

HUMAN CAPITAL PERSISTENCE AND DEVELOPMENT

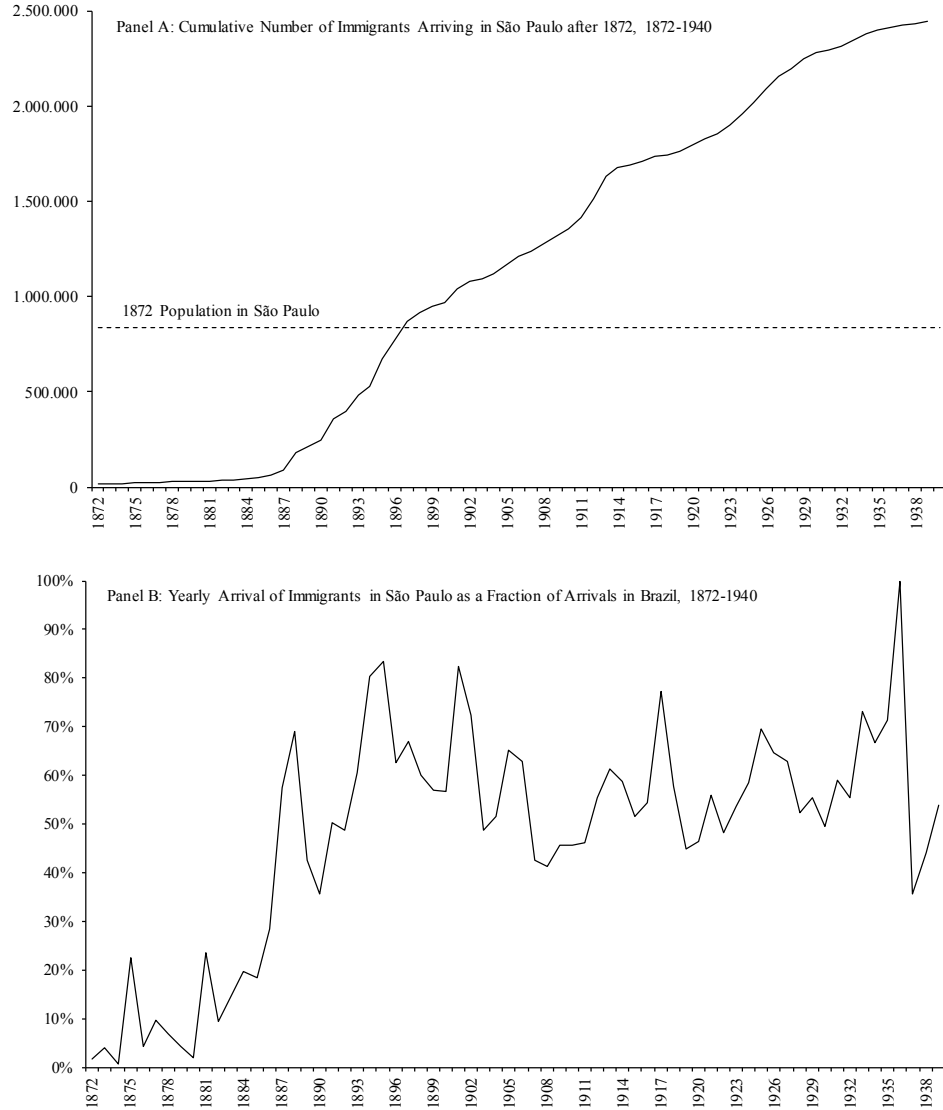
RUDI ROCHA, CLAUDIO FERRAZ AND RODRIGO R. SOARES

APPENDIX

DEFINITION OF VARIABLES AND ORIGINAL SOURCES OF INFORMATION

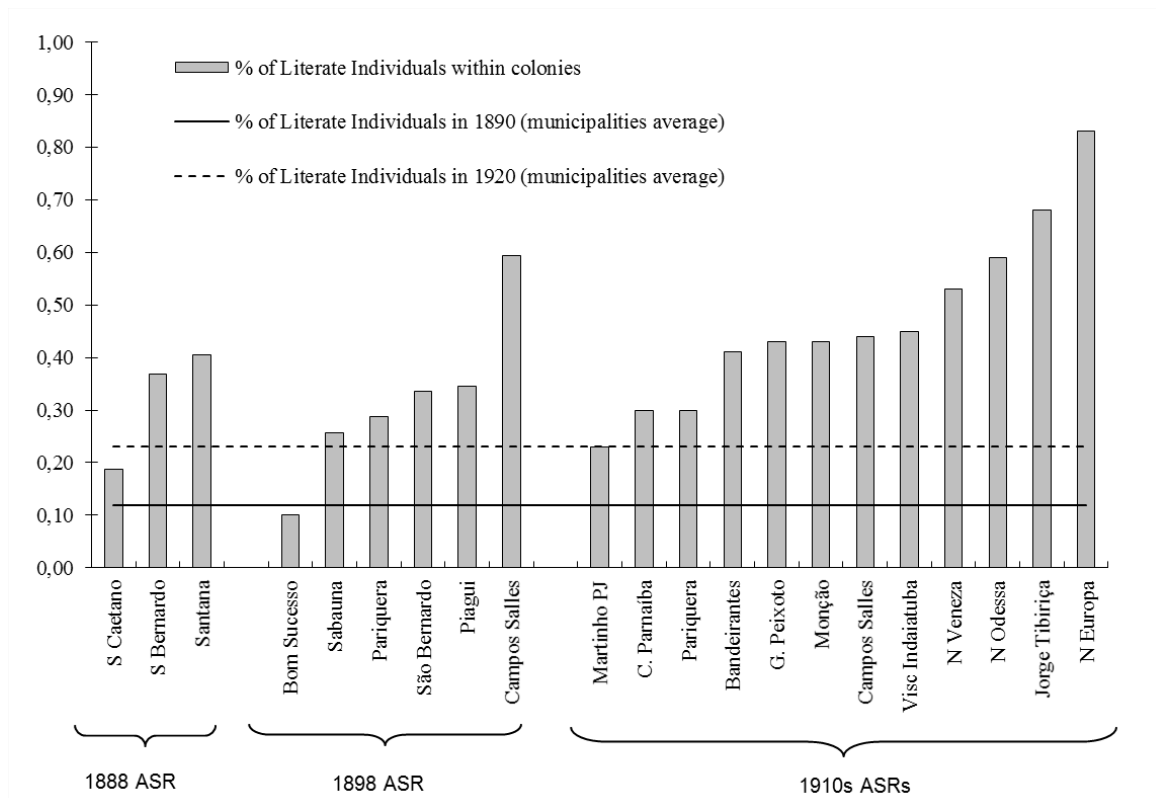
Variables	Description	Source of Information
Socioeconomic Conditions in 1872		
% Literate (aged 6+)	Number of literate aged 6+ over total population aged 6+	1872 Census
% Children attending school	Number of children aged 6-15 attending school over total population aged 6-15	1872 Census
Teachers/children (*1000)	Number of teachers over total number of children aged 6-15 *1000	1872 Census
Population density	Number of individuals by total area in hectares	1872 Census
% Foreigners	Number of foreigners over total population	1872 Census
% Slaves	Number of slaves over total population	1872 Census
Public administration	Total number of workers in public administration over total population *1000	1872 Census
Legal professionals	Total number of workers in legal professions over total population *1000	1872 Census
% Emp agriculture	Share of workers in agriculture over total number of occupied workers	1872 Census
% Emp manufacturing	Share of workers in manufacturing over total number of occupied workers	1872 Census
% Emp services and retail	Share of workers in services and retail over total number of occupied workers	1872 Census
Socioeconomic Conditions in 1920		
% Literate (aged 6+)	Number of literate aged 6+ over total population aged 6+	1920 Census
Schools/children (*1000)	Number of schools over total number of children aged 6-15 *1000	1920 Census
Teachers/children (*1000)	Number of teachers over total number of children aged 6-15 *1000	1920 Census
% Foreigners (and by nationality)	Number of foreigners over total population (and by nationality)	1920 Census
% Literate foreigners	Number of literate foreigners over total foreign population	1920 Census
% Small farms	Number of farms up to 100 hectares over total number of farms	1920 Census
Coffee productivity	Total production of coffee (in tonnes) over total size of rural establishments (in hectares)	1920 Census
Value of farmland	Land price in 1000 Réis/hectare	1920 Census
Wages in construction and agriculture	In Réis	1920 Census
% Emp agriculture	Share of workers in agriculture over total number of occupied workers	1920 Census
% Emp manufacturing	Share of workers in manufacturing over total number of occupied workers	1920 Census
% Emp services and retail	Share of workers in services and retail over total number of occupied workers	1920 Census
Socioeconomic Conditions in 1940		
% Literate (aged 5+)	Number of literate aged 5+ over total population aged 5+	1940 Census
% Children attending school	Number of children aged 7-14 attending school over total population aged 7-14	1940 Census
Schools/children (*1000)	Number of schools over total number of children aged 7-14 *1000	1940 Census
% Emp agriculture	Share of workers in agriculture over total number of occupied workers	1940 Census
% Emp manufacturing	Share of workers in manufacturing over total number of occupied workers	1940 Census
% Emp services and retail	Share of workers in services and retail over total number of occupied workers	1940 Census
Socioeconomic Conditions in 2000		
% Literate (aged 5+)	Number of literate aged 5+ over total population aged 5+	2000 Census
% Children attending school	Number of children aged 7-14 attending school over total population aged 7-14	2000 Census
Schools/children (*1000)	Number of schools over total number of children aged 7-14 *1000	2000 School Census/INEP
Teachers/children (*1000)	Number of teachers over total number of children aged 7-14 *1000	2000 Census
Years of schooling (aged 5+ and by cohorts)	Average number of years of schooling for those aged 5+ (or by cohorts)	2000 Census
Income per capita	Total income over total population	2000 Census
% Emp agriculture	Share of workers in agriculture over total number of occupied workers	2000 Census
% Emp manufacturing	Share of workers in manufacturing over total number of occupied workers	2000 Census
% Emp services and retail	Share of workers in services and retail over total number of occupied workers	2000 Census
Geography and Transportation		
Railway	Dummy indicating presence of railway station in the municipality (for each year)	www.estacoesferroviarias.com.br
Distance to the capital in km (city of São Paulo)	Avg distances of the municip. of the 2000 adm. div. that constitute the original municip in 1872 (1920)	Ipeadata
Latitude/Longitude degrees	Idem, but averaging centroids	Ipeadata
Dummy for latosol	Indicates that the predominant type of soil is latosol	Embrapa
Dummy for argisol	Indicates that the predominant type of soil is argisol	Embrapa
Dummy for cambisol	Indicates that the predominant type of soil is cambisol	Embrapa
Dummy for spondosol	Indicates that the predominant type of soil is spondosol	Embrapa

FIGURE A1: ARRIVAL OF IMMIGRANTS IN THE STATE OF SÃO PAULO, BRAZIL, 1872-1940



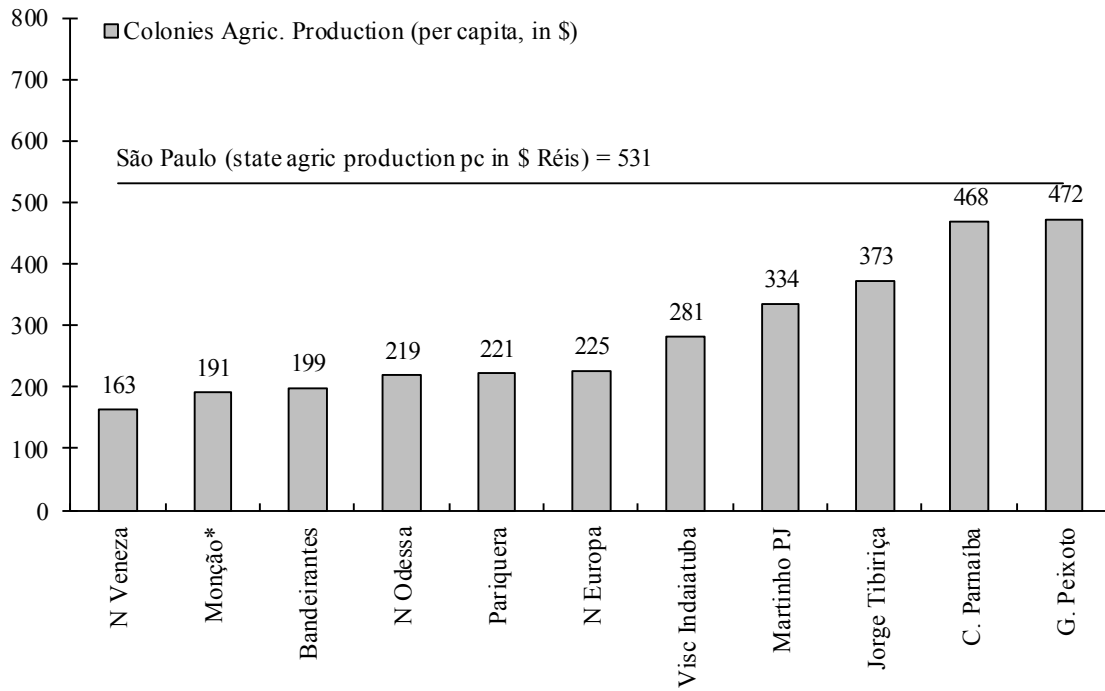
Source: Memorial do Imigrante and Censuses of 1872, 1890, 1900, 1910, 1920 and 1940.

FIGURE A2: LITERACY RATES IN SELECTED STATE SPONSORED SETTLEMENTS, STATE OF SÃO PAULO, BRAZIL, LATE 19TH CENTURY AND EARLY 20TH CENTURY



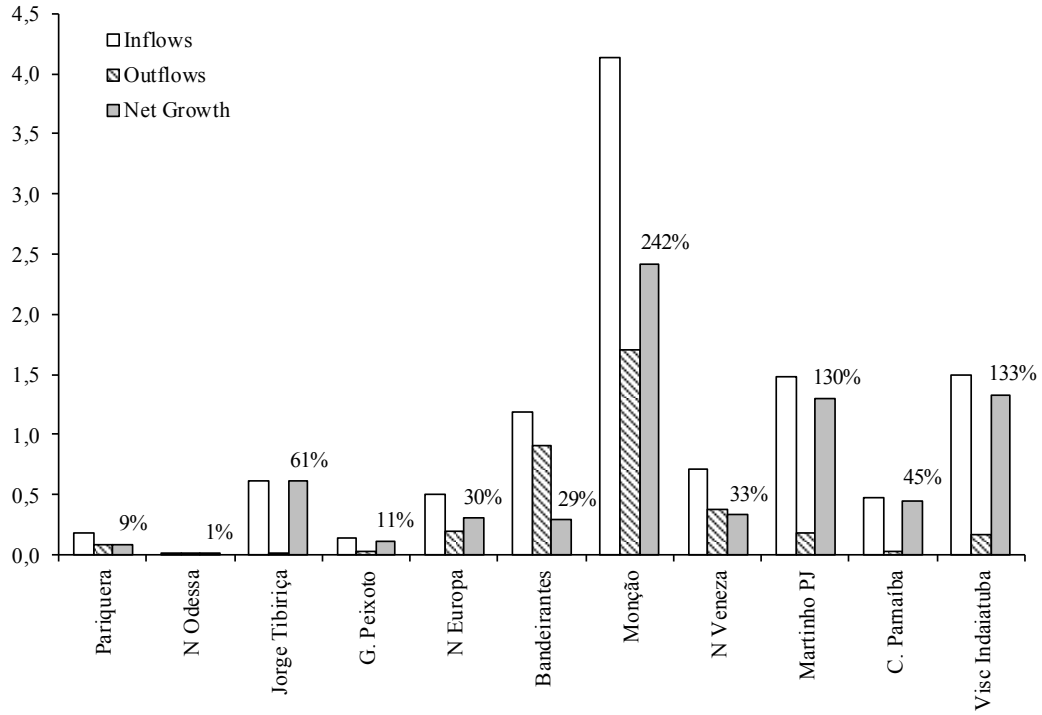
Source: Authors' calculations based on Annual Statistical Reports of São Paulo and Report of Secretariat of Agriculture, various years

FIGURE A3: PRODUCTIVITY IN SELECTED STATE SPONSORED SETTLEMENTS, STATE OF SÃO PAULO, BRAZIL, EARLY 20TH CENTURY



Source: 1920 Census and 1910s ASRs.

FIGURE A4: INFLOW AND OUTFLOW OF IMMIGRANTS AND NET GROWTH IN SELECTED STATE SPONSORED SETTLEMENTS, STATE OF SÃO PAULO, BRAZIL, 1912-1918 (AS FRACTIONS OF 1912 POPULATION)



Source: 1910s ASRs.

TABLE A1: BASIC INFORMATION ON THE STATE SPONSORED SETTLEMENTS CONSIDERED IN THE ANALYSIS, SÃO PAULO, BRAZIL

Settlements	Year of Foundation	Year of Emancipation	Original Municipality	Current Municipality	Initial Area (hectares)	Predominant Nationality	Sources of Information
Glória	1877	N/A	São Paulo	São Paulo	N/A	N/A	Pires (2006)
Santanna	1877	N/A	São Paulo	São Paulo	84	N/A	1877 PPR
São Bernardo	1877	1901	São Bernardo	São Bernardo	1959	Italians, Germans Polish	1877 PPR
São Caetano	1877	N/A	São Bernardo	São Caetano	1909	Italians	Martins (1973)
Canas	1885	1893	Lorena	Canas	1110	Italians, Brazilians	1892 ARSA
Cascalho	1885	1893	Limeira	Cordeirópolis	N/A	Italians	1892 ARSA
Pariquera-açu	1887	1901	Iguape	Pariquera-açu	16603	Germans Brazilians	1898 ARSA
Sen. Antônio Prado	1887	1893	Ribeirão Preto	Ribeirão Preto	1500 aprox	Italians	Silva (2004)
Ribeirão Pires	1887	1893	São Bernardo	Ribeirão Pires	N/A	N/A	Freitas (2008)
Rodrigo Silva	1887	1893	Porto Feliz	Porto Feliz	1601	Belgians	Souza (1978)
Boa Vista	1888	1893	Jacareí	Jacareí	N/A	Italians	
Barão de Jundiá	1887	1893	Jundiá	Jundiá	514 aprox	Italians	Pereira and Fillipini (1988)
Sabaúna	1889	1901	Mogi das Cruzes	Mogi das Cruzes	N/A	Spaniards, Brazilians	1898 ARSA
Quiririm	1890	1893	Taubaté	Taubaté	N/A	Italians	Lorenzo (2002)
Piaguí	1892	1901	Guaratinguetá	Guaratinguetá	1264	N/A	1892 ARSA
Bom Sucesso	1894	1899	Sorocaba	Sorocaba	N/A	Brazilians	1898 ARSA
Campos Sales	1897	1903	Campinas	Cosmópolis	3200	Swiss, Germans	Baldini (2008)
Jorge Tibiriçá	1905	1923-24	Rio Claro	Corumbataí	4356	Spaniards, Italians, Brazilians	1900-1910s ASRAs and ASRs
Nova Odessa	1905	1920-21	Campinas	Nova Odessa	4310	Russians	1900-1910s ASRAs and ASRs
Gavião Peixoto	1907	1923-24	Araraquara	Gavião Peixoto	4840	N/A	1900-1910s ASRAs and ASRs
Nova Paulicéia	1907	1920-21	Araraquara	Gavião Peixoto	4840	N/A	1900-1910s ASRAs and ASRs
Nova Europa	1907	1920-21	Ibitinga/Araraquara	Nova Europa	4840	Germans, Brazilians	1900-1910s ASRAs and ASRs
Bandeirantes	1908	N/A	S. José do Barreiro	S. J. do Barreiro	14104	Germans, Brazilians	1900-1910s ASRAs and ASRs
Monção	1910	N/A	Sta Barbara do Rio Pardo	Iaras	28873	Various	1900-1910s ASRAs and ASRs
Nova Veneza	1910	1919	Campinas	Sumaré	3395	N/A	1900-1910s ASRAs and ASRs
Conde de Parnaíba	1911	1919	Mogi-Mirim	Conchal	4622	N/A	1900-1910s ASRAs and ASRs
Martinho Prado Jr	1911	1920-21	Mogi-Guaçu	Mogi-Mirim	5729	Spaniards, Italians, Brazilians	1900-1910s ASRAs and ASRs
Visc. de Indaiatuba	1911	1923-24	Mogi-Mirim	Mogi-Mirim	2853	N/A	1900-1910s ASRAs and ASRs

Note: N/A (Non-available) indicates that we could not find a source containing the respective information of the settlement.

TABLE A2: BASELINE DIFFERENCES IN CHARACTERISTICS: GEOGRAPHIC AND SOCIOECONOMIC DETERMINANTS OF SETTLEMENT LOCATIONS

Dependent variable: Settlement after 1872	Geography (1)	Demographics (2)	Occupations (3)	All Characteristics (4)
Log Distance to capital (in km)	-0.045 [0.143]			0.029 [0.145]
Latitude	0.032 [0.082]			-0.003 [0.095]
Longitude	-0.021 [0.049]			-0.015 [0.061]
Altitude (in 100m)	-0.033 [0.017]*			-0.024 [0.023]
Dummy for latosol	0.176 [0.249]			0.172 [0.302]
% Literacy rate		0.015 [0.452]		-0.185 [0.582]
% Children attending school		-0.154 [0.320]		-0.043 [0.619]
% Foreigners		4.904 [3.407]		4.043 [4.074]
% Slaves		0.277 [0.499]		0.017 [0.682]
Population density		-0.005 [0.008]		0.001 [0.010]
Railway		0.376 [0.259]		0.507 [0.274]*
Public administration			0.013 [0.046]	0.021 [0.059]
Legal professionals			0.076 [0.059]	0.073 [0.081]
Teachers			-0.036 [0.081]	-0.049 [0.095]
Agriculture			-10.626 [18.054]	-3.669 [18.006]
Manufacturing			-9.793 [18.138]	-2.993 [18.172]
Services and Retail			-11.001 [18.217]	-3.903 [18.198]
F-test (p-value)	0.94 (0.49)	1.44 (0.21)	.99 (0.46)	1.30 (0.20)
Observations	88	88	88	88
R-squared	0.038	0.09	0.055	0.166

Notes: The sample of municipalities that would receive settlements after 1872 contains 19 observations (column Settlements); there are 69 observations in the sample of municipalities that would not receive any settlements (column No Settlements). All variables computed according to the 1872 census boundaries. Data for geographic variables originally from Ipeadata (distance to the capital, latitude, longitude and elevation) and Embrapa Solos (indicator for presence of Latosol). Indicator for railway built on data originally from www.estacoesferroviarias.com.br. The remainder variables were built on data originally available from the 1872 Census. Share of foreigners and slaves computed over total population. Share of literate over population aged 6+. Share of children attending school over total number of children aged 6-15. Population density is total population per hectare. Public administration and Legal profession refer to total number of workers in the given occupation relative to total population *1000. Share of workers in agriculture, manufacturing, services and retail computed over total number of occupied workers.

TABLE A3: THE SHORT-TERM OF SETTLEMENTS USING A CONTINUOUS MEASURE

Dependent Variable:	Literacy rate (%) 1920				Other economic and demographic characteristics 1920						
	No controls	Control for Geography	Control for Charact. 1872	% Foreigners	% Literate foreigners	Population density	% Small farms	Coffee product.	Log land prices	Log wages construc.	Log wages agric.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Share of foreign pop in settlement	0.748 [0.310]**	0.676 [0.320]**	0.484 [0.260]*	0.361 [0.124]***	0.221 [0.279]	12.350 [57.846]	0.653 [0.366]*	-0.009 [0.122]	1.622 [1.811]	-0.057 [0.789]	0.505 [0.622]
Geography controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Characteristics 1872	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	199	199	199	199	199	199	199	199	146	123	169
Adjusted R-squared	0.031	0.254	0.33	0.589	0.166	0.184	0.2	0.46	0.363	0.558	0.256

Notes: This table reports the effects of Settlements on education, foreigners, and other economic characteristics in 1920. All columns report the results from OLS regressions where the dependent variable is: the municipality literacy rate in 1920 (in %) for columns 1-3; the variable listed in the top of the column from columns (4)-(11). The variable share of population in settlement is the share of foreigners in the initial settlement relative to the existing population in the municipality in 1872. Geographic controls include: distance to the capital, latitude, longitude, elevation, and indicators for different types of soil - latosol, argisol, cambisol, and spondosol. The characteristics in 1872 include presence of railway, share of foreigners, share of slaves, share of literate population, share of children attending school, population density, total number of workers in public administration and legal professions relative to total population, share of workers in agriculture, manufacturing, services and retail computed over total number of occupied workers. All regressions estimated for 202 municipalities based on the 1920 census boundaries. Robust standard errors clustered at the 1872 municipality boundaries are displayed in brackets significantly different than zero at 99 (***) , 95 (**), 90 (*) percent confidence.

TABLE A4: THE EFFECTS OF SETTLEMENTS ON EMPLOYMENT AND STRUCTURAL TRANSFORMATION FOR INDIVIDUAL SECTORS

	Sectors in 1920		Sectors in 2000	
	(1)	(2)	(1)	(2)
Manufacturing:				
Construction	0.015	(0.005)***	-0.000	(0.005)
Clothing	0.039	(0.015)**	-0.000	(0.012)
Wood and furniture	0.003	(0.002)	0.002	(0.005)
Metal	0.007	(0.003)**	0.004	(0.003)
Chemical	0.001	(0.000)**	0.004	(0.002)*
Food	0.002	(0.001)**	-0.007	(0.003)**
Machines and Equip	-	-	0.003	(0.003)
Automobile	-	-	0.006	(0.005)
Plastics	-	-	-0.005	(0.005)
Retail and Services:				
Transportation	-	-	0.003	(0.003)
Banks and finance	-	-	0.014	(0.005)**
Education	-	-	0.007	(0.002)***
Health	-	-	0.008	(0.003)**
Personal services	-	-	0.002	(0.001)**
Culture	-	-	0.001	(0.001)
Political activism	-	-	-0.000	(0.000)
Automotive retail and repairing	-	-	0.006	(0.002)***
Agricultural goods	-	-	0.000	(0.001)
Food retail and supermarkets	-	-	0.003	(0.002)
Textile and clothing	-	-	0.002	(0.001)
Dept stores	-	-	0.019	(0.006)***
General repairing	-	-	0.001	(0.000)
Personal goods and hygiene retail	-	-	0.003	(0.001)***
Housing	-	-	0.002	(0.001)
Machines and equipments	-	-	0.001	(0.001)**

Notes: Robust standard errors in parentheses, clustered at the 1872 census boundaries: *** p<0.01, ** p<0.05, * p<0.1. Each coefficient reports the result of a distinct regression where the dependent variable is regressed on the variable settlement, a dummy that equals one if the municipality received at least one state sponsored settlement before 1920. Dependent variables are computed as the total number of workers in the given sector over total number of occupied individuals. Column 1 (2) reports the results for dependent variables in 1920 (2000). All specifications include geographic controls (distance to the capital, latitude, longitude, elevation, and indicators for different types of soil - latosol, argisol, cambisol, and spondosol) and controls for the presence of railway and other characteristics in 1872 (share of foreigners, share of slaves, share of literate population, share of children attending school, population density, total number of workers in public administration and legal professions relative to total population, share of workers in agriculture, manufacturing, services and retail computed over total number of occupied workers). All variables computed according to the 1920 census boundaries.

TABLE A5: PERSISTENT EFFECTS OF SETTLEMENTS ON ENTREPRENEURSHIP, BY COHORTS OF INDIVIDUALS

	All Cohorts	1920-29	1930-39	1940-49	1950-59	1960-69
	(1)	(2)	(3)	(4)	(5)	(6)
<u>Panel A. % employers for individuals born in municipality</u>						
Settlement	-0.000 (0.003)	0.023 (0.031)	0.010 (0.015)	0.008 (0.008)	-0.005 (0.005)	0.002 (0.004)
Mean Dependent Variable	0.0349	0.0946	0.0651	0.0554	0.0504	0.0396
<u>Panel B. % entrepreneurs for individuals born in municipality</u>						
Settlement	0.001 (0.009)	0.052 (0.064)	0.031 (0.029)	0.041 (0.016)**	0.014 (0.014)	-0.002 (0.011)
Mean Dependent Variable	0.222	0.464	0.398	0.332	0.272	0.225
Geographic controls	Yes	Yes	Yes	Yes	Yes	Yes
Characteristics 1872	Yes	Yes	Yes	Yes	Yes	Yes
Observations	202	194	202	202	202	202

Notes: Robust standard errors in parentheses, clustered at the 1872 census boundaries: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In all columns the variable settlement is a dummy that equals one if the municipality received at least one state sponsored settlement before 1920. Dependent variables are average share of entrepreneurs (employers in Panel A and self-employed in Panel B) for each cohort of individuals born in the municipality, as defined at the top of each column, and built on microdata originally from the 2000 census. All specifications include geographic controls (distance to the capital, latitude, longitude, elevation, and indicators for different types of soil - latosol, argisol, cambisol, and spondosol) and controls for the presence of railway and other characteristics in 1872 (share of foreigners, share of slaves, share of literate population, share of children attending school, population density, total number of workers in public administration and legal professions relative to total population, share of workers in agriculture, manufacturing, services and retail computed over total number of occupied workers). All variables computed according to the 1920 census boundaries.