

# Immigrant Diversity and Long-Run Development: Evidence from the Age of Mass Migration in Brazil

Luigi Minale (Universidad Carlos III Madrid, CReAM and IZA)

Rudi Rocha (FGV)

Bruno Vigna (IE-UFRJ)

EEA Congress  
30 August, 2023  
Barcelona

# Motivation

- ▶ Immigration and the continuous rise in social heterogeneity are still some of the most important challenges facing modern societies.

# Motivation

- ▶ Immigration and the continuous rise in social heterogeneity are still some of the most important challenges facing modern societies.
- ▶ Putnam (1997 SPS): *the most certain prediction that we can make about almost any modern society is that it will be more diverse a generation from now than it is today.*

# Motivation

- ▶ Immigration and the continuous rise in social heterogeneity are still some of the most important challenges facing modern societies.
- ▶ Putnam (1997 SPS): *the most certain prediction that we can make about almost any modern society is that it will be more diverse a generation from now than it is today.*
- ▶ Whether diversity is instrumental or detrimental to development is a question relevant for understanding the development process and to inform discussion about migration policies today.

# Motivation

- ▶ **Theory** provides mixed answers:

# Motivation

- ▶ **Theory** provides mixed answers:
  - ▶ Diversity can enhance productivity through innovation, skill complementarities, increased creativity, trade, and product variety.

# Motivation

- ▶ **Theory** provides mixed answers:
  - ▶ Diversity can enhance productivity through innovation, skill complementarities, increased creativity, trade, and product variety.
  - ▶ Diversity can trigger mutual distrust and lack of social cohesion, thus generating an inefficient provision of public goods, ethnically biased policies, and conflict.

# Motivation

- ▶ **Theory** provides mixed answers:
  - ▶ Diversity can enhance productivity through innovation, skill complementarities, increased creativity, trade, and product variety.
  - ▶ Diversity can trigger mutual distrust and lack of social cohesion, thus generating an inefficient provision of public goods, ethnically biased policies, and conflict.
  - ▶ Some mechanisms might be more relevant in the short and medium-term, others in the long-term.



# Motivation

- ▶ **Theory** provides mixed answers:
  - ▶ Diversity can enhance productivity through innovation, skill complementarities, increased creativity, trade, and product variety.
  - ▶ Diversity can trigger mutual distrust and lack of social cohesion, thus generating an inefficient provision of public goods, ethnically biased policies, and conflict.
  - ▶ Some mechanisms might be more relevant in the short and medium-term, others in the long-term.
- ▶ **Empirical evidence** on birthplace diversity is either based on cross-country studies or it focuses on short and medium-run effects.

# This Paper (I)

- ▶ This article investigates the **long-run** economic effects of the **immigrant population's composition** (birthplace diversity).

# This Paper (I)

- ▶ This article investigates the **long-run** economic effects of the **immigrant population's composition** (birthplace diversity).
- ▶ Focus on the large immigration wave experienced by **Brazil** between 1880 and 1920. Immigrant inflow into the State of São Paulo during the period was twice the size of initial population.

# This Paper (I)

- ▶ This article investigates the **long-run** economic effects of the **immigrant population's composition** (birthplace diversity).
- ▶ Focus on the large immigration wave experienced by **Brazil** between 1880 and 1920. Immigrant inflow into the State of São Paulo during the period was twice the size of initial population.
- ▶ Do municipalities that received a more **diverse mix of immigrant population** ended up having better long-term economic outcomes and why so?

## This Paper (II)

- ▶ **Identification Strategy.** Develop an instrumental variable combining over-time variation in the origin-country composition of immigrants arriving from overseas with the timing of the railway network expansion.

## This Paper (II)

- ▶ **Identification Strategy.** Develop an instrumental variable combining over-time variation in the origin-country composition of immigrants arriving from overseas with the timing of the railway network expansion.
- ▶ **Data.** Use unique historical data (passenger arrival records) about 1.2 million immigrants arriving in São Paulo between 1880 and 1920.

## This Paper (II)

- ▶ **Identification Strategy.** Develop an instrumental variable combining over-time variation in the origin-country composition of immigrants arriving from overseas with the timing of the railway network expansion.
- ▶ **Data.** Use unique historical data (passenger arrival records) about 1.2 million immigrants arriving in São Paulo between 1880 and 1920.
- ▶ **Preview of results.** A 1 SD increase in accumulated immigrant diversity in 1920 is associated with a 7-8% percent higher income per capita in 2000.

## This Paper (II)

- ▶ **Identification Strategy.** Develop an instrumental variable combining over-time variation in the origin-country composition of immigrants arriving from overseas with the timing of the railway network expansion.
- ▶ **Data.** Use unique historical data (passenger arrival records) about 1.2 million immigrants arriving in São Paulo between 1880 and 1920.
- ▶ **Preview of results.** A 1 SD increase in accumulated immigrant diversity in 1920 is associated with a 7-8% percent higher income per capita in 2000.
- ▶ **Mechanisms.** Municipalities with a more diverse stock of migrants experienced (i) faster structural transformation; (ii) higher municipal spending on education.



# Literature and Contribution

- ▶ Empirical literature on the **effect of cultural and birthplace diversity**:
  - ▶ Mostly based on cross-country studies or rather descriptive in nature: Alesina and La Ferrara (JEL 2005); Alesina Harnoss Rapoport (JEG 2016); Bove and Elia (WD 2017); Bahar Rapoport Turati (Research Policy 2020); Docquier et al. (JoEG 2020).; Montalvo and Reynal-Querol (Restat 2022).
  - ▶ Focuses on short and medium-run effects.: Ottaviano and Peri (JoEG 2006); Sparber (Regional Studies 2010).

# Literature and Contribution

- ▶ Empirical literature on the **effect of cultural and birthplace diversity**:
  - ▶ Mostly based on cross-country studies or rather descriptive in nature: Alesina and La Ferrara (JEL 2005); Alesina Harnoss Rapoport (JEG 2016); Bove and Elia (WD 2017); Bahar Rapoport Turati (Research Policy 2020); Docquier et al. (JoEG 2020).; Montalvo and Reynal-Querol (Restat 2022).
  - ▶ Focuses on short and medium-run effects.: Ottaviano and Peri (JoEG 2006); Sparber (Regional Studies 2010).
- ▶ Literature on the **long-run effects of migration**.
  - ▶ Usually the focus is the mass migration episode in the US and the size of immigrant shock rather than its composition: Sequeira Nunn Qian (Restud 2019); Tabellini (Restud 2020); Giuliano and Tabellini (unpublished 2021).
  - ▶ Exceptions: Ager and Bruckner (EER 2013); Burchardi Chaney Hassan (Restud 2019); Fulford Pwtkov Schiantarelli (JEG 2020); Droller (EJ 2017).

# Literature and Contribution

- ▶ Empirical literature on the **effect of cultural and birthplace diversity**:
  - ▶ Mostly based on cross-country studies or rather descriptive in nature: Alesina and La Ferrara (JEL 2005); Alesina Harnoss Rapoport (JEG 2016); Bove and Elia (WD 2017); Bahar Rapoport Turati (Research Policy 2020); Docquier et al. (JoEG 2020).; Montalvo and Reynal-Querol (Restat 2022).
  - ▶ Focuses on short and medium-run effects.: Ottaviano and Peri (JoEG 2006); Sparber (Regional Studies 2010).
- ▶ Literature on the **long-run effects of migration**.
  - ▶ Usually the focus is the mass migration episode in the US and the size of immigrant shock rather than its composition: Sequeira Nunn Qian (Restud 2019); Tabellini (Restud 2020); Giuliano and Tabellini (unpublished 2021).
  - ▶ Exceptions: Ager and Bruckner (EER 2013); Burchardi Chaney Hassan (Restud 2019); Fulford Pwtkov Schiantarelli (JEG 2020); Droller (EJ 2017).
- ▶ Literature on **nation building**:
  - ▶ Bandiera et al. (EJ 2019); Murard and Sakalli (unpublished 2019).

## Brief Historical Context

- ▶ Brazil has received about 4.1 million immigrants between 1872 and 1920, of which nearly 3 million came from Italy, Portugal, Spain, Germany and Japan, and the rest from around 70 other countries.
- ▶ This wave of immigration drastically changed the makeup of Brazilian population: in 1872 the total population was about 10 million with less than 400 thousand foreigners or naturalized.

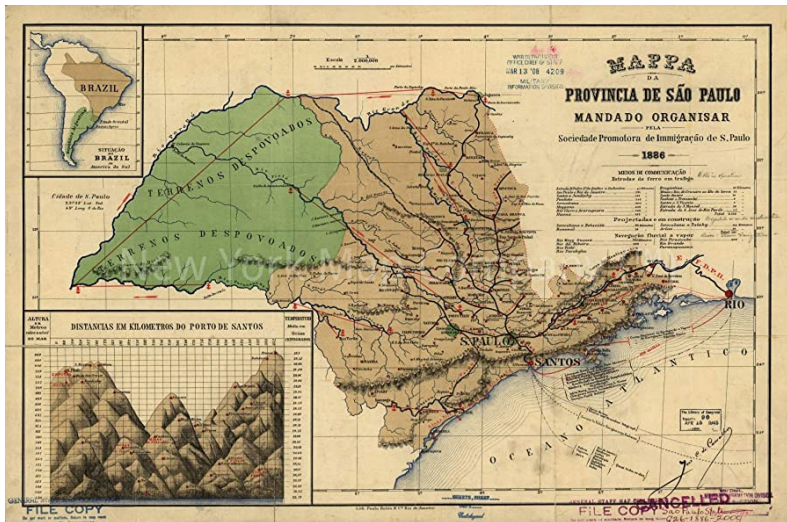
## Brief Historical Context

- ▶ Pull factor: **Expansion of coffee production** driven by growing international demand
- ▶ Pull factor: **Shortage of labor supply** also because of slave trade abolished (Law of the Free Womb 1831, enacted 1871, final slave abolished in 1881).
- ▶ Push factor: European population grew from 144 to 486 million between 1800-1920. Hunger and poverty in rural areas.
- ▶ Push factor: Decrease in cost of the trip (steam ships)

# Focus: São Paulo State



# Focus: São Paulo State



## Data: Immigrant's *Hospedaria* in São Paulo

- ▶ An hostel in the city of São Paulo, which served as a gateway, a inspection station, and marshaling place for the immigrants for their first few days or weeks in Brazil.



## Data: Immigrant's *Hospedaria* in São Paulo

- ▶ An hostel in the city of São Paulo, which served as a gateway, a inspection station, and marshaling place for the immigrants for their first few days or weeks in Brazil.
- ▶ Digitalised logbooks from the Immigrant's *Hospedaria* in São Paulo covering around **1.2 million arrivals** between 1882 and 1920.

## Data: Immigrant's *Hospedaria* in São Paulo

- ▶ An hostel in the city of São Paulo, which served as a gateway, a inspection station, and marshaling place for the immigrants for their first few days or weeks in Brazil.
- ▶ Digitalised logbooks from the Immigrant's *Hospedaria* in São Paulo covering around **1.2 million arrivals** between 1882 and 1920.
- ▶ **Individual level information**: first and last names, gender, civil status, intrafamily kinship, nationality, profession, religion, last country of residence, last country of residence, information about the financing of their travel (If subsidized migrants or not), and **destination in Brazil (municipality, train station or farm)**.

## Data: Immigrant's *Hospedaria* in São Paulo

- ▶ An hostel in the city of São Paulo, which served as a gateway, a inspection station, and marshaling place for the immigrants for their first few days or weeks in Brazil.
- ▶ Digitalised logbooks from the Immigrant's *Hospedaria* in São Paulo covering around **1.2 million arrivals** between 1882 and 1920.
- ▶ **Individual level information**: first and last names, gender, civil status, intrafamily kinship, nationality, profession, religion, last country of residence, last country of residence, information about the financing of their travel (If subsidized migrants or not), and **destination in Brazil (municipality, train station or farm)**.
- ▶ Use the data to construct a **database with the annual numbers of immigrants by origin country and destination municipality**.

## Data: other data

- ▶ Census data for 1872, 1920, 1940 and 2000 at the municipal level. 1872 used for baseline characteristics; 1920, 1940 and 2000 for outcome variables.
- ▶ Geographic characteristics of municipalities: Ipeadata.
- ▶ Information about the expansion of railroad infrastructure: historical data from [www.estacoesferroviarias.com.br](http://www.estacoesferroviarias.com.br).

# Immigrant Diversity Index

- ▶ To measure birthplace diversity among the immigrant population we follow Alesina, Harnoss e Rapoport (2016):

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

Where  $s_j$  indicates the proportion of immigrants with origin  $j$  among the total number of foreigners in municipality  $m$ .

# Descriptive Statistics

| VARIABLES                                 | (1)<br>Avg. | (2)<br>St.Dev. | (3)<br>Min | (4)<br>Max | (5)<br>N |
|---|-------------|----------------|------------|------------|----------|
| <b>A. Diversity e Share of Foreigners</b> |             |                |            |            |          |
| 1920 Birthplace Diversity Index           | 0.431       | 0.284          | 0.000      | 0.821      | 202      |
| 1872 Share of Foreigners                  | 0.012       | 0.016          | 0.000      | 0.081      | 202      |
| 1920 Share of Foreigners                  | 0.125       | 0.096          | 0.000      | 0.357      | 202      |
| 1872 Dummy railway                        | 0.024       | 0.156          | 0.000      | 1.000      | 202      |
| 1920 Dummy railway                        | 0.738       | 0.441          | 0.000      | 1.000      | 202      |
| <b>C. 1872 Census Variables</b>           |             |                |            |            |          |
| Proportion of slaves                      | 0.154       | 0.086          | 0.039      | 0.531      | 202      |
| Proportion of literate (>6 y.o.)          | 0.202       | 0.110          | 0.048      | 0.452      | 202      |
| Proportion of children enrolment          | 0.145       | 0.101          | 0.027      | 0.764      | 202      |
| Population (× 1,000)                      | 11.147      | 7.198          | 1.566      | 41.751     | 202      |
| Proportion in agriculture                 | 0.594       | 0.099          | 0.351      | 0.908      | 202      |
| Proportion in industry                    | 0.109       | 0.044          | 0.020      | 0.244      | 202      |
| Proportion in services/retail             | 0.296       | 0.091          | 0.058      | 0.569      | 202      |

# Descriptive Statistics

| VARIABLES                             | (1)<br>Avg. | (2)<br>St.Dev. | (3)<br>Min | (4)<br>Max | (5)<br>N |
|---------------------------------------|-------------|----------------|------------|------------|----------|
| <b>D. 1920 Census Variables</b>       |             |                |            |            |          |
| Proportion of literate (>6 y.o.)      | 0.296       | 0.099          | 0.103      | 0.699      | 202      |
| Schools/school-age child (x 1,000)    | 0.387       | 0.377          | 0.000      | 1.863      | 202      |
| Professors/school-age child (x 1,000) | 9.562       | 6.938          | 0.826      | 45.955     | 202      |
| Population (x 1,000)                  | 22.635      | 43.101         | 2.917      | 577.621    | 202      |
| Proportion in agriculture             | 0.784       | 0.133          | 0.071      | 0.957      | 202      |
| Proportion in industry                | 0.091       | 0.071          | 0.005      | 0.495      | 202      |
| Proportion in services/retail         | 0.125       | 0.072          | 0.035      | 0.549      | 202      |
| Proportion of catholics               | 0.849       | 0.182          | 0.098      | 1.000      | 202      |
| Proportion of protestants (1940)      | 0.019       | 0.015          | 0.000      | 0.076      | 202      |
| <b>E. 2000 Census Variables</b>       |             |                |            |            |          |
| Proportion of literate (>6 y.o.)      | 0.952       | 0.016          | 0.854      | 0.989      | 202      |
| Proportion of children enrolment      | 0.962       | 0.018          | 0.863      | 0.991      | 202      |
| Schools/school-age child (x 1,000)    | 0.387       | 0.377          | 0.000      | 1.863      | 202      |
| Professors/school-age child (x 1,000) | 82.833      | 21.228         | 19.670     | 152.683    | 202      |
| Years of schooling (>5 y.o.)          | 5.538       | 0.660          | 3.532      | 7.109      | 202      |
| Population (x 1,000)                  | 183.329     | 818.392        | 2.867      | 11,086.8   | 202      |
| Per capita income (log)               | 5.693       | 0.272          | 4.791      | 6.392      | 202      |
| Proportion in agriculture             | 0.197       | 0.134          | 0.003      | 0.595      | 202      |
| Proportion in industry                | 0.252       | 0.087          | 0.097      | 0.545      | 202      |
| Proportion in services/retail         | 0.533       | 0.098          | 0.254      | 0.784      | 202      |

# Preliminary Evidence: OLS

We start by estimating the following OLS equation:

$$Y_{m,2000} = \beta_0 + \beta_1 Div_{m,1920} + X'_m \gamma + e_m \quad (2)$$

- ▶  $y_{m,2000}$ : per capita income in 2000 of municipality  $m$ .
- ▶  $Div_{m,1920}$ : birthplace diversity index of immigrants.
- ▶  $X_m$ : includes geographic characteristics, such as latitude, longitude, distance from the capital and elevation, soil type. Socioeconomic characteristics at baseline (1872 census): the proportion of slaves, children enrolled in school, literacy rate of the total population, the proportion of workers occupied by sector, and population density.



## Preliminary Evidence: OLS

| Dependent Variable:            | Ln per capita Income in 2000 |                     |                     |                     |                     |                     |
|--------------------------------|------------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|                                | (1)                          | (2)                 | (3)                 | (4)                 | (5)                 | (6)                 |
| 1920 Immigrant Diversity Index | 0.097<br>(0.014)***          | 0.096<br>(0.014)*** | 0.085<br>(0.015)*** | 0.074<br>(0.021)*** | 0.074<br>(0.021)*** | 0.071<br>(0.021)*** |
| Observations                   | 202                          | 202                 | 202                 | 202                 | 202                 | 202                 |
| R-squared                      | 0.566                        | 0.566               | 0.576               | 0.573               | 0.573               | 0.579               |
| Geographical Controls          | YES                          | YES                 | YES                 | YES                 | YES                 | YES                 |
| Socioecon Controls (1972)      | YES                          | YES                 | YES                 | YES                 | YES                 | YES                 |
| Share of Foreigners (1872)     |                              | YES                 |                     |                     | YES                 |                     |
| Share of Foreigners (1920)     |                              |                     | YES                 |                     |                     | YES                 |
| Railway Connection Timing      |                              |                     |                     | YES                 | YES                 | YES                 |

- ▶ Estimates robust when controlling for immigrant population size.
- ▶ Railway Connection Timing controls for the time elapsed since the municipality was connected to the network.
- ▶ A 1 SD increase in the Diversity Index in 1920 is associated with a 7.1-7.4 percent higher income per capita in year 2000.

# Sources of Bias in OLS

- ▶ OLS estimates are very likely biased due to the endogeneity of migrant diversity.

## Sources of Bias in OLS

- ▶ OLS estimates are very likely biased due to the endogeneity of migrant diversity.
- ▶ On the one hand, destinations with high economic potential might attract many migrants, regardless of their origin country, ending up with both high migrant share and high diversity index. Positive spurious correlation between diversity and growth.

## Sources of Bias in OLS

- ▶ OLS estimates are very likely biased due to the endogeneity of migrant diversity.
- ▶ On the one hand, destinations with high economic potential might attract many migrants, regardless of their origin country, ending up with both high migrant share and high diversity index. Positive spurious correlation between diversity and growth.
- ▶ On the other hand, early arrived migrant groups in high economic potential destinations might prevent other immigrant groups from coming in. Negative spurious correlation between diversity and growth.

## Sources of Bias in OLS

- ▶ OLS estimates are very likely biased due to the endogeneity of migrant diversity.
- ▶ On the one hand, destinations with high economic potential might attract many migrants, regardless of their origin country, ending up with both high migrant share and high diversity index. Positive spurious correlation between diversity and growth.
- ▶ On the other hand, early arrived migrant groups in high economic potential destinations might prevent other immigrant groups from coming in. Negative spurious correlation between diversity and growth.
- ▶ Finally, migrants might endogenously sort into areas with high/low diversity according to their characteristics and productivity.

# Empirical Strategy: Intuition

1. The total amount and the composition of immigrant inflow in terms of origin countries varied considerably over time. Driven mostly by shocks at home. [shocks origin countries](#)

# Empirical Strategy: Intuition

1. The total amount and the composition of immigrant inflow in terms of origin countries varied considerably over time. Driven mostly by shocks at home. [shocks origin countries](#)
2. The size and the composition of immigrant inflows display independent variation. [quantities vs diversity](#)

# Empirical Strategy: Intuition

1. The total amount and the composition of immigrant inflow in terms of origin countries varied considerably over time. Driven mostly by shocks at home. [shocks origin countries](#)
2. The size and the composition of immigrant inflows display independent variation. [quantities vs diversity](#)
3. Period under study coincides with the expansion of the railway network in the São Paulo State. [rail network expansion](#)



# Empirical Strategy: Intuition

- ▶ **Example:** Municipalities connected to the network when migrant population inflow is more heterogeneous ended up with a more diverse pool of migrants in 1920.

# Empirical Strategy: Intuition

- ▶ **Example:** Municipalities connected to the network when migrant population inflow is more heterogeneous ended up with a more diverse pool of migrants in 1920.
- ▶ **Advantage of combining two sources of variation:** the interaction between the two produces variation that is unlikely to affect our contemporary outcomes of interest other than through historical immigrant composition to the municipality.

# Empirical Strategy: Intuition

- ▶ **Example:** Municipalities connected to the network when migrant population inflow is more heterogeneous ended up with a more diverse pool of migrants in 1920.
- ▶ **Advantage of combining two sources of variation:** the interaction between the two produces variation that is unlikely to affect our contemporary outcomes of interest other than through historical immigrant composition to the municipality.
- ▶ **Proceed in 3 steps:** a zero-stage to construct the cross-sectional 1920 instrument from the 1880-1920 panel dataset, and the usual first and second stage.

## Zero-Stage: Synthetic Diversity Index Construction

We estimate a set of zero-stage equations using yearly panel data at the municipality-country of origin level:

$$Imm_{m,t}^c = \alpha_0 + \alpha_1 ImmFlow_t^c \times RailAccess_{m,t} + \alpha_2 RailAccess_{m,t} + \kappa_m + \mu_t + u_{m,t}^c \quad (3)$$

- ▶  $ImmFlow_t^c$ : the number of immigrants arriving into São Paulo from country  $c$  in year  $t$
- ▶  $RailAccess_{m,t}$ : an indicator for the municipality being connected to the railway network at time  $t$
- ▶ The parameter  $\alpha_1$  captures the differential effect that connection to the railway has on immigrant settlement during periods of high aggregate immigration relative to periods of low aggregate immigration, from each specific country.

## Zero-Stage: Results

| Dependent variable     | Number of immigrants arriving to the municipality |                     |                    |                     |                    |
|------------------------|---|---------------------|--------------------|---------------------|--------------------|
|                        | ITA   | ESP                 | POR                | GER                 | JAP                |
|                        | (1)   | (2)                 | (3)                | (4)                 | (5)                |
| Imm Flow x Rail Access | 0.011<br>(0.002)***                               | 0.008<br>(0.001)*** | 0.007<br>(0.003)** | 0.006<br>(0.002)*** | 0.008<br>(0.003)** |
| Rail Access            | -18.826<br>(7.549)**                              | -12.081<br>(10.786) | -9.484<br>(11.703) | -0.487<br>(0.611)   | -0.312<br>(2.217)  |
| Observations           | 7,878   | 7,878               | 7,878              | 7,878               | 7,878              |
| R-squared              | 0.456   | 0.381               | 0.210              | 0.274               | 0.113              |
| Year FE                | YES   | YES                 | YES                | YES                 | YES                |
| Municipality FE        | YES   | YES                 | YES                | YES                 | YES                |

## Zero-Stage: Synthetic Diversity Index Construction

We then predict, using only the interaction term, the number of immigrants who have settled by municipality and origin country, for all the years from 1882 to 1920 as follows:

$$\widehat{Imm}_{m,t}^c = \widehat{\alpha}_1 ImmFlow_t^c \times RailAccess_{m,t} \quad (4)$$

- ▶ Add up the predicted number of immigrants over the entire period 1882-1920.
- ▶ Finally, calculate the Synthetic Diversity Index at the municipality level and use it as an IV for the observed Diversity Index.

# TSLS and Identification

$$\begin{aligned} \underline{1stStage} : Diversity_{m,1920} = & \delta_0 + \delta_1 SyntheticDiversity_{m,1920} + \\ & + \delta_2 ShareForeigners_m + \delta_3 RailTiming_m + X'\theta_m + v_m \end{aligned} \quad (5)$$

$$\begin{aligned} \underline{2ndStage} : Y_{m,2000} = & \gamma_0 + \gamma_1 \widehat{Diversity}_{m,1920} + \\ & + \gamma_2 ShareForeigners_m + \gamma_3 RailTiming_m + X'\eta_m + \epsilon_m \end{aligned} \quad (6)$$

- ▶  $ShareForeigners_m$ : share of foreigners either in 1872 or 1920.
- ▶  $RailTiming_m$ : number of years elapsed since railway connection.
- ▶  $X'$ : Same geographic and economic municipality controls measured at baseline as in OLS regressions.

# TSLS and Identification

$$\begin{aligned} \underline{1stStage} : Diversity_{m,1920} = & \delta_0 + \delta_1 SyntheticDiversity_{m,1920} + \\ & + \delta_2 ShareForeigners_m + \delta_3 RailTiming_m + X'\theta_m + v_m \end{aligned} \quad (5)$$

$$\begin{aligned} \underline{2ndStage} : Y_{m,2000} = & \gamma_0 + \gamma_1 \widehat{Diversity}_{m,1920} + \\ & + \gamma_2 ShareForeigners_m + \gamma_3 RailTiming_m + X'\eta_m + \epsilon_m \end{aligned} \quad (6)$$

- ▶  $ShareForeigners_m$ : share of foreigners either in 1872 or 1920.
- ▶  $RailTiming_m$ : number of years elapsed since railway connection.
- ▶  $X'$ : Same geographic and economic municipality controls measured at baseline as in OLS regressions.
- ▶ **Identifying assumption**:  $SyntheticDiversity$  affects long-term outcomes only through observed diversity in 1920, once conditioning on municipality controls, the timing of railway connection, and the total share of immigrants.



# First-Stage Results

| Dependent Variable                    | 1920 Observed Diversity Index |                     |                     |                     |                     |                     |
|---------------------------------------|-------------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|                                       | (1)                           | (2)                 | (3)                 | (4)                 | (5)                 | (6)                 |
| 1920 Synthetic Diversity Index        | 0.764<br>(0.061)***           | 0.753<br>(0.061)*** | 0.703<br>(0.068)*** | 0.487<br>(0.076)*** | 0.491<br>(0.076)*** | 0.486<br>(0.076)*** |
| Share of Foreigners (1872)            |                               | 0.007<br>(0.004)*   |                     |                     | 0.003<br>(0.003)    |                     |
| Share of Foreigners (1920)            |                               |                     | 0.575<br>(0.238)**  |                     |                     | 0.026<br>(0.239)    |
| Railway Connection Timing             |                               |                     |                     | 0.009<br>(0.001)*** | 0.008<br>(0.001)*** | 0.008<br>(0.001)*** |
| Observations                          | 202                           | 202                 | 202                 | 202                 | 202                 | 202                 |
| Partial-F (Synthetic Diversity Index) | 156.07                        | 149.96              | 106.47              | 40.84               | 41.25               | 40.87               |
| R-squared                             | 0.548                         | 0.560               | 0.560               | 0.650               | 0.651               | 0.650               |
| Geo Controls                          | YES                           | YES                 | YES                 | YES                 | YES                 | YES                 |
| Socioecon Controls (1872)             | YES                           | YES                 | YES                 | YES                 | YES                 | YES                 |

- ▶ The instrument is a strong predictor of actual birthplace Diversity Index in 1920. F-stat always above 40.
- ▶ Coefficients almost unaffected by the inclusion of share of foreigners: instrument loads on variation uncorrelated with the immigrant stock's size.
- ▶ Instrument remains strong also when controlling for Railway Connection Timing.

# The Impact of Immigrant Diversity on Income (TSLS)

| Dependent Variable:                   | Ln per capita Income in 2000 |                     |                     |                    |                    |                   |
|---------------------------------------|------------------------------|---------------------|---------------------|--------------------|--------------------|-------------------|
|                                       | (1)                          | (2)                 | (3)                 | (4)                | (5)                | (6)               |
| 1920 Immigrant Diversity Index        | 0.107<br>(0.018)***          | 0.106<br>(0.019)*** | 0.087<br>(0.023)*** | 0.081<br>(0.039)** | 0.081<br>(0.038)** | 0.068<br>(0.039)* |
| Share of Foreigners (1872)            |                              | 0.004<br>(0.018)    |                     |                    | 0.002<br>(0.018)   |                   |
| Share of Foreigners (1920)            |                              |                     | 0.048<br>(0.028)*   |                    |                    | 0.040<br>(0.026)  |
| Railway Connection Timing             |                              |                     |                     | 0.002<br>(0.002)   | 0.002<br>(0.002)   | 0.002<br>(0.002)  |
| Observations                          | 202                          | 202                 | 202                 | 202                | 202                | 202               |
| Partial-F (Synthetic Diversity Index) | 156.07                       | 149.96              | 106.47              | 40.84              | 41.25              | 40.87             |
| R-squared                             | 0.565                        | 0.573               | 0.579               | 0.573              | 0.576              | 0.565             |
| Geo Controls                          | YES                          | YES                 | YES                 | YES                | YES                | YES               |
| Socioecon Controls (1872)             | YES                          | YES                 | YES                 | YES                | YES                | YES               |

- ▶ TSLS confirm positive long-term impact of Immigrant Diversity on income.
- ▶ SE gets larger when including Railway Connection Timing but size of impact fairly stable.
- ▶ A 1 SD increase in the Diversity Index in 1920 is associated with a 6.8 to 8.1 percent higher income per capita in year 2000.

# Structural Transformation and Occupational Diversity

- ▶ Birthplace diversity might have **enhanced productivity through complementarities** among cultures, skills, professions and trades brought from the places of origin (Murard 2018; Menyhert 2018).

# Structural Transformation and Occupational Diversity

- ▶ Birthplace diversity might have **enhanced productivity through complementarities** among cultures, skills, professions and trades brought from the places of origin (Murard 2018; Menyhert 2018).
- ▶ We test (using census data from 1920, 1940 and 2000,) whether municipalities with a more diverse mix of immigrants have shifted faster from agriculture to manufacturing and services, and if they have developed an higher degree of industrial occupational diversity.

# Structural Transformation and Occupational Diversity

|   | 1920                 | 1940                | 2000                 | 1920                | 1940              | 2000                 |
|---|----------------------|---------------------|----------------------|---------------------|-------------------|----------------------|
|   | OLS                  | OLS                 | OLS                  | 2SLS                | 2SLS              | 2SLS                 |
|   | (1)                  | (2)                 | (3)                  | (4)                 | (5)               | (6)                  |
| <i>Panel A: Share Agricultural Occupation</i>     |                      |                     |                      |                     |                   |                      |
| 1920 Imm. Diversity Index                         | -0.023<br>(0.009)*** | -0.025<br>(0.012)** | -0.036<br>(0.011)*** | -0.030<br>(0.019)   | -0.033<br>(0.028) | -0.062<br>(0.024)*** |
| <i>Panel B: Share Industrial Occupation</i>       |                      |                     |                      |                     |                   |                      |
| 1920 Imm. Diversity Index                         | 0.012<br>(0.004)***  | 0.004<br>(0.006)    | 0.000<br>(0.007)     | 0.009<br>(0.010)    | 0.027<br>(0.020)  | 0.034<br>(0.018)*    |
| <i>Panel C: Share Services/Retail Occupation</i>  |                      |                     |                      |                     |                   |                      |
| 1920 Imm. Diversity Index                         | 0.011<br>(0.005)**   | 0.022<br>(0.008)*** | 0.034<br>(0.009)***  | 0.020<br>(0.011)*   | 0.007<br>(0.014)  | 0.024<br>(0.017)     |
| <i>Panel D: Industrial Occupational Diversity</i> |                      |                     |                      |                     |                   |                      |
| 1920 Imm. Diversity Index                         | 0.006<br>(0.009)     |                     | 0.010<br>(0.008)     | 0.058<br>(0.019)*** |                   | 0.052<br>(0.020)***  |
| Observations                                      | 202                  | 202                 | 202                  | 202                 | 202               | 202                  |

# Schooling Investment and Human Capital Accumulation

1. The faster structural transformation might **increase incentives to invest in human capital**.
2. Higher public investment in schooling might also be a response to high birthplace diversity, a **state-building tool to foster homogenization** (Bandiera et al 2018).

# Schooling Investment and Human Capital Accumulation

1. The faster structural transformation might **increase incentives to invest in human capital**.
  2. Higher public investment in schooling might also be a response to high birthplace diversity, a **state-building tool to foster homogenization** (Bandiera et al 2018).
- ▶ We test the impact on government investment in schooling and educational outcomes.

# Schooling Investment and Human Capital Accumulation

|                           | Schools per<br>capita 1920 | Schools per<br>capita 1940 | Education<br>expenditure<br>per capita<br>1940 | Enrollment<br>rate 7-14 y.o<br>1940 | Years of<br>education<br>2000 |
|---------------------------|----------------------------|----------------------------|--|-------------------------------------|-------------------------------|
|                           | (1)                        | (2)                        | (3)  | (4)                                 | (5)                           |
| <i>Panel A: OLS</i>       |                            |                            |  |                                     |                               |
| 1920 Imm. Diversity Index | 0.027<br>(0.040)           | 0.287<br>(0.162)*          | 0.196<br>(0.068)***                            | 0.027<br>(0.009)***                 | 0.138<br>(0.049)***           |
| <i>Panel B: 2SLS</i>      |                            |                            |  |                                     |                               |
| 1920 Imm. Diversity Index | 0.249<br>(0.077)***        | 0.761<br>(0.392)*          | 0.373<br>(0.124)***                            | 0.040<br>(0.024)*                   | 0.259<br>(0.111)**            |
| Observations              | 202                        | 202                        | 202  | 202                                 | 202                           |



# Immigrant Diversity and Inequality

| Dependent Variable:            | Gini coefficient in 2000 |                   |                    |                  |                  |                  |
|--------------------------------|--------------------------|-------------------|--------------------|------------------|------------------|------------------|
|                                | OLS<br>(1)               | OLS<br>(2)        | OLS<br>(3)         | 2SLS<br>(4)      | 2SLS<br>(5)      | 2SLS<br>(6)      |
| 1920 Immigrant Diversity Index | 0.007<br>(0.004)*        | 0.007<br>(0.004)* | 0.009<br>(0.004)** | 0.004<br>(0.010) | 0.004<br>(0.010) | 0.011<br>(0.010) |
| Observations                   | 202                      | 202               | 202                | 202              | 202              | 202              |
| Geo Controls                   | YES                      | YES               | YES                | YES              | YES              | YES              |
| Socioecon Controls (1872)      | YES                      | YES               | YES                | YES              | YES              | YES              |
| Railway Connection Timing      | YES                      | YES               | YES                | YES              | YES              | YES              |
| Share of Foreigners (1872)     |                          | YES               |                    |                  | YES              |                  |
| Share of Foreigners (1920)     |                          |                   | YES                |                  |                  | YES              |

- ▶ OLS: Positive relationship between Immigrant Diversity in 1920 and inequality in 2000.
- ▶ TSLS: Coefficients still positive but imprecisely estimated.

## Interaction between Size and Composition

Does the impact of diversity depend on the share of the immigrant population on the total?

| Dependent Variable:                   | Ln per capita Income 2000 |                     |                    |                    |
|---------------------------------------|---------------------------|---------------------|--------------------|--------------------|
|                                       | OLS<br>(1)                | OLS<br>(2)          | 2SLS<br>(3)        | 2SLS<br>(4)        |
| 1920 Immigrant Diversity Index        | 0.114<br>(0.021)***       | 0.092<br>(0.025)*** | 0.140<br>(0.067)** | 0.144<br>(0.071)** |
| Diversity Index * Share of Foreigners | 0.060<br>(0.053)          | 0.070<br>(0.053)    | 0.098<br>(0.189)   | 0.184<br>(0.205)   |
| Share of Foreigners                   | -0.030<br>(0.031)         | -0.040<br>(0.030)   | -0.055<br>(0.112)  | -0.109<br>(0.122)  |
| Observations                          | 202                       | 202                 | 202                | 202                |
| Geo Controls                          | YES                       | YES                 | YES                | YES                |
| Socioecon Controls (1872)             | YES                       | YES                 | YES                | YES                |
| Railway Connection Timing             |                           | YES                 |                    | YES                |

- ▶ Share of foreigners calculated in 1872 to avoid endogeneity.
- ▶ The impact of diversity does not seem to be larger or smaller when immigrant share is large.

## Conclusion and Next Steps

- ▶ This paper tackles the question: is migrant composition in terms of birthplace diversity beneficial for long-run development?
- ▶ Looks at the mass migration in Brazil 1880-1920 and uses newly available data about more than 1 million immigrants.
- ▶ Finds a positive effect on income today of birthplace diversity.

## Conclusion and Next Steps

- ▶ This paper tackles the question: is migrant composition in terms of birthplace diversity beneficial for long-run development?
- ▶ Looks at the mass migration in Brazil 1880-1920 and uses newly available data about more than 1 million immigrants.
- ▶ Finds a positive effect on income today of birthplace diversity.
- ▶ Next steps
  - ▶ What diversity is beneficial for development? Construct measures of language distance, cultural diversity, etc.
  - ▶ Explore heterogeneity in characteristics of countries of origin (policies individuals had been exposed to before migration, etc. )
  - ▶ Do migrants took values with them?

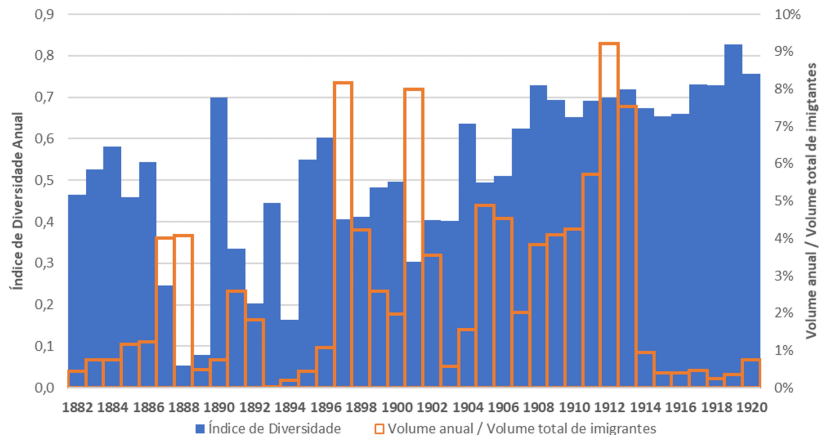
# Appendix

## Identification - Exogenous push factors

- ▶ Italy: *Decreto Perinetti* of 1902. Prevented the so-called subsidizes emigration directed especially to Brazil.
- ▶ Portugal: Republican revolution in 1910; Between 1908 e 1913 34% of the total inflow registered during the 77 years of the Hospedaria.
- ▶ Japan: Legislation change in 1912 allowed land acquisition. *Alien Land Act of 1913* in the US would forbid it. Average annual inflow from Japan grows 8 times from 1912 on compared with 1908-1912 period.
- ▶ Germany: political instability between 1889-1893 (fall of Bismarck). Flow grows six fold.
- ▶ Spain: great drought between 1893-1897.

[Back to main](#)

# Identification - Volume VS. Diversity



**Figure:** Volume VS. diversity of immigrants between 1882 and 1920

# Identification - Railway network expansion



**Figure:** Rail network year 1882



# Identification - Railway network expansion



**Figure:** Rail network year 1890

# Identification - Railway network expansion



**Figure:** Rail network year 1900

# Identification - Railway network expansion



**Figure:** Rail network year 1910

# Identification - Railway network expansion



**Figure:** Rail network year 1920

# Immigrant Diversity and Long-Run Development: Evidence from the Age of Mass Migration in Brazil

Luigi Minale (Universidad Carlos III Madrid, CReAM and IZA)

Rudi Rocha (FGV)

Bruno Vigna (IE-UFRJ)

EEA Congress  
30 August, 2023  
Barcelona