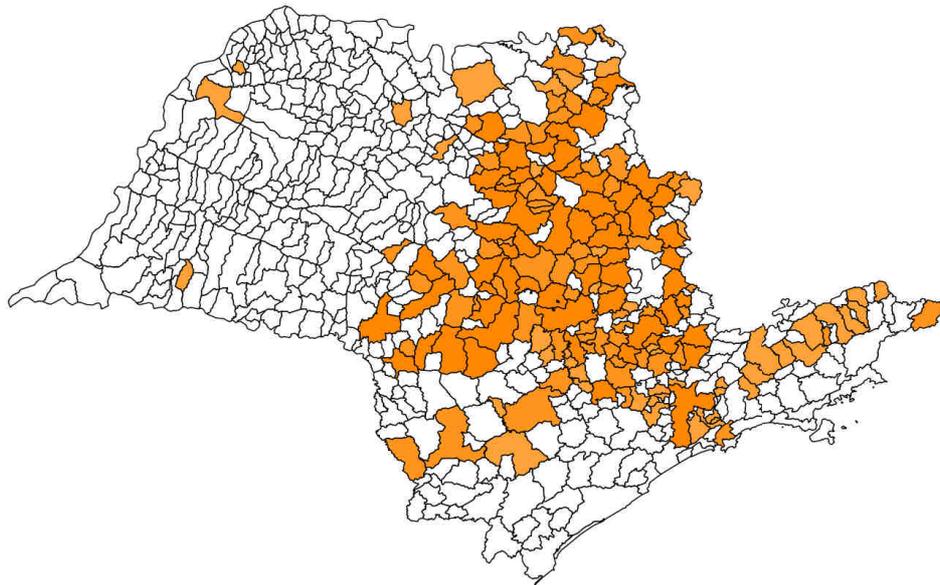


Figura A.25: Dispersão dos espanhóis - São Paulo, fluxo acumulado até 1910
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

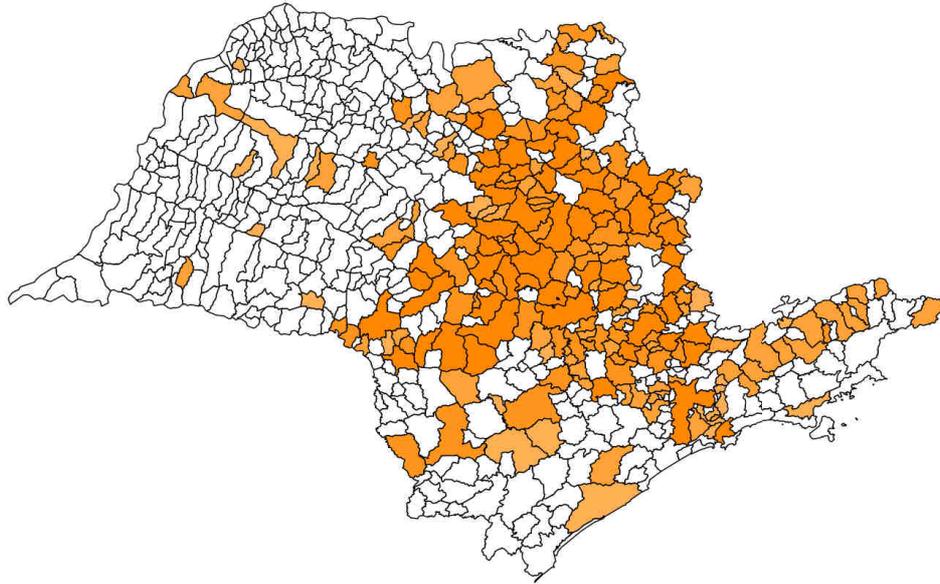


Figura A.26: Dispersão dos espanhóis - São Paulo, fluxo acumulado até 1915
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

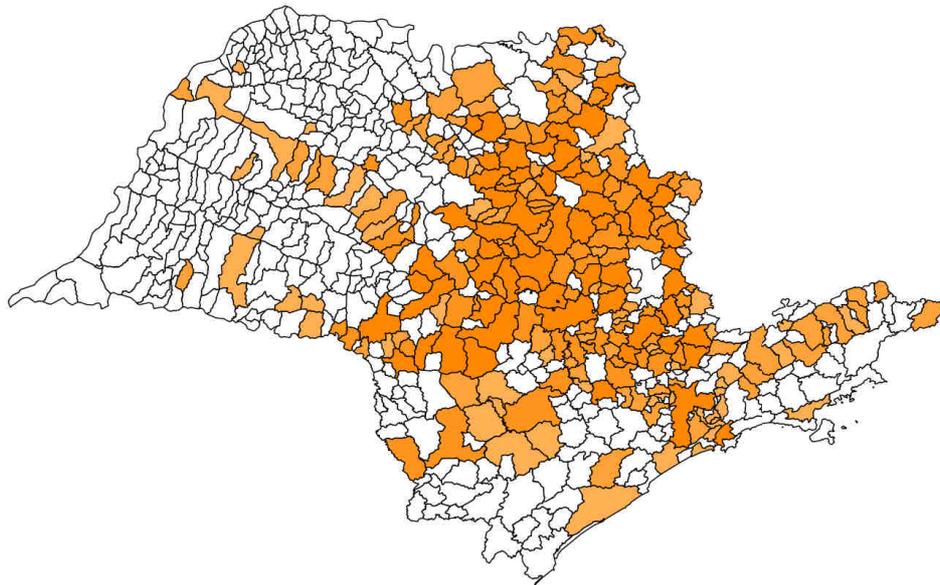


Figura A.27: Dispersão dos espanhóis - São Paulo, fluxo acumulado até 1920
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

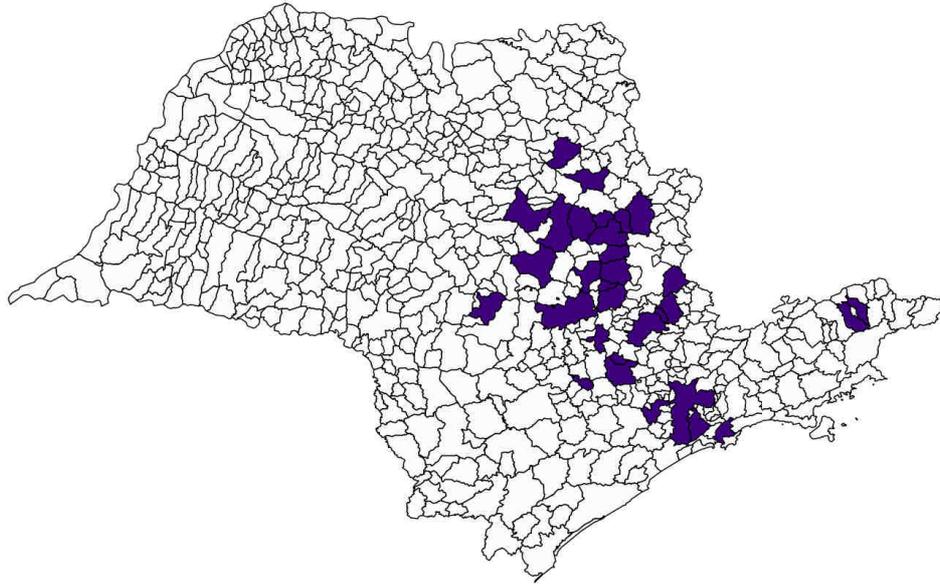


Figura A.28: Dispersão dos portugueses - São Paulo, fluxo acumulado até 1885
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

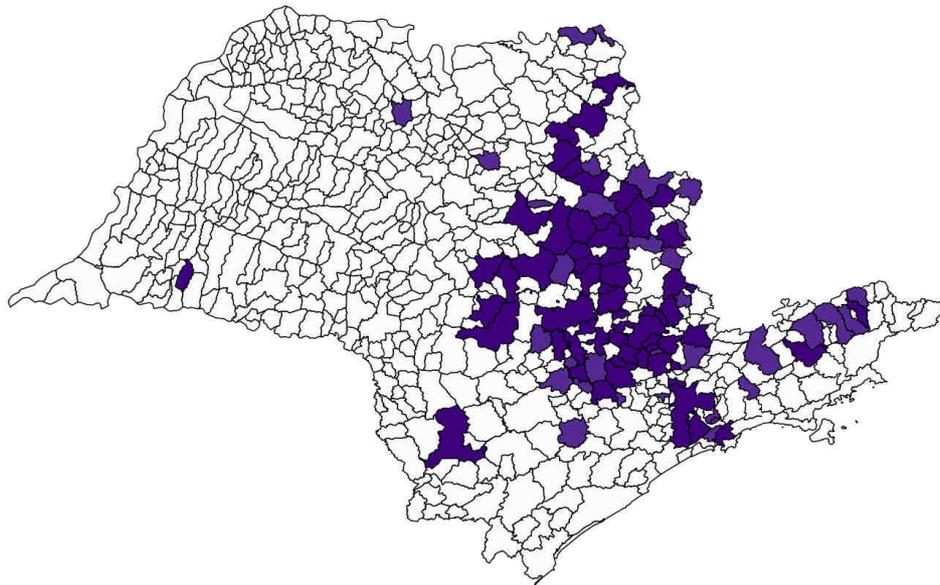


Figura A.29: Dispersão dos portugueses - São Paulo, fluxo acumulado até 1890
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

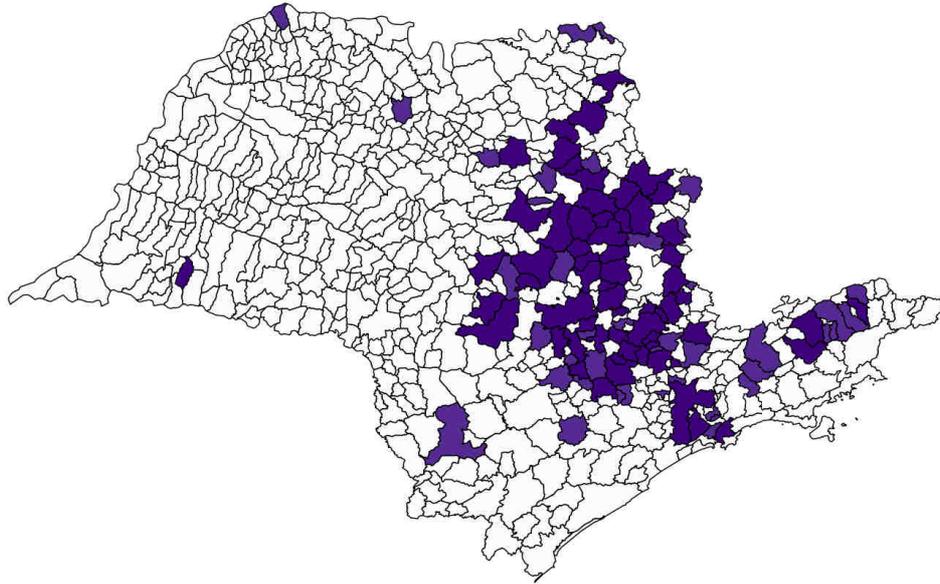


Figura A.30: Dispersão dos portugueses - São Paulo, fluxo acumulado até 1895
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

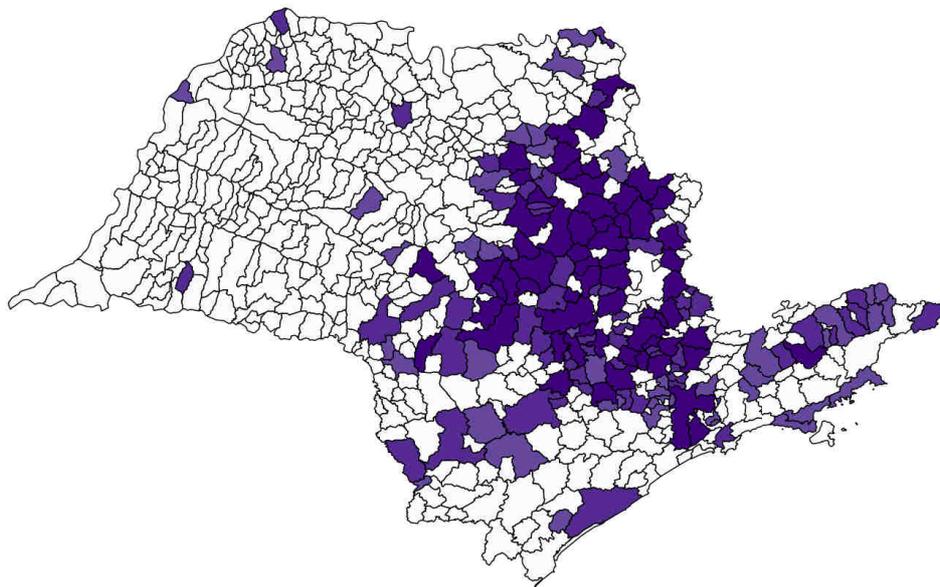


Figura A.31: Dispersão dos portugueses - São Paulo, fluxo acumulado até 1900
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

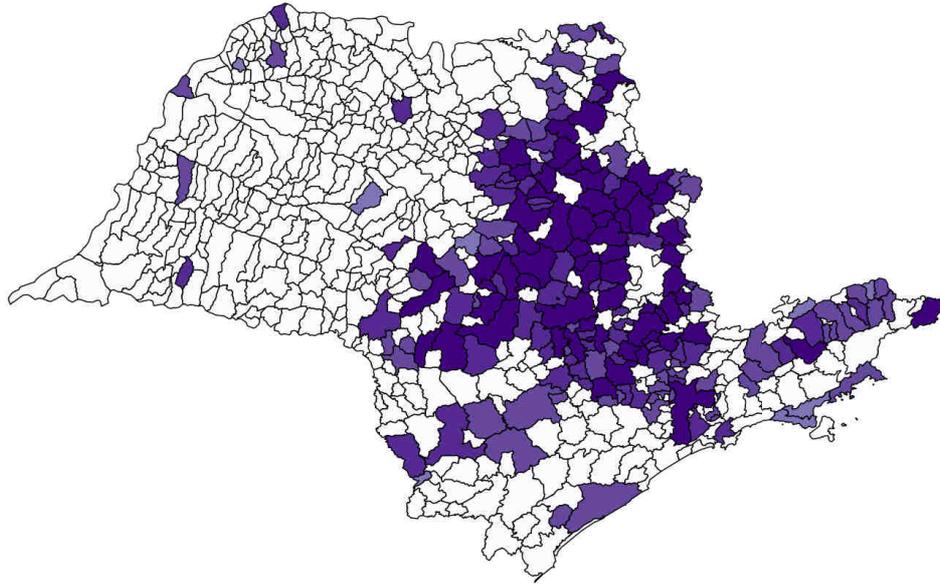


Figura A.32: Dispersão dos portugueses - São Paulo, fluxo acumulado até 1905
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

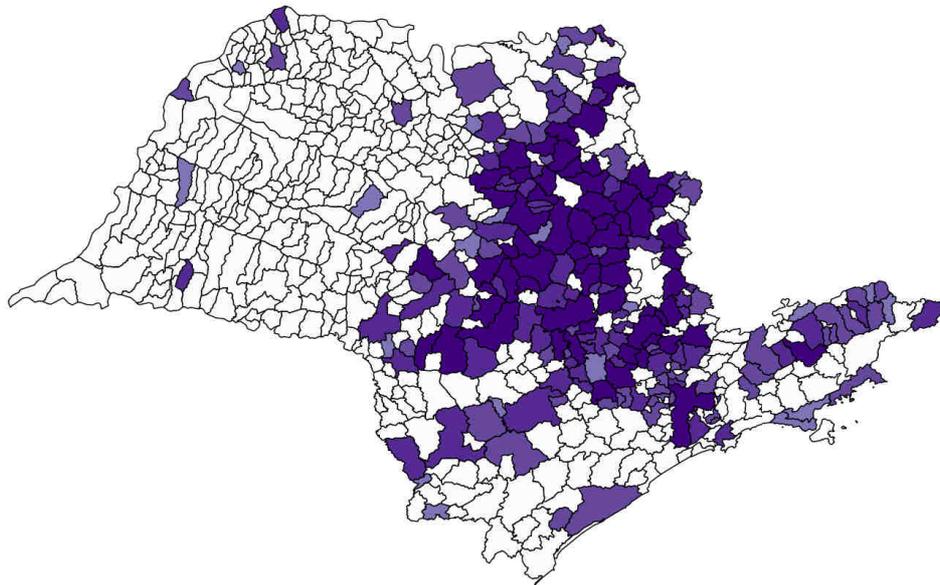


Figura A.33: Dispersão dos portugueses - São Paulo, fluxo acumulado até 1910
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

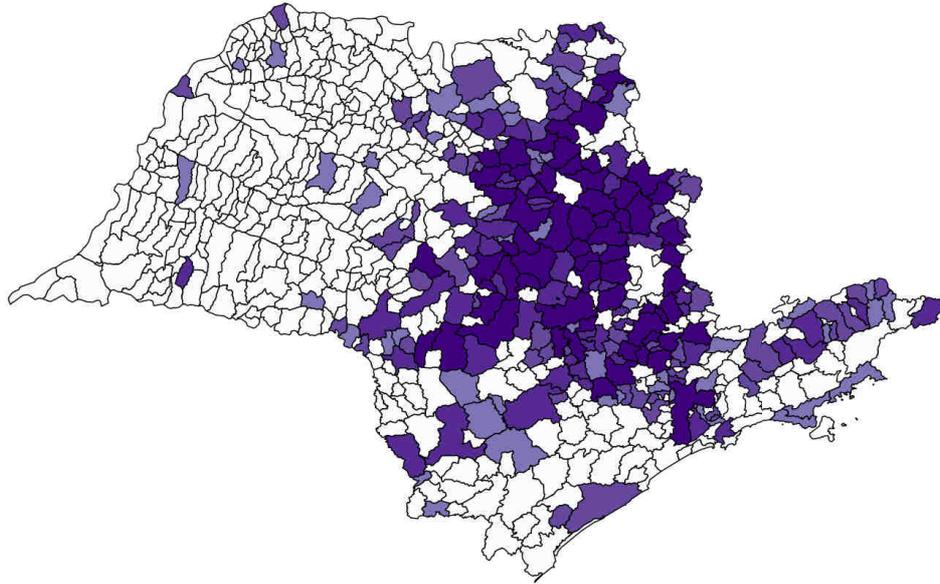


Figura A.34: Dispersão dos portugueses - São Paulo, fluxo acumulado até 1915
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

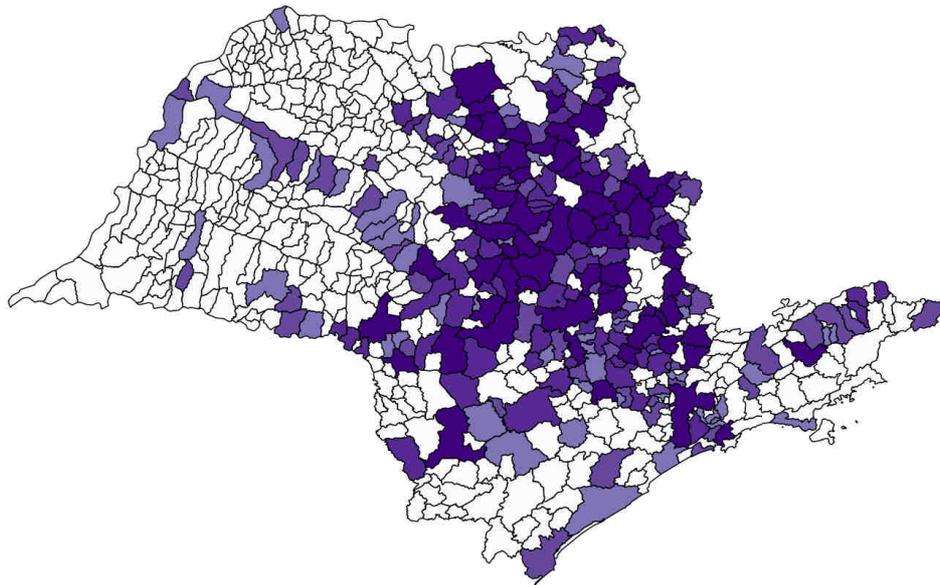


Figura A.35: Dispersão dos portugueses - São Paulo, fluxo acumulado até 1920
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

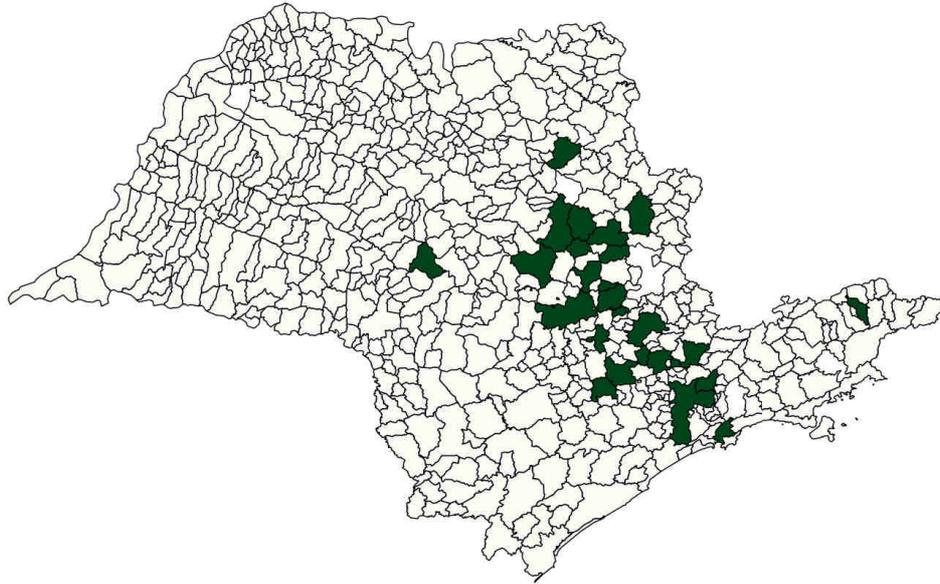


Figura A.36: Dispersão dos alemães - São Paulo, fluxo acumulado até 1885
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

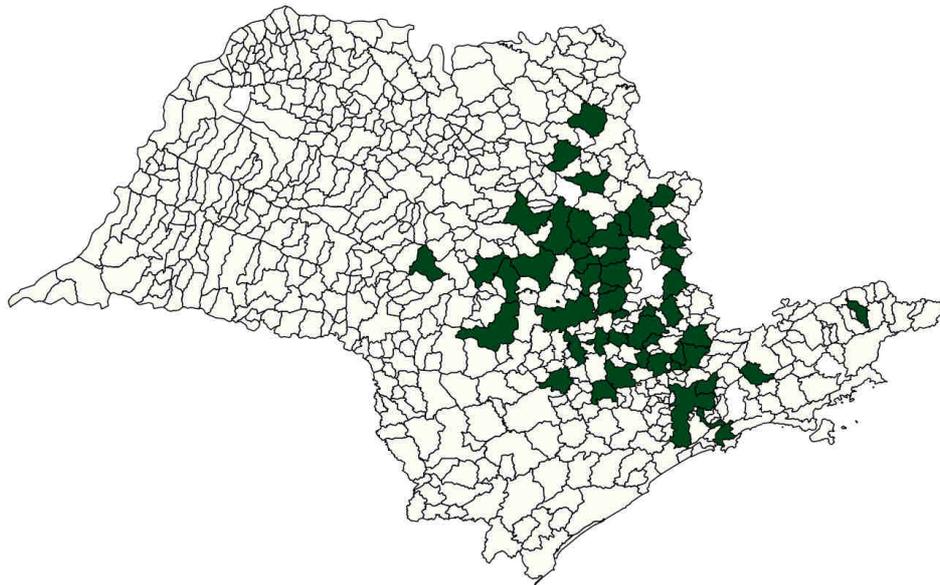


Figura A.37: Dispersão dos alemães - São Paulo, fluxo acumulado até 1890
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

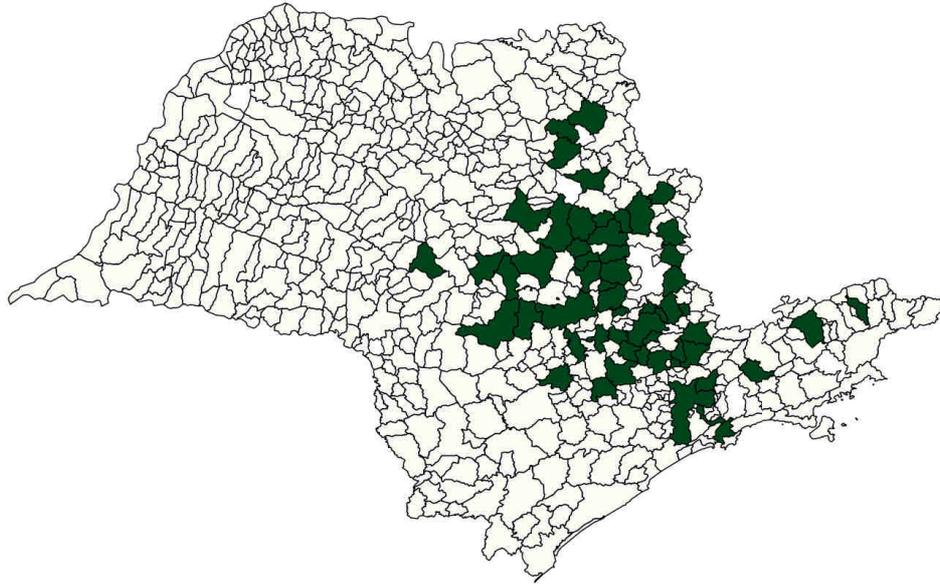


Figura A.38: Dispersão dos alemães - São Paulo, fluxo acumulado até 1895
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

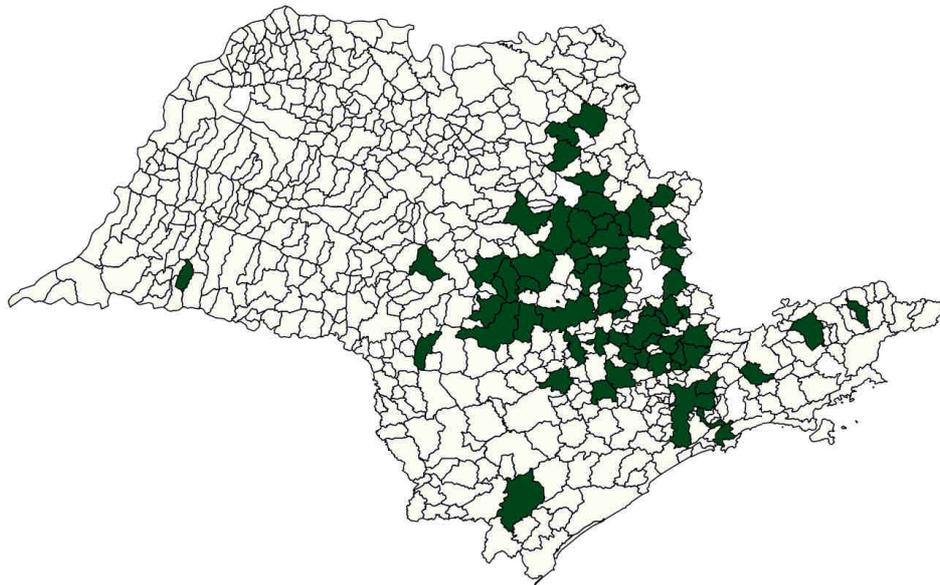


Figura A.39: Dispersão dos alemães - São Paulo, fluxo acumulado até 1900
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

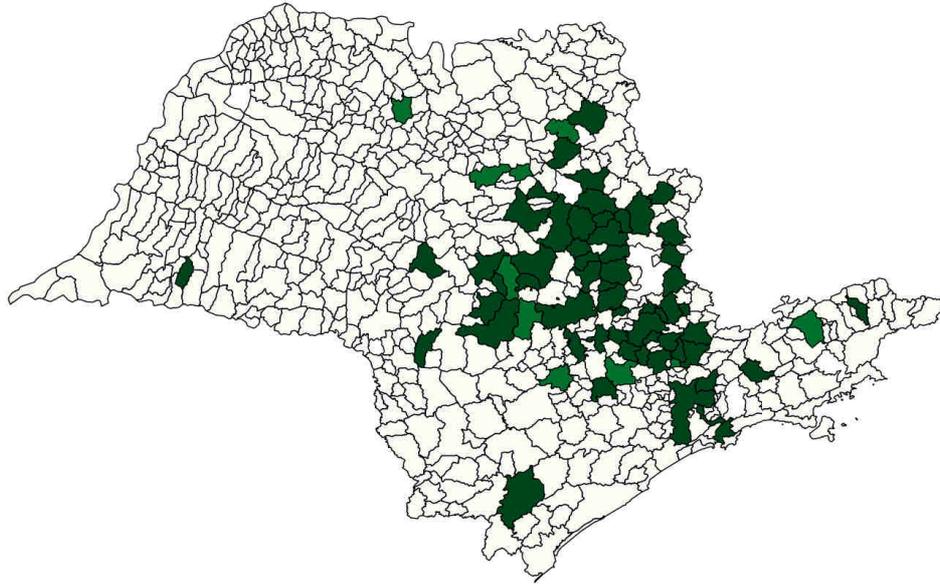


Figura A.40: Dispersão dos alemães - São Paulo, fluxo acumulado até 1905
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

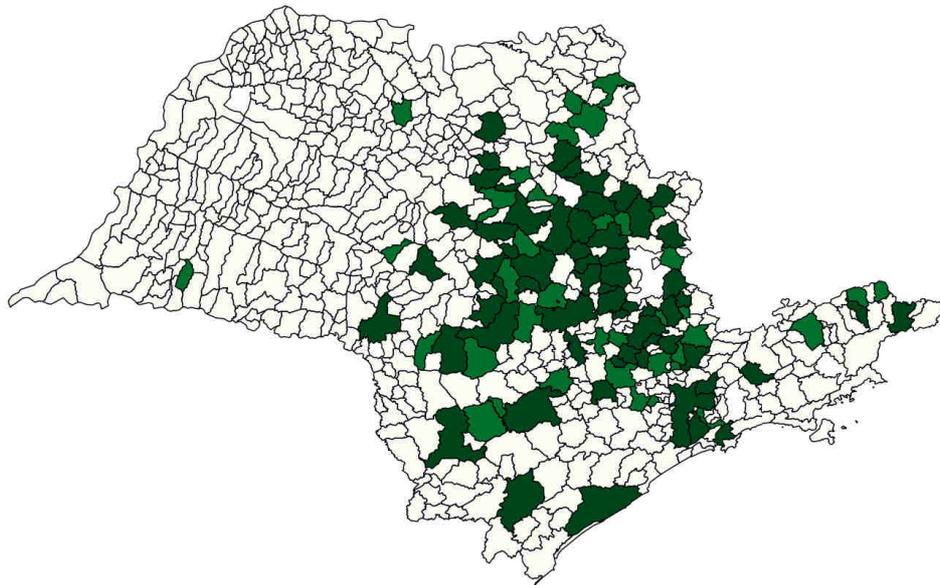


Figura A.41: Dispersão dos alemães - São Paulo, fluxo acumulado até 1910
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

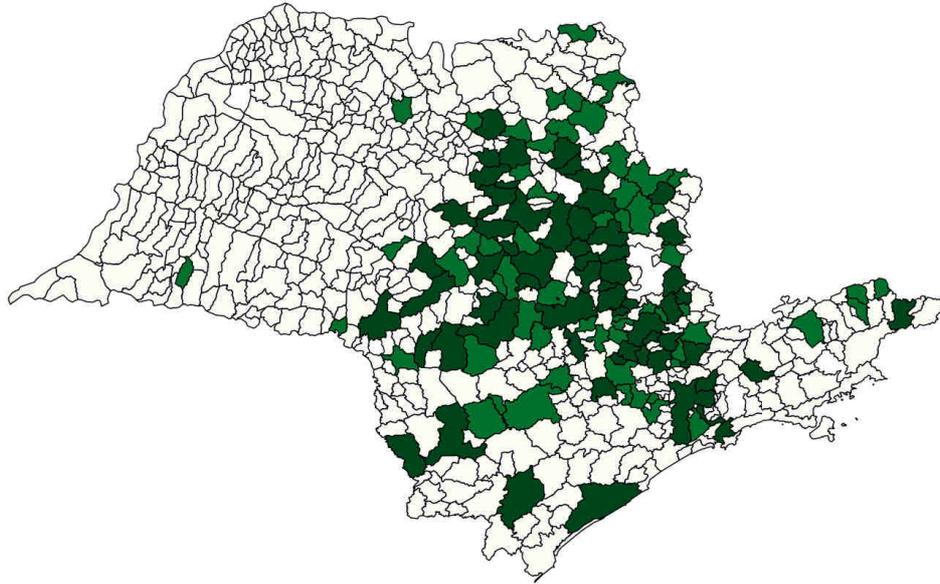


Figura A.42: Dispersão dos alemães - São Paulo, fluxo acumulado até 1915
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

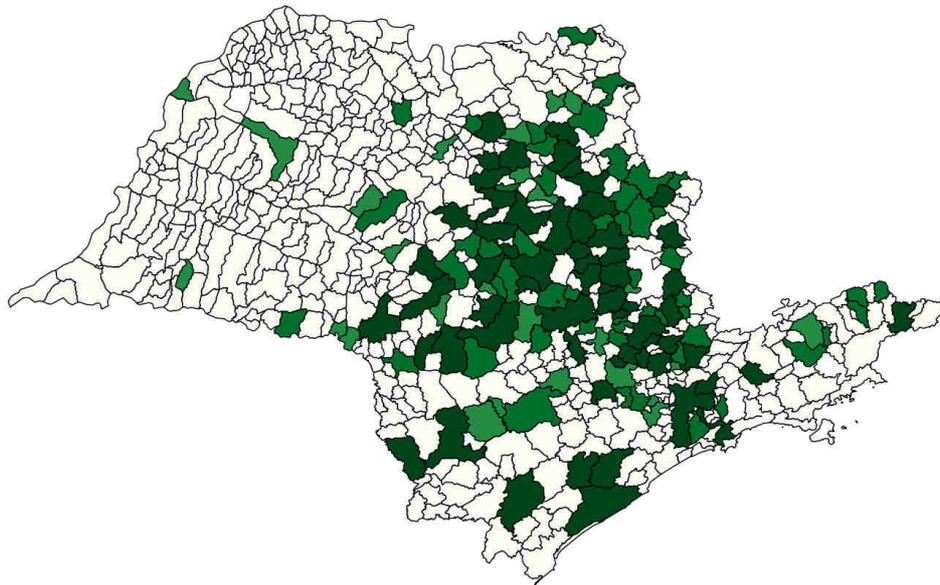


Figura A.43: Dispersão dos alemães - São Paulo, fluxo acumulado até 1920
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.



Figura A.44: Dispersão dos japoneses - São Paulo, fluxo acumulado até 1905
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

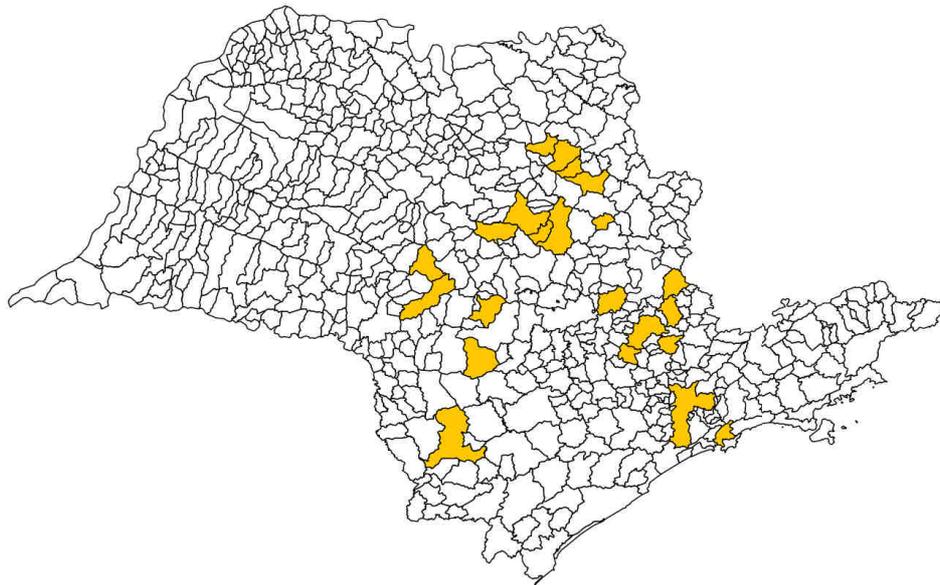


Figura A.45: Dispersão dos japoneses - São Paulo, fluxo acumulado até 1910
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

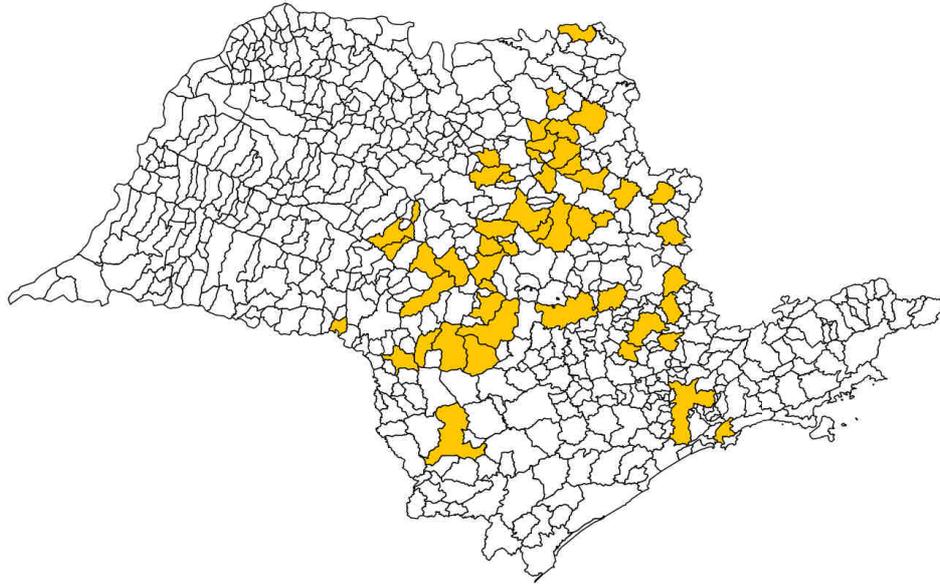


Figura A.46: Dispersão dos japoneses - São Paulo, fluxo acumulado até 1915
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

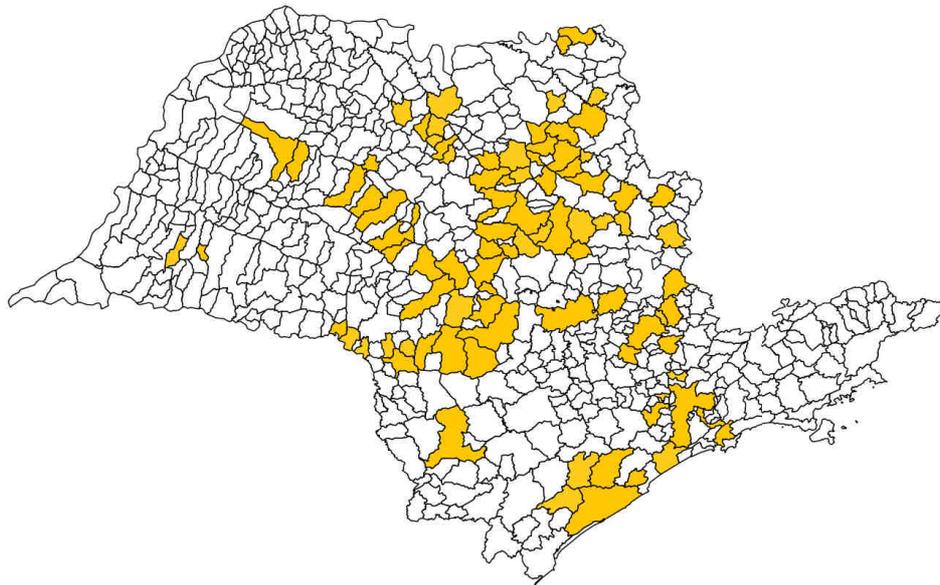


Figura A.47: Dispersão dos japoneses - São Paulo, fluxo acumulado até 1920
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

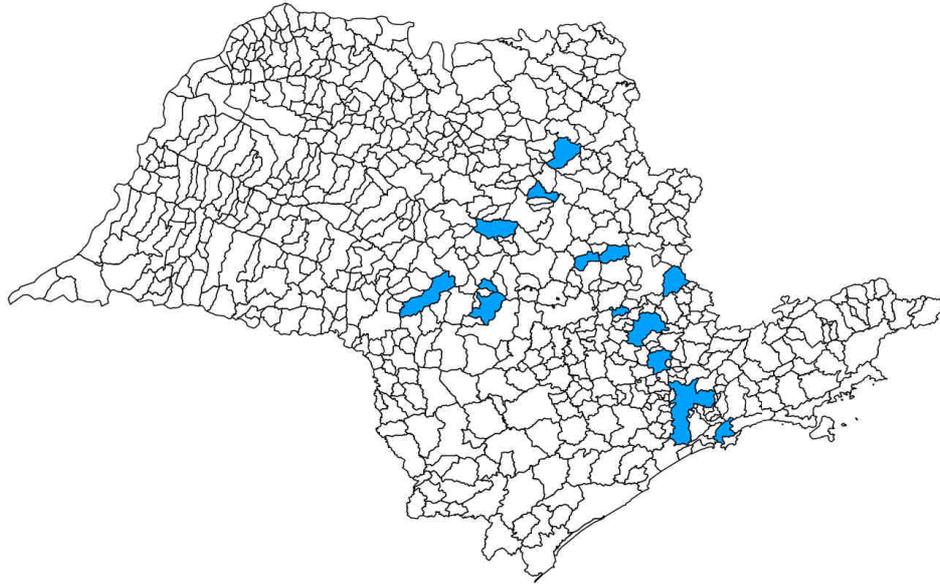


Figura A.48: Dispersão dos russos - São Paulo, fluxo acumulado até 1905
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

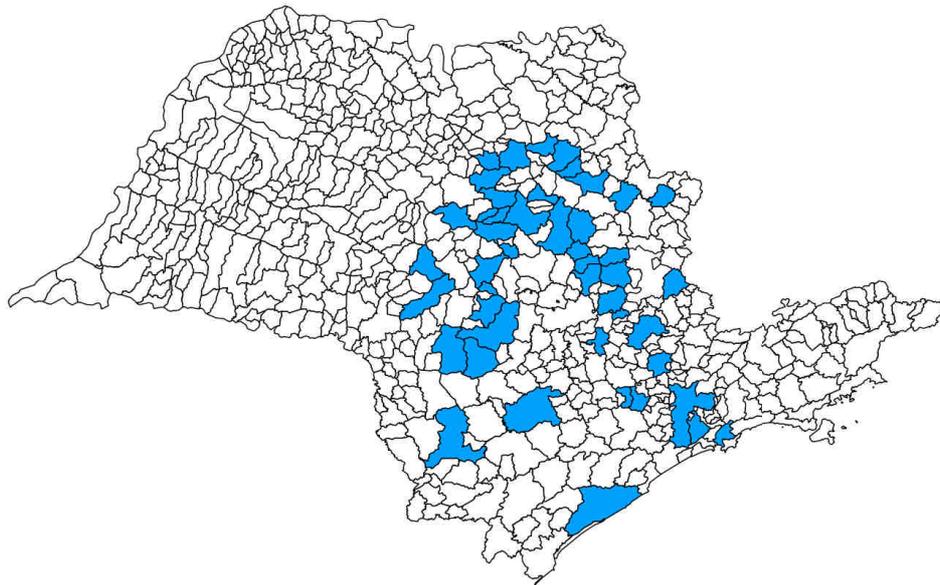


Figura A.49: Dispersão dos russos - São Paulo, fluxo acumulado até 1910
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

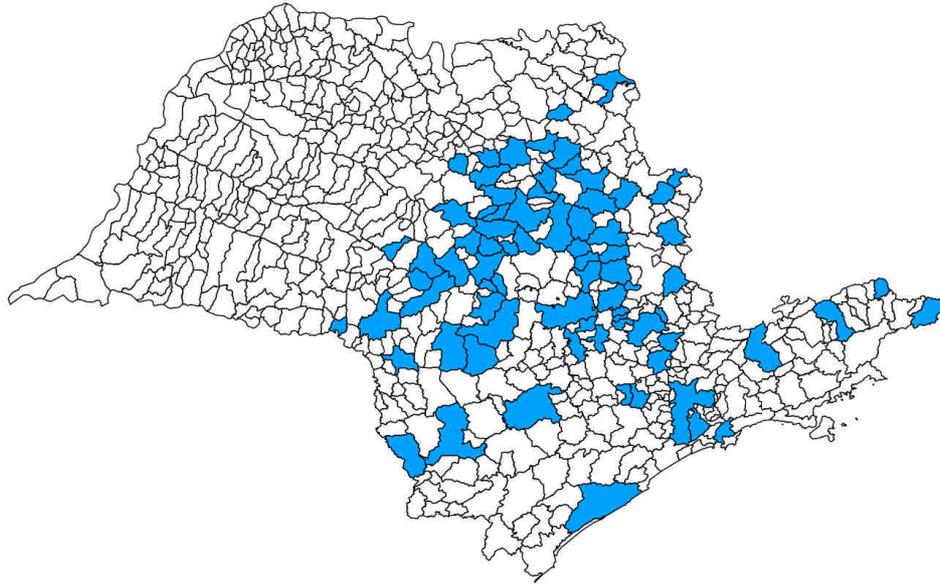


Figura A.50: Dispersão dos russos - São Paulo, fluxo acumulado até 1915
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

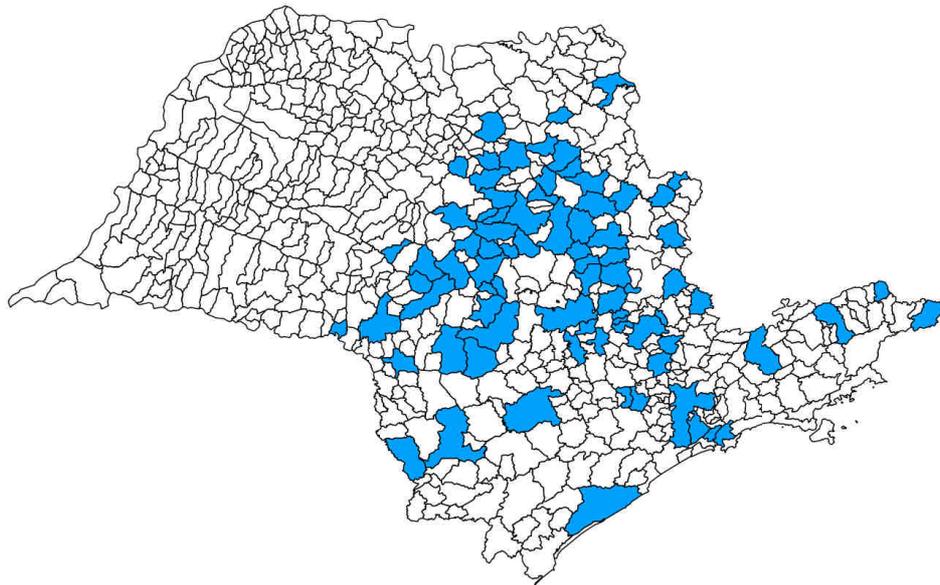


Figura A.51: Dispersão dos russos - São Paulo, fluxo acumulado até 1920
Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

1890

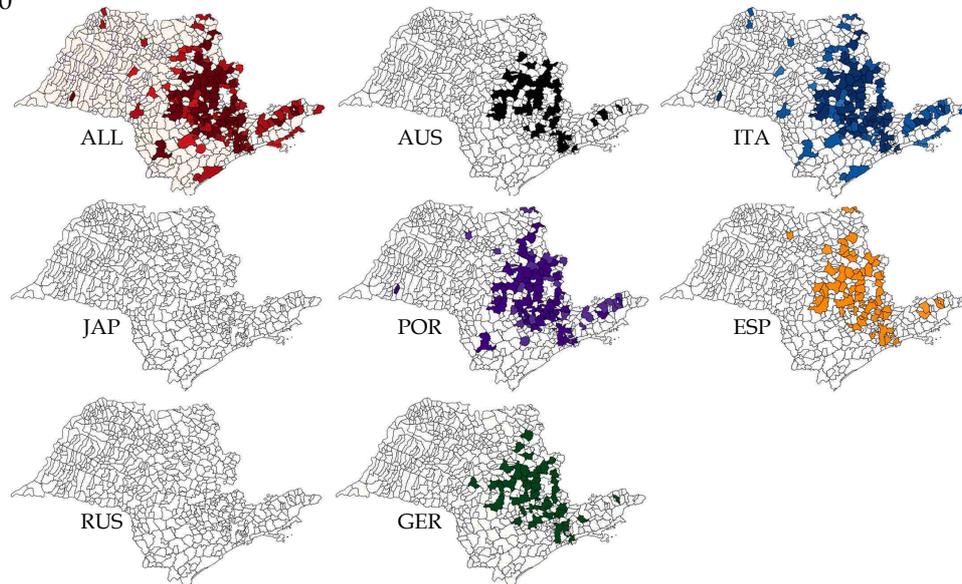


Figura A.52: Dispersão dos imigrantes por nacionalidade - São Paulo, fluxo acumulado até 1890

Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

1900

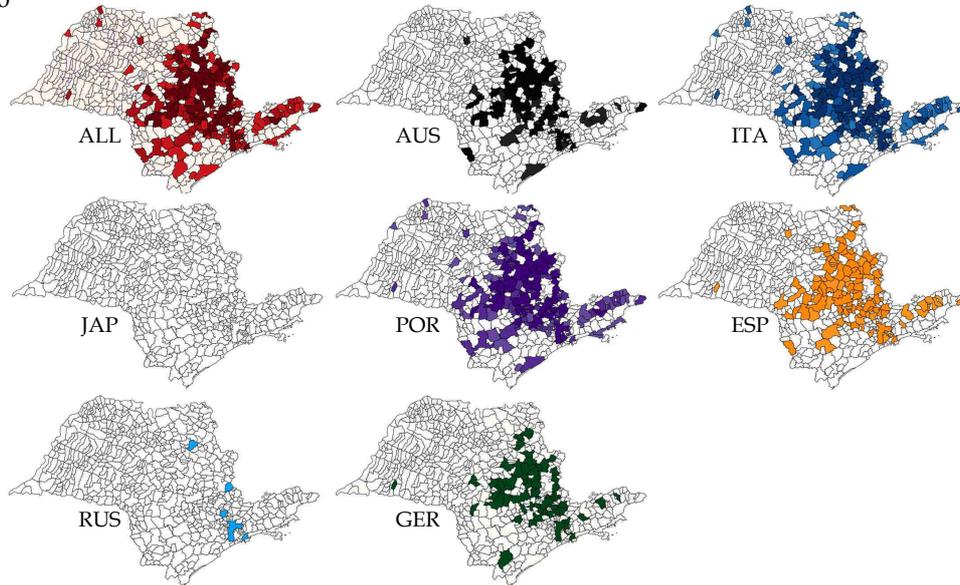


Figura A.53: Dispersão dos imigrantes por nacionalidade - São Paulo, fluxo acumulado até 1900

Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

1910

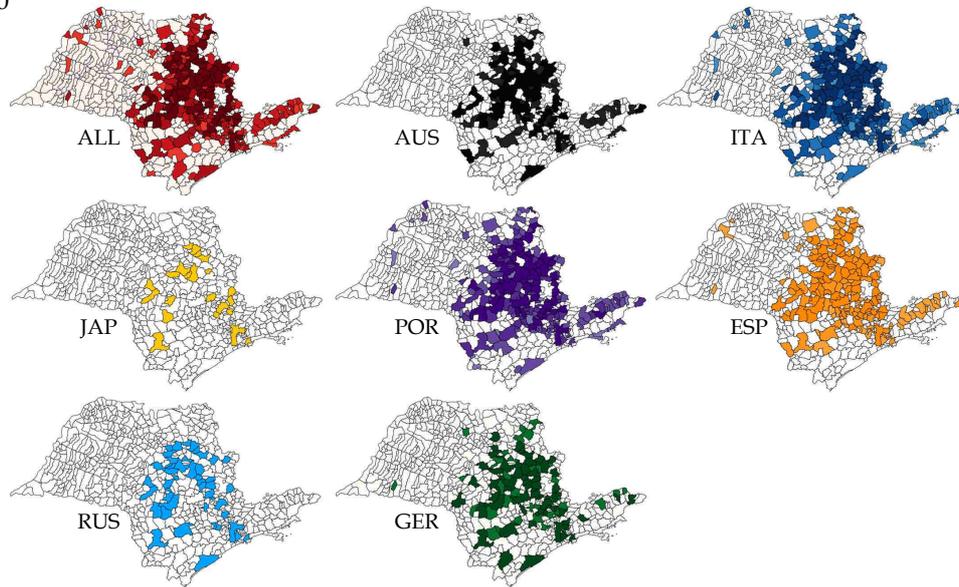


Figura A.54: Dispersão dos imigrantes por nacionalidade - São Paulo, fluxo acumulado até 1910

Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

1920

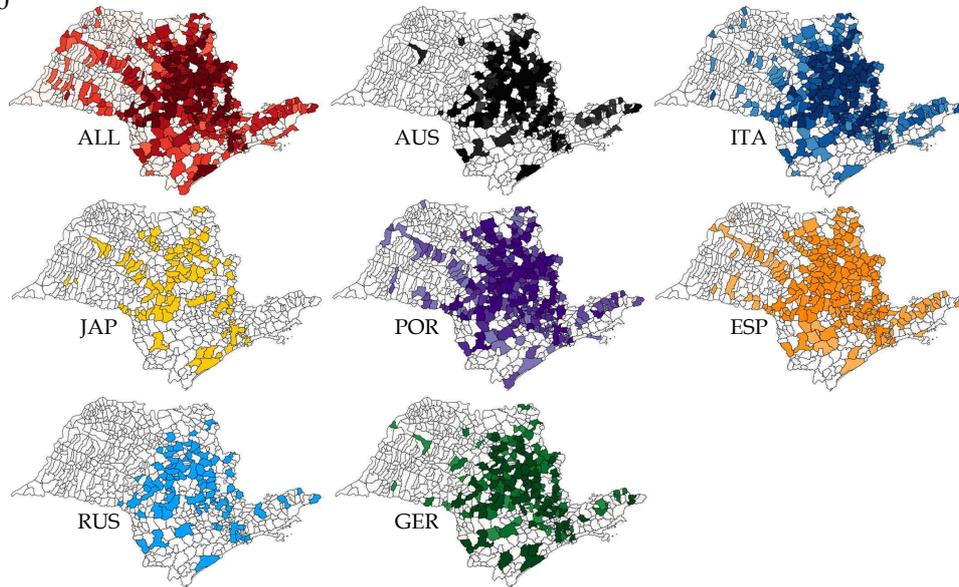


Figura A.55: Dispersão dos imigrantes por nacionalidade - São Paulo, fluxo acumulado até 1920

Fonte: Museu da Imigração de São Paulo. Elaboração do autor.

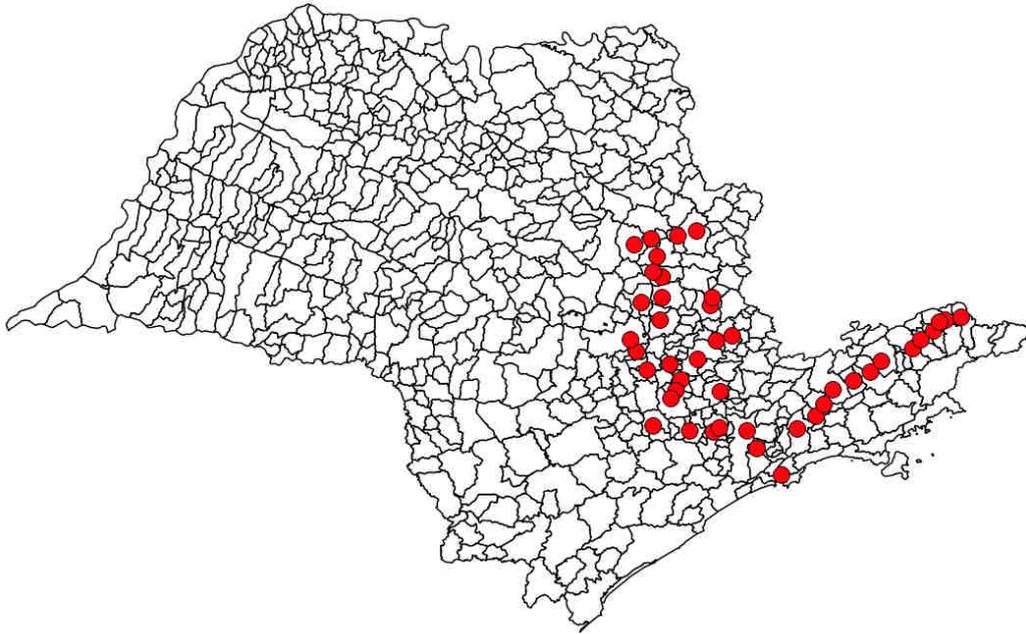


Figura A.56: Dispersão das estações ferroviárias - São Paulo, 1882
Fonte: www.estacoesferroviarias.com.br. Elaboração do autor.



Figura A.57: Dispersão das estações ferroviárias - São Paulo, 1890
Fonte: www.estacoesferroviarias.com.br. Elaboração do autor.



Figura A.58: Dispersão das estações ferroviárias - São Paulo, 1900
Fonte: www.estacoesferroviarias.com.br. Elaboração do autor.

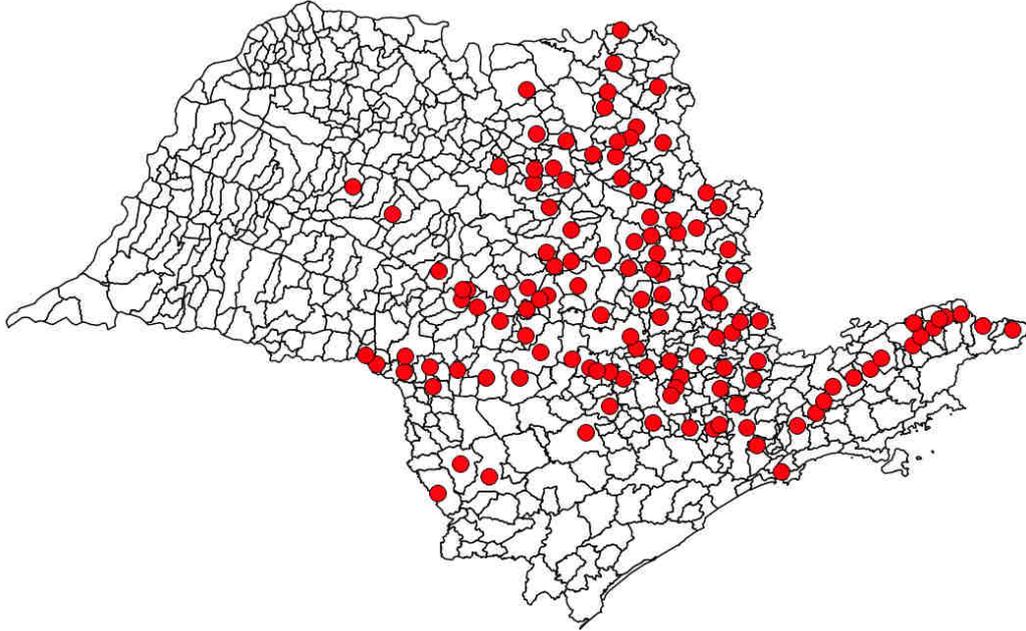


Figura A.59: Dispersão das estações ferroviárias - São Paulo, 1910
Fonte: www.estacoesferroviarias.com.br. Elaboração do autor.

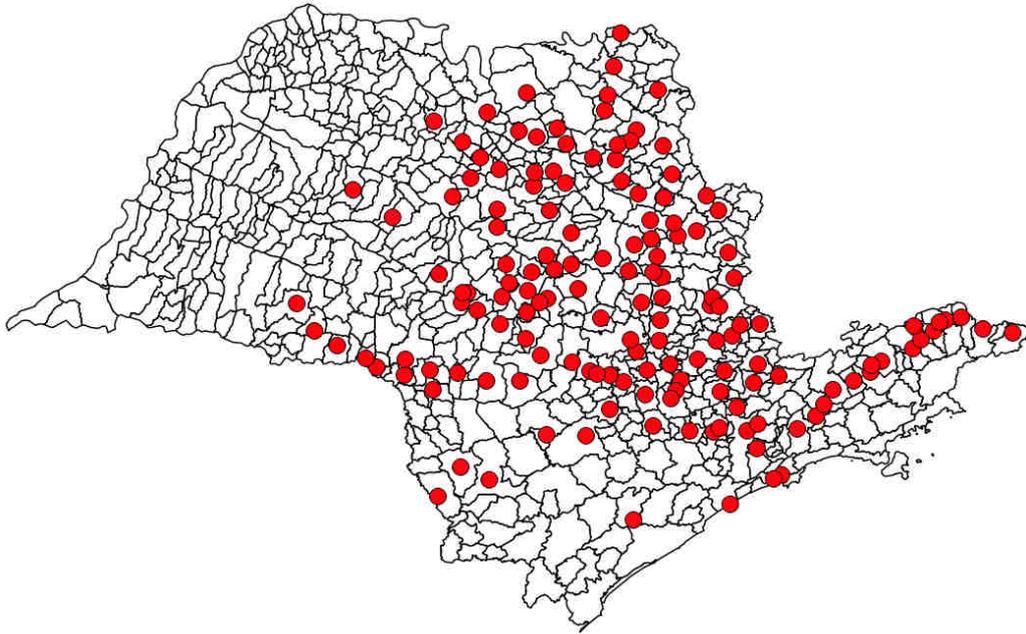


Figura A.60: Dispersão das estações ferroviárias - São Paulo, 1920
Fonte: www.estacoesferroviarias.com.br. Elaboração do autor.

Tabela A.1: Quantidades de imigrantes registrados na Hospedaria do Brás, em São Paulo, por nacionalidade declarada 1882-1920 (15 mais frequentes)

Nacionalidade	Quantidades
Italianos	715.491
Espanhóis	258.840
Portugueses	146.916
Austríacos	28.257
Alemães	16.421
Japoneses	11.133
Russos	6.384
Franceses	4.115
Poloneses	3.314
Belgas	1.633
Suecos	1.433
Húngaros	1.296
Suiços	1.037
Gregos	992
Canadenses	959

Tabela A.2: Quantidades de imigrantes registrados na Hospedaria do Brás, em São Paulo, por município de destino 1882-1920 (15 mais frequentes)

Município	Quantidades
São Paulo	83.636
Ribeirão Preto	45.258
Campinas	32.859
São Carlos	27.257
Jaú	26.735
São Simão	23.560
Araraquara	21.303
Bauru	15.803
São Manuel	15.593
Descalvado	14.445
S.J. Rio Pardo	13.830
Jaboticabal	13.200
Cravinhos	12.994
Botucatu	12.447
Rio Claro	12.231

Tabela A.3: Estações ferroviárias: município a que pertencia em 1920, ano de inauguração, linha da qual fazia parte, coordenadas geográficas

Município 1920	Ano Inaug	Linha férrea	Coord. Sul	Coord. Oeste
Santos	1867	2-SP Railway	-23.927306	-46.340668
Jundiaí	1867	2-SP Railway	-23.195083	-46.872389
São Bernardo	1867	2-SP Railway	-23.698139	-46.552917
São Paulo	1867	2-SP Railway	-23.534998	-46.635297
Campinas	1872	3-Paulista	-22.908500	-47.066806
Indaiatuba	1872	4-Sorocabana	-23.092028	-47.216778
Salto	1873	4-Sorocabana	-23.189389	-47.255639
Itú	1873	4-Sorocabana	-23.256556	-47.305722
Lavrinhas	1874	1-Central	-22.567833	-44.906000
Queluz	1874	1-Central	-22.539833	-44.769556
Amparo	1875	5-Mogiana	-22.702833	-46.766194
Araras	1875	3-Paulista	-22.366944	-47.379750
Cachoeira Pta.	1875	1-Central	-22.660306	-45.008750
Capivari	1875	4-Sorocabana	-23.004611	-47.506139
Cotia	1875	4-Sorocabana	-23.545556	-46.935250
Mogi das Cruzes	1875	1-Central	-23.521167	-46.196667
Mogi Mirim	1875	5-Mogiana	-22.435528	-46.953361
Monte Mor	1875	4-Sorocabana	-22.951139	-47.312722
Parnaíba	1875	4-Sorocabana	-23.512750	-46.875722
Pedreira	1875	5-Mogiana	-22.741917	-46.902361
São Roque	1875	4-Sorocabana	-23.535889	-47.140722
Sorocaba	1875	4-Sorocabana	-23.496250	-47.454556
Caçapava	1876	1-Central	-23.099333	-45.705028
Guararema	1876	1-Central	-23.413806	-46.031528
Jacareí	1876	1-Central	-23.304333	-45.966417
Limeira	1876	3-Paulista	-22.564000	-47.397111
Rio Claro	1876	3-Paulista	-22.409028	-47.558056
Rio das Pedras	1876	4-Sorocabana	-22.843750	-47.607528
Sao Jose Campos	1876	1-Central	-23.173833	-45.885694
Taubaté	1876	1-Central	-23.022361	-45.560306
Guaratinguetá	1877	1-Central	-22.813917	-45.191417
Leme	1877	3-Paulista	-22.184250	-47.382417
Lorena	1877	1-Central	-22.732778	-45.119611
Pindamonhangaba	1877	1-Central	-22.928028	-45.463056
Piracicaba	1877	4-Sorocabana	-22.735500	-47.655306
Casa Branca	1878	5-Mogiana	-21.779417	-47.078000
Cruzeiro	1878	1-Central	-22.577722	-44.957194
Mogi Guaçú	1878	5-Mogiana	-22.369583	-46.940611
Pirassununga	1878	3-Paulista	-22.004472	-47.426472
Sta. Cruz Conceição	1878	3-Paulista	-22.137778	-47.453833
Porto Ferreira	1880	3-Paulista	-21.851472	-47.477861
Descalvado	1881	3-Paulista	-21.901889	-47.624333

Município 1920	Ano Inaug	Linha férrea	Coord. Sul	Coord. Oeste
Santa Cruz Palmeiras	1881	3-Paulista	-21.825361	-47.243000
Itapira	1882	5-Mogiana	-22.444028	-46.880444
São Simão	1882	5-Mogiana	-21.455000	-47.583306
Tambau	1882	5-Mogiana	-21.708833	-47.275028
Cravinhos	1883	5-Mogiana	-21.343000	-47.735389
Ribeirão Preto	1883	5-Mogiana	-21.149417	-47.784056
Tietê	1883	4-Sorocabana	-23.109917	-47.717389
Analândia	1884	3-Paulista	-22.134280	-47.675703
Atibaia	1884	3-Paulista	-23.114583	-46.579917
Bragança Paulista	1884	3-Paulista	-22.952944	-46.547278
Santa Rita Passa Quatro	1884	3-Paulista	-21.684364	-47.487817
São Carlos	1884	3-Paulista	-22.022972	-47.895667
Araraquara	1885	3-Paulista	-21.793778	-48.171250
Brotas	1885	3-Paulista	-22.288154	-48.112007
Batatais	1886	5-Mogiana	-20.888124	-47.603115
Dois Córregos	1886	3-Paulista	-22.371194	-48.378828
Laranjal Paulista	1886	4-Sorocabana	-23.046117	-47.838787
Sao João Boa Vista	1886	5-Mogiana	-21.965236	-46.800623
Conchas	1887	4-Sorocabana	-23.011514	-48.010073
Franca	1887	5-Mogiana	-20.533541	-47.412864
Jaú	1887	3-Paulista	-22.304839	-48.556000
Mineiros do Tietê	1887	3-Paulista	-22.411410	-48.450902
Sao José do Rio Pardo	1887	5-Mogiana	-21.595319	-46.892679
Anhembi	1888	4-Sorocabana	-22.933964	-48.162500
Bananal	1888	1-Central	-22.680000	-44.319583
Franco da Rocha	1888	2-SP Railway	-23.329634	-46.726421
São Manuel	1888	4-Sorocabana	-22.728301	-48.567534
Botucatu	1889	4-Sorocabana	-22.873917	-48.444278
Espírito Santo Pinhal	1889	5-Mogiana	-22.195514	-46.751538
Tatuí	1889	4-Sorocabana	-23.350167	-47.834806
Itatiba	1890	3-Paulista	-23.015172	-46.837058
Mococa	1890	5-Mogiana	-21.473275	-46.990731
Brodowski	1892	5-Mogiana	-20.989160	-47.659421
Guariba	1892	3-Paulista	-22.642953	-44.577244
São José Barreiro	1892	1-Central	-22.642953	-44.577244
Serra Negra	1892	5-Mogiana	-22.609011	-46.702736
Jaboticabal	1893	3-Paulista	-21.252400	-48.325832
São Pedro	1893	4-Sorocabana	-22.548449	-47.915112
Ribeirão Bonito	1894	3-Paulista	-22.069308	-48.175384
Itapetininga	1895	4-Sorocabana	-23.583500	-48.047674
Avaré	1896	4-Sorocabana	-23.097600	-48.916450
Cerqueira César	1896	4-Sorocabana	-23.033950	-49.164246
Itatinga	1896	4-Sorocabana	-23.097683	-48.622077

Município 1920	Ano Inaug	Linha férrea	Coord. Sul	Coord. Oeste
Pereiras	1896	4-Sorocabana	-23.035293	-47.947131
Lençóis Paulista	1898	4-Sorocabana	-22.594991	-48.798359
Santa Rosa Viterbo	1898	5-Mogiana	-21.484499	-47.360067
Barra Bonita	1899	3-Paulista	-22.497735	-48.560011
Jardinópolis	1899	5-Mogiana	-21.018014	-47.770337
Matão	1899	8-EF Araraquara	-21.599035	-48.360802
Sertãozinho	1899	5-Mogiana	-21.133426	-47.987616
Dourado	1900	3-Paulista	-22.114910	-48.317388
Orlândia	1901	5-Mogiana	-20.720545	-47.881556
São Joaquim Barra	1901	5-Mogiana	-20.582555	-47.857864
Taquaritinga	1901	8-EF Araraquara	-21.383782	-48.499086
Bebedouro	1902	3-Paulista	-20.948111	-48.475608
Monte Alto	1902	3-Paulista	-21.259052	-48.497021
Agudos	1903	4-Sorocabana	-22.475289	-48.989367
Boa Esperança Sul	1903	3-Paulista	-21.997593	-48.390235
Ituverava	1903	5-Mogiana	-20.332911	-47.802283
Pederneiras	1903	3-Paulista	-22.358472	-48.774267
Pitangueiras	1903	3-Paulista	-21.010795	-48.217973
Bauru	1905	4-Sorocabana	-22.323278	-49.078389
Igarapava	1905	5-Mogiana	-20.036749	-47.744093
Piratininga	1905	3-Paulista	-22.408090	-49.133205
Avaí	1906	6-EF Noroeste	-22.156758	-49.327935
Piquete	1906	1-Central	-22.615514	-45.176873
Pirajú	1906	4-Sorocabana	-23.176889	-49.384611
Lins	1908	6-EF Noroeste	-21.655320	-49.733165
Ipaussú	1908	4-Sorocabana	-23.045637	-49.633304
Óleo	1908	4-Sorocabana	-23.004912	-49.404013
Penápolis	1908	6-EF Noroeste	-21.418220	-50.078934
Santa Cruz Rio Pardo	1908	4-Sorocabana	-22.910389	-49.621917
Altinópolis	1909	7-SP-Minas	-21.026834	-47.373109
Barretos	1909	3-Paulista	-20.564443	-48.563524
Itapeva	1909	4-Sorocabana	-23.966523	-48.889597
Itaberá	1909	4-Sorocabana	-23.857685	-49.143407
Itararé	1909	4-Sorocabana	-24.118583	-49.338278
Ourinhos	1909	4-Sorocabana	-22.976056	-49.870361
Pirajuí	1909	6-EF Noroeste	-22.319556	-49.120556
Salto Grande	1909	4-Sorocabana	-22.901944	-49.971250
Santa Adélia	1909	8-EF Araraquara	-21.235111	-48.802417
Socorro	1909	5-Mogiana	-22.597250	-46.525667
Catanduva	1910	8-EF Araraquara	-21.130611	-48.969861

Município 1920	Ano Inaug	Linha férrea	Coord. Sul	Coord. Oeste
Ibitinga	1910	3-Paulista	-21.746000	-48.824639
Itajobi	1910	3-Paulista	-21.313444	-49.055972
Monte Azul Pta.	1910	3-Paulista	-20.897169	-48.634881
Bocaina	1910	3-Paulista	-22.142520	-48.521873
Bariri	1911	3-Paulista	-22.076438	-48.739089
Rio Preto	1911	8-Araraquara	-20.809250	-49.376639
Tabapuã	1911	8-Araraquara	-20.996011	-49.123781
Itapuí	1912	3-Paulista	-22.234136	-48.718847
Cajuru	1912	5-Mogiana	-21.276108	-47.297910
Angatuba	1913	4-Sorocabana	-23.574096	-48.387155
Viradouro	1913	3-Paulista	-20.874500	-48.297333
Assis	1914	4-Sorocabana	-22.661879	-50.420567
Itanhaém	1914	2-SP Railway	-24.182146	-46.787575
Olímpia	1914	3-Paulista	-20.738166	-48.907806
Palmital	1914	4-Sorocabana	-22.787599	-50.221843
Piracaia	1914	3-Paulista	-23.052400	-46.361806
São Vicente	1914	2-SP Railway	-23.960785	-46.409523
Tremembé	1914	1-Central	-22.963716	-45.549171
Guarulhos	1915	1-Central	-23.482814	-46.546416
Iguape	1915	2-SP Railway	-24.322304	-47.627085
Itápolis	1915	4-Sorocabana	-21.593020	-48.822772
Conceição M. Alegre	1916	4-Sorocabana	-22.416348	-50.573273
Sta. Bárbara D'Oeste	1917	3-Paulista	-22.749881	-47.409973
Porto Feliz	1920	4-Sorocabana	-23.218443	-47.525961
Novo Horizonte	1939	3-Paulista	-21.476256	-49.209288