Inherited cultural diversity and wages in Brazil*

Philipp Ehrl
Universidade Católica de Brasília

Leonardo Monasterio
Universidade Católica de Brasília, IPEA

June 10, 2017

Abstract

This paper estimates the long-term impact of immigration to Rio Grande do Sul/Brazil on contemporary wages. Based on a unique micro-data panel that includes the names of workers, we apply machine learning algorithms to classify surnames and infer each workers' ancestry in order to calculate the inherited cultural diversity in the workforce by municipality. We address the endogeneity of cultural diversity by using three sets of instruments: distance to settlements created by the government for non-Iberian immigrants between 1824-1918, share of street names with foreign surnames and share of foreigners in 1920. Our IV-estimations prove robust to human capital differences, institutions, geography, the spatial sorting of workers based on intrinsic abilities and the diffusion of knowledge through imports. The results clearly indicate that an increase in diversity – exclusively transmitted through the share of workers with non-Iberian ancestry – leads to a positive wage externality.

Keywords: Ancestry, Immigration, Cultural Diversity, Agglomeration Economies, Brazil
JEL Classification: J31, R12, C26, N16

* *We are grateful to Irineu de Carvalho Filho for providing the data on colonies and to Rodrigo Soares for providing the data on institutions. We also thank participants in the UCLA's Economic History Seminar, the UEA Copenhagen, the IPEA Brasília and the UCB for their comments and suggestions that helped to improve the paper. Monasterio gratefully acknowledges financial support from CNPq grant no. 445040/2015-7.*

1 Introduction

Foreign immigration has various positive effects on the economic performance of countries (Alesina and Ferrara 2005) and regions (Kemeny 2014). Besides consumption amenities, a number of recent micro-level studies document that cultural diversity generates favorable localized effects on the productivity of native workers (Bakens *et al.* 2013; Ottaviano and Peri 2006).[1] The skill level of migrants is found to be pivotal (Südekum *et al.* 2014), as well as the assimilation of the immigrants (Ottaviano and Peri 2005). These paper are in agreement that the share of immigrants in the population is at least an equally important component of cultural diversity as the diversity within the group of immigrants. In this paper, we address the question in how far these of cultural assets are inheritable by future generations and whether a foreign cultural background still generates measurable effects on the local labor market?

This question has direct implications for national migration policies and laws. If foreign immigrants generate positive externalities not only in the present but also in the future through their offspring, it would be rational for countries to incur a certain amount of costs for the attraction and assimilation of foreigners. At the micro-level, Dohmen *et al.* (2012) show that some attitudes like trust and risk are inherited from the parents. Regarding culture, Borjas (1995) already showed that convergence is slow and that ethnic skill difference persist over generations. Bisin and Verdier (2000) demonstrate that persistence can be accomplished by intra-ehnicity marriages and higher socialization efforts by minority groups. There is also a large literature on the deeper causes of cross-country income differences, recently surveyed by Spolaore and Wacziarg (2013), where the current consensus seems to be that culture, human capital (and other things European settlers brought with them) are more important for the development of nations than institutions or geography.

In contrast to most of the previous literature, we study a middle-income country. Brazil is an interesting country to analyze because it experienced mass immigration waves before World War I, as many other countries did, but nowadays its share of foreigners is below most industrialized countries. It is thus a perfect case for studying whether this substantial, positive migration shock generated persistent cultural diversity and whether it still affects economic outcomes. This study focuses only on the Federal State of Rio Grande do Sul for several reasons. First of all, the unique history of the state is an important source of exogeneity in our identification strategy. The government created several settlements for subsidized non-Iberian immigrants and, as we will argue in detail, their location was chosen for mostly non-economic reasons. Second, exploiting only intra-state variation means that, compared to cross-country studies, we have to deal with less confounding factors. Finally, it allows us to work with a panel that comprises the entire working population without being constrained by computational limitations.[2]

[1] Apart from the positive effect of workers productivity, Niebuhr (2010) finds higher innovation rates in culturally more diverse regions and Trax *et al.* (2015) confirm that firms also experience a higher productivity.

[2] Our individual/yearly panel data of all workers in Rio Grande do Sul from 2008–2013 already has more than 10.8 million observations; the same database for Brazil would exceed 300 million observations.

The present paper applies machine learning algorithms to infer the ancestry of the current Brazilian workforce from their surnames. We then use the ancestry classification to calculate measures for the diversity of cultural backgrounds in analogy to the previous literature.[3] It turns out to be crucial to decompose the common fractionalization index[4] into the share of the majoritarian cultural group (Iberians in our case) and the fractionalization of the remaining cultures. The effect of these diversity measures is estimated in several individual wage regressions. To isolate the productivity externality of inherited cultural diversity we account for the endogeneity of the regional diversity variables using long-lagged instruments based on the location of official colonies for immigrants. Other instruments exploits the diversity of street names and the diversity of immigrants in 1920 in the respective region. In addition, we combine the 2SLS estimation with a prior a estimation step where individual fixed effects are controlled for, similar to the approach in Combes *et al.* (2008). This approach makes sure that the identified effect is neither due to any unobservable intrinsic skill or human capital differences between people with different ancestry, nor due to the spatial sorting of workers based on individual heterogeneity. We argue that the remaining effect is really an externality that arises from the composition of diverse cultural backgrounds. Other variables that potentially transmit a persistent productivity effect or may be interdependent with foreign immigration, such as geography, institutions, barriers to the diffusion of knowledge and internal migration are controlled for in several robustness checks.

Our paper reveals good news for countries that have or had open borders: Cultural diversity creates a positive wage externality that accrues to all individuals in the local labor market. Note that all workers in our sample are Brazilian citizens, thus the significant effect establishes the transmission of cultural traits within families. This kind of evidence is not provided by cross-country studies. The present results thus support the notion that immigrants as well as their ancestors posses different skills and attitudes that complement the abilities of workers from the majoritarian ancestry group in the production process. The decomposition of the fractionalization diversity index shows that the effect is exclusively transmitted through the share of workers with non-Iberian ancestry. This result suggest that it does not matter how diverse as long as there *is* cultural diversity in the sense of any deviation from the majoritarian ancestry group.The IV estimates suggest that a 10 percentage point increase in the share of workers with non-Iberian ancestry leads to a wage increase of 9%. Even though the estimations also reveal that more productive workers sort into more diverse areas, the wage externality of ancestral diversity remains but the wage increase drops to the order of 2–3%. The remaining robustness checks confirm that immigration of foreigners and the subsequent increase of cultural diversity not only has a positive effect through the current generation of immigrants but it will also have a

[3] According to the definition in Boyd and Richerson (1988: 33), "culture is information capable of affecting individuals' behavior that they acquire from other members of their species through teaching, imitation, and other forms of social transmission". For the rest of the paper, the term "cultural diversity" will refer to the diversity of workers' inherited cultural background according to their ancestral classification.

[4] The fractionalization index is equal to one minus the formula of the Herfindahl index, i.e., one minus the sum of squared shares of each ancestral group, cf. eq. (1).

persistent positive effect on future generations.[5]

Among the contributions to the economic effects of immigrant diversity at the regional level, the paper closest to ours is Bakens *et al.* (2013). The authors also account for the endogeneity bias and the sorting of workers, however, the latter is found to have the largest effect on wages. Hence, their final result is that cultural diversity itself merely has a small positive effect and, surprisingly, not so in the largest cities, i.e., the result is "not robust throughout the different samples" (Bakens *et al.* 2013: 27). Part of the divergences with respect to the present findings may be explained by the use of different instruments and because there is no distinction of the diversity measure (the fractionalization index) into the share of natives and the diversity among immigrants, as we decompose it. Among others, Bakens *et al.* (2013) rely on the shift-share migration instrument but question its exogeneity (because the coefficient on diversity becomes highly negative and significant). Regarding the 25 largest municipalities in our sample, we find that the coefficient of the share of non-Iberian ancestry is almost twice as large as in the full sample.

Most of the papers on the economic value of cultural diversity estimate either employment or housing rent equations alongside the wage regressions in order to disentangle whether the diversity effect operates via the demand side (productive externality) or the supply side (consumption amenity) of the labor market. Both Ottaviano and Peri (2006) and Ottaviano and Peri (2005) develop a theoretical framework using the Roback (1982) model to properly distinguish between both possible effects. In addition to the wage equation, the former yields an employment equation, while the latter model requires the estimation of an additional housing rent equation. All of the studies we are aware of that adopt either of the two approaches, namely Kemeny (2012), Bakens *et al.* (2013) and Südekum *et al.* (2014), confirm the dominance of the productivity channel. Given the consolidated position on this issue, we see no further need to include employment or housing rent regressions.

Our paper is also related to Ager and Brückner (2013) in the sense that they analyze the impact of immigration to the United States during the age of mass migration (1870-1920) through the channel of cultural diversity. However, their dependent variable is economic growth during that time while ours is the contemporary individual wage level, since we are concerned with the persistence effects of immigration on productivity. There are at least three more papers on the long-term effects of mass immigration to Brazil. de Carvalho Filho and Colistete (2010) find that the presence of immigrants in the municipalities of Sao Paulo increased the supply of public education. They also show that, one century later, students in these places still get higher scores on standard tests and income per capita is higher. Stolz *et al.* (2013) look at the arrival of immigrants in Brazilian states as a positive human capital shock. They argue that immigration caused increases on educational expenditures and estimate substantial economic long-term effects.[6] Rocha *et al.* (forthcoming) make another interesting contribution on the effect of immigration on the long-run development

[5] In this regard, our results support the importance of assimilation in Ottaviano and Peri (2005) because natives with an foreign ancestry are in a certain way assimilated immigrants.

[6] Musacchio *et al.* (2014) found no relationship between immigration and expenditures at the state level in Brazil. The authors argue that Brazilian states with egalitarian institutions that went trough commodity booms where able to spent their tax revenues on public education.

of Brazil. They identify official settlements in the State of Sao Paulo and document the local persistence of elevated human capital in these regions. While this is certainly one of the advantages from immigration, our estimation strategy rules out that the identified wage differentials are caused by individuals' human capital.

Recent studies have used names to investigate long-term social themes: the spatial distribution of families (Rodríguez-Díaz *et al.* 2015); investigation of social mobility (Clark 2015; Clark *et al.* 2015; Güell *et al.* 2014); and the use of first names to study immigrant assimilation in the United States (Carneiro *et al.* 2015). Concerning the applications of the classification of surnames by ancestry, there are papers in epidemiology (Lakha *et al.* 2011; Petersen *et al.* 2011) and contributions in the field of Political Science (Dancygier 2014) and of Anthropology (Susewind 2015). As far as we know, this is the first study to address the long-term impact of immigration and diversity using surnames.

The present paper is organized as follows. Section 2 contains the historical facts about the immigration to Rio Grande do Sul and the installation of state sponsored colonies. Section 3 explains the classification of ancestry by surnames as well as the details on the data sources and the construction of the most important variables. The fourth section presents the descriptive statistics and simple correlations between diversity, wages and the proposed instrumental variables. The fifth section explains the empirical strategy, whose results and robustness checks are in section 6. Section 7 concludes the paper.

2 Historical Background

2.1 Subsidized immigration to Rio Grande do Sul

Local and national governments subsidized immigration to Rio Grande do Sul, the southernmost state of Brazil, between 1824 and 1918.[7] In 1808, the state, which has an area comparable to continental Italy, was inhabited by 87,000 people. In 1830, its population had doubled; by 1872 the state reached 446,000 inhabitants; and by the 1920s the population of the state was above 2.1 million (IBGE 1990, 2003). Immigration played a major role in this population boom: in 1900, foreigners were around 12% of residents in the state Levy (1974). Nowadays, the share of immigrants in the state is negligible.

There are three main reasons for the government willingness to subsidize non-Iberian European immigration to Rio Grande do Sul. First, there were strategic motivations. The state was the site of several wars in the 19th century with its neighboring countries, as well as bloody internal conflicts. Second, there was a clear intent to raise the labor/land ratio for economic reasons (Lago 2014). Complaints about the high cost of labor were frequent, especially after the coffee boom in Rio de Janeiro and Sao Paulo. Third, elites complained about the racial composition of the Brazilian population. Many racist leaders and intellectuals declared that their fellow countrymen where too dark and too mixed,

[7] Parts of this and the next section are based on de Carvalho Filho and Monasterio (2012).

and non-Iberian European immigration was a solution to this "problem" (Andrews 1988; Seyferth 2002).

There was little consistency of immigration policies throughout 1824–1918 and legal (and de facto) abrupt changes were common (Roche 1969). In the beginning, the Empire had a leading role, while after 1845 the provincial government started to take part in the settlement projects. In general, the government was responsible for choosing the place of the settlement, cleaning the land, and paying for the trip from the European port to the final destination. The settlers got their plots for free, or bought them with subsidized credit. Although there was no monotonic change in the strength of the incentives, the general trend was toward the reduction of the subsidies to the immigrants.[8]

After decades of mass immigration, particularly from Germany and Italy, the previous attitude of the government to cluster immigrants from the same nation changed. The unification of Germany in the 1870s was seen as a potential threat to Brazil (Gertz 1991). Fearing the creation of a *Neue Deutschland* in the state, they decided to establish colonies that mixed non-Iberian European settlers from several countries. de Carvalho Filho and Monasterio (2012) identify and locate 49 official settlements, of which 18 were settled by German colonists, 21 by a combination of German and other nationalities, and 10 without Germans.[9]

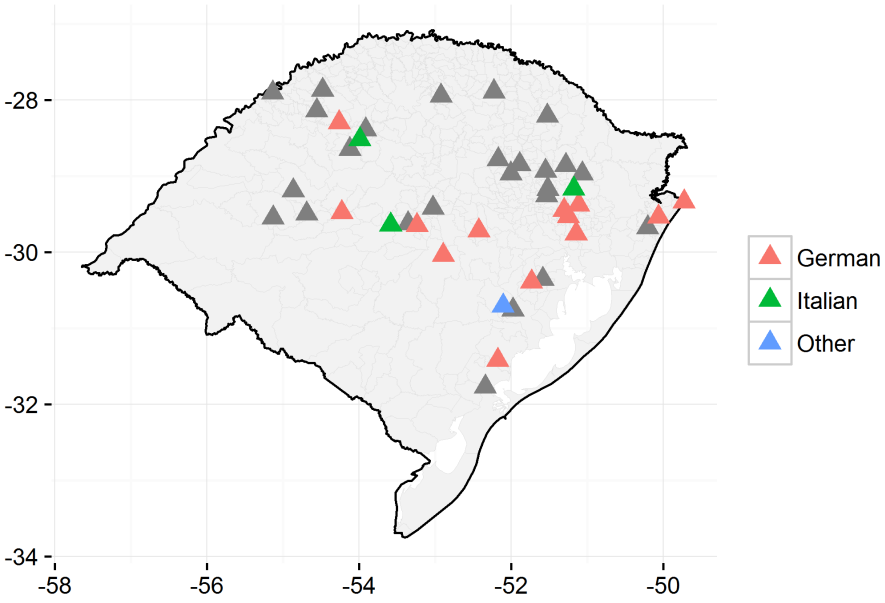
2.2 Exogeneity of official settlements

We have good reasons to suppose that the location of official colonies was exogenous to contemporary local productivity. The general purpose of Brazil's location policy, consistent with the reasons for supporting immigration, was to scatter official settlements all over the state. Provincial and National governments expected that European settlements would become "seeds" of development throughout Rio Grande do Sul (Roche 1969). Some of them were entrepot on the way to other regions, while others were "islands", i.e., isolated settlements created to occupy uninhabited areas (Amstad 1999; Roche 1969). Their development was also very different. As one would expect from randomly located settlements, some colonies failed, surviving only in the historical records. Others thrived and become prosperous cities. Furthermore, anecdotal evidence shows that sometimes settlers had to be deceived in order to accept the location of their plots of land (Amstad 1999: 87).

The locations of the official colonies are depicted in figure 1. Visual inspection suggests that their location was somewhat evenly spread all over the state. With the exception of the *Campanha* (Prairie) region, near the southern border, which was mostly settled by extensive cattle ranching before the 19th century, official colonies seem to be well-distributed over Rio Grande do Sul. We must point out that there were private colonies in Rio Grande do Sul but we are careful to focus just on the official ones. Private ones would jeopardize our identification strategy because their location was probably endogenous to their economic viability.

[8] See Iotti (2001) for the full changes in the historical legislation regarding immigration.
[9] The complete list of settlements is available in de Carvalho Filho and Monasterio (2012: 797).

Figure 1: Location of homogeneous and mixed official settlements



Notes: This map displays the location of official colonies in the State of Rio Grande do Sul. The axis indicate latitude and longitude. The color of the triangles indicate the origin of the settlers (as indicated in the graph's legend) and the grey triangles symbolize colonies where immigrants from several origins were mixed together. Furthermore, the light grey lines depict the borders of the current municipalities.

The settlers were not required to stay in the official colonies. Very high fertility rates among settlers families pushed their descendants to occupy new lands, often near the original colonies (Roche 1969).[10] It will be shown in section 4 that even today the share of people with non-Iberian surnames tends to fall as one moves away from the official settlements.

3 Surnames, ancestry and data

3.1 Surnames, race/color and ancestry in Brazil

People in Brazil usually have two surnames. The order is the inverse of Spanish speaking countries: the first is inherited from the mother and second from the father. Married woman can legally keep their maiden name, but most of the time they drop the surname of the mother and add that of the husband.[11] Following the literature on ancestry, we have analyzed only the patrilineal lineage.

Official census surveys and administrative data (such as RAIS, cf. section 3.4) in Brazil do not register information on the people's ancestry. Traditionally censuses have used just five categories to determine color/race, namely, white (*branco*), mixed (*pardo*), black (*preto*),

[10] The depletion of land productivity also may have contributed to the occupation of new areas, see Amstad (1999); Roche (1954); Waibel (1950).

[11] In fact, married men may also assume legally the wife's surname after marriage, but this is very rare.

far-eastern (*amarelo for the Chinese and Japanese*) and native-Brazilian (i.e. Amerindians, *indio*). Although those categories do have social significance, they are often far too broad to allow for specific applications such as ours.[12]

The next subsection describes how we obtain the ancestry classification. A surname classification algorithm is used, except when the individuals in question are registered in the RAIS database as being "black", "mixed" or "native". In such cases the original classification was maintained. In that way 8 categories of ancestry/skin color are obtained, namely, Iberian (IBR), Japanese (JPN), Italian (ITA), German (GER), Eastern European (EAS), black (BLK), mixed (MXD) and native (NAT).

3.2 The classification of surnames and street names

Monasterio (2017) describes in detail the process of classification of ancestry based on surnames. In summary, the process involves the following steps:

1. The building of a reference database that links surnames and ancestry. Based on both historical migration records and contemporary data, a reference document is created which associates surnames to ancestry. This file has a double function: to serve as the base for fuzzy matching and as training dataset for machine learning algorithms. From more than 5 million individual observations, 71 thousand unique names were identified and linked to an ancestry. This is the reference database;

2. The surnames in the reference database were fuzzy matched to the surnames of the workers. Not all names were classified in this initial step because many of them are not sufficiently similar to those in the reference table, but fuzzy matching classified the majority of individuals. Although only 293,634 of the 531,009 unique surnames found in the RAIS data were identified by fuzzy matching, that number corresponds to 96,4% of the workers. That is because the identified names are far more popular than those that were not identified.

3. The language classification method proposed by Cavnar and Trenkle (1994) was applied to the unmatched surnames of step #2. This algorithm was calibrated by means of the reference data created on step #1. Several machine learning methods were applied but the highest accuracy was reached by the relatively simple Cavnar and Trenkle (1994) method. The intuition of algorithm is the following: (i) a linguistic profile of each surname/ancestry is created based on the reference database. This profile contains the frequency of 3-grams (sequence of characters from the words) in each surname/ancestry; (ii) each new surname is broken into n-grams and compared to the each ancestry profile; (iii) the category with the least total distance in relation to the profiles is assigned to the new surname.

[12] Color and race in Brazil is a tricky issue and it would be impossible to deal with it here. Piza and Rosemberg (1999) and Osório (2004) discuss color and race in the Brazilian official surveys. Nogueira (1985) is the classic work that compares the different criteria for race identity in Brazil and USA.

In Monasterio (2017) and in this paper, the gathered historical and contemporary data records allow the consideration of Iberian (Spaniards and Portuguese), Italian, German, East European and Japanese ancestries, reflecting the main countries of origin of immigrants (Levy 1974). In Brazil's case, it is inappropriate to use surname methods to identify the ancestry of native Brazilians or Afro-Brazilians descendants because they have adopted or rather have been forced to adopt Iberian surnames. To classify the ancestry of these individuals, we resort to the official classification of the IBGE (Brazilian Institute of Geography and Statistics) according to the skin color which is provided in the individual RAIS data: "black", "mixed" (*pardo*) and "native".

Despite its simplicity, the Cavnar and Trenkle (1994) algorithm proved to be reasonably accurate. In the case of the Japanese surnames, 95% of the test data were classified correctly. The worst results were obtained for the Eastern European (EAS) surnames, where only 66% of the surnames in that category were effectively classified as such. Overall accuracy estimated accuracy was 80%. It should be borne in mind that, in spite of the classificatory algorithm's somewhat poor performance when compared to other papers, its classification errors only affected 3.6% of the individuals in the RAIS database, i.e. those not classified by fuzzy matching.

It is not possible to guarantee that the classification of surnames reflects precisely the cultural or genomic ancestry of workers. Even when both of the individual's surnames are used there is an inevitable loss of the matrilineal lineage and, furthermore, adoptions, name changes when marrying and other events can contribute to reducing the accuracy of that particular indicator. In the case of aggregated data, however, those idiosyncrasies tend to cancel one another and the overall trend is towards enhanced precision.

The same algorithm described in the previously section was applied to the street names of Rio Grande do Sul. As we show in section 5, the diversity of non-Iberian street names by municipality will be instrument variables in our estimations. In Brazil, it is traditional for street names to celebrate local or national distinguished individuals or people who have lived on those streets. Planned cities are rare and grid-based naming systems are even rarer. By law, the city council is responsible for choosing the names of public places and it can be change them, but most of the time they do preserve they official names. Furthermore, there is a national law that forbids naming streets after living people.

The names of streets were assembled from RAIS-Estabelecimentos, a database with contains the addresses of all firms in Brazil. For Rio Grande do Sul, there were 105,576 unique street name entries. The name of addresses that contained no reference to individuals, such as names that celebrate historical dates, where excluded. The mapping between ancestry and surnames was then applied to the last word of the street name. This allowed us to calculate an index for the (inherited) cultural diversity of the municipality according to the same procedure that we use to calculate the ancestral diversity index, cf. eq. (1).

3.3 Historical instruments

The Brazilian Census of 1920 has detailed information on the number of foreigners by municipality. Unfortunately, no information about color/race was collected at that time, however information about the nationality of foreigners is provided. Thus, instead of using ancestry data, we can calculate the share of each nationality group in each municipality. In order to calculate a diversity index for the year 1920, we infer the shares of each color/race of Brazilians from the 1940 Census. Although this may bring some bias to the diversity index in 1920, it is very probable that this is not a serious issue because there were no large migration flows between 1920 and 1940.

A notorious difficulty with historical data is that the boundaries of municipalities change over time. Particularly, the evolution of municipalities occurs in a non-linear fashion so that a one-to-one mapping between municipalities in 1920 and 2013 is often not possible. To assign the correct values of the historical diversity data to the current municipalities, we follow Ehrl (2017) and use stable, so called "Minimum Comparable Areas" (AMCs) for the 1920–2013 period.

To build the distance to colony variable, we first calculate a matrix with the distance in miles between the centers of all municipalities. Then we calculate the minimum between a given municipality and an official colony, applying this procedure to each of the three colony types. If the municipality itself contained an ex-colony, the distance is equal to zero. Finally, the instrument we use in the following estimations is the mean of the distance to each of the three official colony types.

3.4 RAIS

All individual data come from the Annual Social Information Report (Relação Anual de Informações Sociais-MIGRA) hereafter referred to as the RAIS (Ministério do Trabalho e Emprego 2016). The database is a very large restricted-access administrative file that contains full names, demographic and wage data of all Brazilians workers in the formal labor market. The quality and coverage of the database has increased, especially during the last decade.

Workers in the informal economy – mostly poor – are not registered in the RAIS database. However, this is not a major concern because Rio Grande de Sul has the second highest income per capita among the 26 Brazilian states, only behind São Paulo. So, even in very small municipalities there is a considerable number of formal workers. To check the validity of this argument, we perform the same analysis with the Census data from 2010 which contains all types of working individuals: formal workers, informal workers, employers and self-employed.

Every year there are around 4 million observations in the raw data of RAIS for Rio Grande do Sul. We gathered RAIS-MIGRA for each year between 2008 and 2013, that ads up to 10.8 million observations. We have cleaned the data excluding whoever receives less than

one minimum wage (probably interns), foreign born and we keep only the main employment of the workers (ie. the one with higher wage or more hours).

4 Descriptive Statistics

Table 1 summarizes the share of each ancestry group, as well as their average wage, population size and the values of the diversity measures used in the following. Obviously, people with an Iberian ancestry record are by far the largest group in the population of Rio Grande do Sul. The two next largest groups are Italian and German descendants. Together these two shares sum up to almost 1/3 of the population. In some municipalities the share of either German or Italian descendants is still as high as 40% and 53%, respectively. The two other immigrant groups from Eastern Europe and Japan play a limited role in comparison. In contrast to the population in most other Brazilian states, African descendants (categories MXD and BLK) only account for 6,4% of the working population in Rio Grande do Sul whereas, according to the Census in 2010, their combined share is slightly above 50%. This fact shows once more how strong the influence of European immigrants was and still is in this Federal State.

It is also worthwhile to note that the average wages as well as the current and past diversity measures vary substantially across the municipalities. The last line of table 1 points out that the average population of 21.407 in the municipalities of Rio Grande do Sul is rather low. Despite the capital Porto Alegre, there are only 8 cities with a population above 200.000 and no other city with a population above 500.000.

Table 1: Summary statistics – municipality aggregates

variable	mean	std. dev.	min.	max.
log(wage)	3.378	0.159	2.881	4.150
diversity	0.575	0.098	0.271	0.678
diversity non-IBR	0.636	0.078	0.308	0.776
diversity non-IBR (1920)	0.622	0.116	0.196	0.765
mean distance to colony	30.44	20.82	0.000	161.6
street name diversity	0.447	0.162	0.000	0.727
share IBR	0.582	0.112	0.356	0.850
share ITA	0.173	0.100	0.036	0.532
share GER	0.154	0.079	0.022	0.396
share EAS	0.024	0.012	0.004	0.051
share JPN	0.003	0.001	0.001	0.008
share MXD	0.038	0.021	0.004	0.161
share BLK	0.026	0.019	0.007	0.092
share NAT	0.001	0.002	0.000	0.020
population	21407	75105	1216	1394270

Notes: The number of observations is equal to 496. The acronyms of ancestry groups are the following: Iberian (IBR), Italian (ITA), German (GER), Eastern European (EAS), Japan (JPN), mixed race (MXD), black (BLK), Amerindians (NAT).

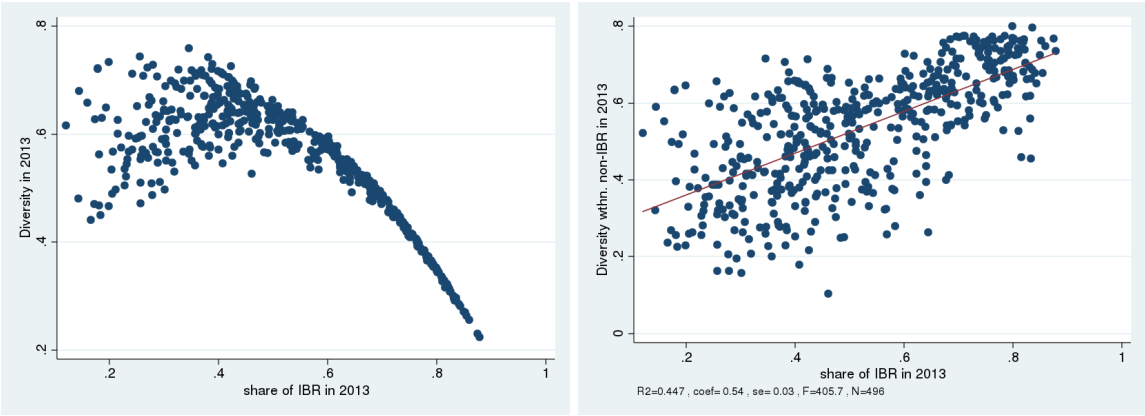
Following the previous literature on the economic effects of cultural diversity, we choose

the fractionalization index as our primary diversity measure, see Kemeny (2014). The diversity value in each municipality k is defined as 1 minus the Herfindahl index, i.e.,

$$diversity_k = 1 - \sum_c s_{ck}^2 \tag{1}$$

where s_{ck} is the share of an ancestry group c in region k and we sum across all ancestry groups $c \in IBR, ITA, GER, EAS, JPN, MXD, BLK, NAT$, cf. table 1. Our intention was to treat each of these groups equally but the properties of the Herfindahl, i.e., fractionalization index become undesirable when one group is disproportionately larger then the others. The left graph in figure 2 reveals that for values of the IBR-share above 0.6, there is almost a 1:1 mapping between the IBR-share and this "naive" fractionalization index. This mechanical relation between the two variables is unfortunate because in the interpretation of our results we will not be able to indicate whether a possibly positive effect of diversity stems from the fact that the share of Iberian descendants is low or indeed stems from the diversity between all ancestry groups. Therefore, we disentangle the diversity variable defined in eq. (1) into the share of Iberian descendants and the diversity among the remainder groups (*non-IBR* henceforth), defined analog to eq. (1). As the second graph in figure 2 shows, these two decomposed diversity measures are also highly correlated but this time the correlation is not mechanical but due to empirical circumstances. Thus it seems that in regions where the share of Iberian descendants is high, the diversity between the remaining cultural groups is also larger. Vice versa, this regularity suggests that in municipalities where the IBR share is below average, one of the immigrant groups dominates over the others, leading to low diversity values. This observation makes sense considering that location patterns are quite stable over time and that the immigrants in the early 20^{th} century spatially clustered together, as is still the case today, cf. the literature on migration networks following Bartel (1989).

Figure 2: Decomposing the naive diversity measure

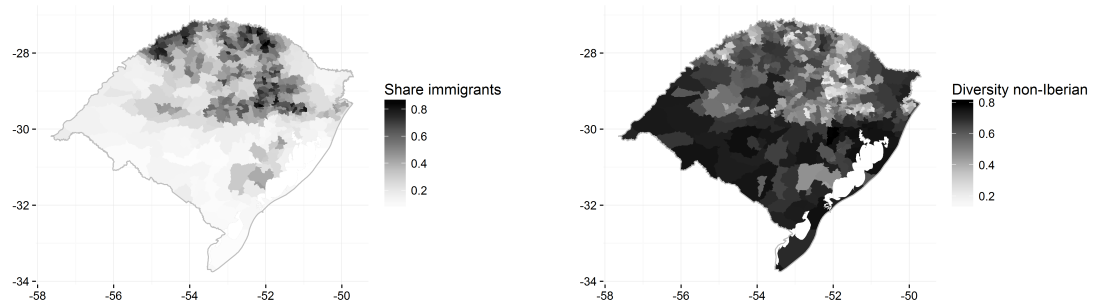


Notes: Both scatter plots are at the current municipality-level. The results from the corresponding linear regression are reported below the right graph.

Both diversity components, the IBR-share and the non-IBR diversity, are our main vari-

ables of interest. Due to their high correlation (0.72) and the consequent multicollinearity concerns, we run separate specifications with only either of the two variables. The high correlation can also be inferred from figure 3 that displays the spatial variation of the two diversity components.[13] It is especially interesting to consider these maps in conjunction with the location of the official colonies in figure 1. The left graph clearly shows that the share of immigrants' descendants in 2013 is higher (darker tones) in the vicinity of the official colonies. Conversely, the municipalities near the southern border, which were used for cattle raising prior to the immigration waves, present the lowest shares of immigrants' descendants. The right graph confirms that the diversity between non-Iberian descendants is much higher in areas that are close to the official colonies.

Figure 3: Spatial variation of the diversity components



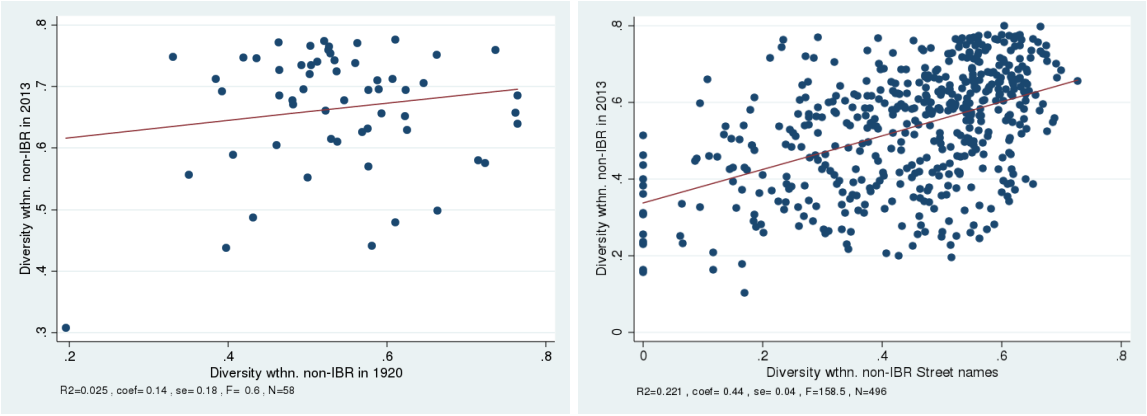
Notes: The left map indicates the value of the diversity index between non-Iberian descendants for each of the current municipalities in Rio Grade do Sul. The map on the right illustrates the spatial variation of the share of immigrants (= 1− the IBR share).

The instruments introduced in section 3 have different usefulness for the diversity variables. We have at least one or two instruments available for each of the three diversity variables. The relations for the two main diversity variables are illustrated in figures 4 and 5. As expected, the cultural diversity in 1920 and the inverse of the mean distance to the official colonies are good predictors for the current ancestral diversity in Rio Grade do Sul. This observation clearly demonstrates the persistence of location patterns and that the installation of the official settlements still has an effect on the spatial distribution of immigrants' descendants. While the diversity between non-Brazilians in 1920 is not really a relevant indicator for the diversity between non-Iberian descendants in 2013, the street name diversity performs much better. In fact, the right graph in figure 4 shows that both variables have a high and positive correlation. The F-value and R^2 of the corresponding simple regression are the highest at the level of municipalities. Recall that due to the different delineation of municipalities in 1920, the variables based on the Census in 1920 only have variation at the level of Minimum Comparable Areas (AMCs), of which there are 58. Figure 5 indicates that both available instruments for the IBR-share are highly relevant. Even though the share of Brazilians citizens was already quite high in 1920, a

[13] The right graph, seems to be a color reversion of the left one. Indeed if we showed the share of Iberian descendants instead of the non-Iberian share, one would have difficulty to distinguish the two graphs in figure 3 from each other.

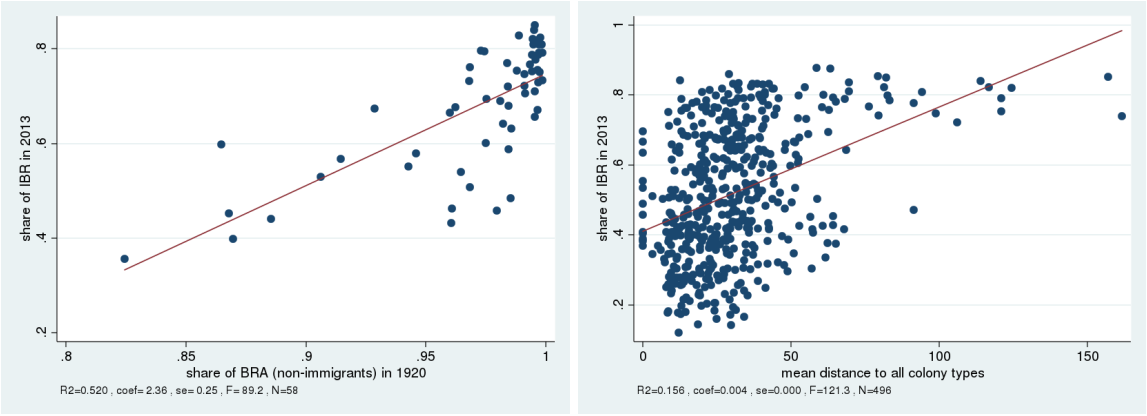
larger concentration of foreigners in the past still predicts which regions have a higher share of non-IBR ancestry today.[14] Moreover, those regions that are farthest away from the official colonies present higher shares of Iberian descendants today.[15]

Figure 4: Instruments for diversity between non-Iberian descendants



Notes: The left graph is at the AMC-level and the right graph is at the current municipality-level. The results from the corresponding linear regression are reported below each graph.

Figure 5: Instruments for share of Iberian descendants



Notes: The left graph is at the AMC-level and the right graph is at the current municipality-level. The results from the corresponding linear regression are reported below each graph.

5 Empirical Strategy

The purpose of the paper is to assess the wage effects of the composition of individuals with different ancestral background in the labor market. To this end, we run wage regressions on

[14] In 1920, the share of foreigners in Rio Grande do Sul was around 7%. The fall from the 12% peak observed in 1900 is explained in part by the softening of legal restrictions to naturalization during the Republic (Seyferth 1997, 2000, 2002).

[15] Some values are apparent outliers in the right graph of figure 5 but even without those observations with mean distances above 100, the regression indicates a significant correlation and an F-value of 88.

region-specific measures of ancestral diversity controlling for a maximum of worker-specific wage determinants. As a first approximation, we estimate the following regression by OLS

$$w_i = \beta_0 + \beta \cdot diversity_k + \cdot Z_i\gamma + \varsigma_s + \epsilon_i \tag{2}$$

where w_i is the log hourly wage of individual i, ς_s are sector fixed effects and ϵ_i is the individual error term. The vector Z_i contains the following individual wage determinants: gender, age, age squared and dummies for formal education level, race/color, illiteracy and deficiency. Our main interest in eq. (2) is the coefficient β, the effect of ancestral diversity in municipality k. The share of Iberian descendants ($IBR\text{-}share$) and the diversity between non-Iberian descendants ($div.\ non\text{-}IBR$) are alternatively used instead of the common diversity/fractionalization index defined in 1. As explained in the previous section, the two latter variables allow for a more detailed decomposition of the diversity effect.

The coefficient of each of these three diversity variables of interest may suffer from endogeneity. At the heart of the endogeneity concern is the general fact that more productive regions are more attractive than others. However, even the direction of the possible bias is not entirely predictable because it depends on the mobility of different population groups, both at present and in the past. Three different consideration are important: First, if descendants of non-Iberian immigrants are more mobile *nowadays*, then we would expect that members of this group are particularly attracted to more productive cities. In this case, the estimate of ancestral diversity would be biased upward and the IBR-share coefficient would be biased downwards. If, vice versa, Iberian descendants are more mobile or for whatever reason more attracted to productive cities, the direction of the biases would be opposite. Yet we are not aware of studies that provide a connection between peoples ancestry and their intra-national mobility. Second, regarding the migration of non-Iberian immigrants *in the past*, it also depends if they were rather attracted to rural or urban regions and whether these location patters were stable over time. The latter statement about persistence is easy to confirm. Thus OLS estimation of eq. (2) generates a downward bias in the diversity measure and an upward bias in the IBR-share if the immigrants in the early 20^{th} century were rather attracted to rural regions where they could engage in agriculture. In any case, it is unlikely that current regional productivity advantages were already foreseeable because the industrial revolution and the technological changes were still in its infancy in Brazil in 1920 (Ehrl and Monasterio 2016). Finally, if both Iberian and non-Iberian descendants are equally mobile and supposed that both groups are equally attracted by productive regions, the fact that the majority of Brazilians are Iberian descendants implies that their share is also be higher in attractive regions, whereas the diversity between non-Iberian descendants would be unaffected. This constellation also results in an upward bias in the IBR-share. Based on these arguments we have reason to believe that this direction of bias is the empirically more likely case.

Note that the endogeneity bias is more severe in studies that rely on the diversity of current immigrants because this group of people is clearly more mobile than the natives and the former thus have a larger share in urban regions compared to rural regions. Following

Ottaviano and Peri (2006), most of the preceding papers on this topic employ lagged diversity values and past migration flows to construct instrumental variables. We also resort to this estimation technique to tackle the endogeneity bias but in contrast we use the long-lagged instrumental variables described in section 3: cultural diversity in 1920, the share of Brazilian citizens in 1920, the diversity of street names in each municipality and the mean distance to historic official colony settlements.

The relevance of each of these instruments was already discussed in the previous section. Their exogeneity assumptions differ slightly from each other and thus will be explained one by one. As argued in section 2.2, the installation of the official colonies was based on strategic and military reasons, hence for our purposes can be considered random and unrelated to possible inherent productivity advantages of the specific municipality or its neighbor regions. Consequently, the mean distance to the closest of each of the three colony types (German, Italian or East European) is exogenous to current wage and productivity determinants of the municipality, even to those that might be time-invariant.

The validity of the cultural diversity and the share of Brazilian citizens in 1920 as IVs requires that after the initial distribution of immigrants, their subsequent relocation, i.e., migration within the state of Rio Grande do Sul until 1920 was not based on any factor related to the remuneration within the region that persists over the period 1920–2013. Recall that the industrial structure of Brazil has changed substantially so that any of the Marshallian productivity externalities in its current form should be out of question. Still, some first-nature advantages that have and still do facilitate the production, such as access to the sea, persist. Without consideration, these factors compromise the exogeneity of these two instruments from 1920. To mitigate this concern, we control for the latitude and longitude of municipalities.

Variation in the street names across municipalities depends on the different people who lived in or were considered influential for a respective municipality. In particular, we expect that cities with a higher cultural diversity in the past have street names that reflect these different cultural backgrounds. The choice to name/re-name a certain street after a person is taken by the municipality council and therefore depends on a variety of unpredictable factors, which however should not show any relation to persistent wage determinants except for the (persistent) cultural diversity or other intrinsic differences between natives and immigrants. Because we have more instruments than endogenous variables, it is possible to assess their exogeneity by Hansen-Sargan overidentification tests.

A different but also potentially severe problem stems from the fact that we cannot observe workers' intrinsic motivation and their inherent abilities. Suppose that the error term in eq. (2) is composed of an individual-specific component μ_i and a random, time-varying component ε_{it}. If people with either Iberian or non-Iberian ancestry had systematically different values of μ_i, then the coefficients of our diversity variables would also be biased. Immigrants' descendants may have a higher level of human capital than the average Brazilian for a number of reasons. Immigrants had a higher level of human capital (Rocha *et al.* forthcoming), higher income, entrepreneurial skills, etc. which may be inherited by fu-

ture generations. It is also plausible that people with non-Iberian ancestry have a higher propensity to learn a different language or have international experience makes their labor supply more valuable for firms. Another source of bias comes from the consolidated observation that, in general, people with high unobservable abilities have a preference for living in more diverse regions. As Ottaviano and Peri (2006: 39) note: "people with higher education, higher international experience, and higher exposure to culture and news may be more appreciative of diversity". This endogeneity problem is not resolved by the instrumental variable estimation to the extent that due to the stability of settlements, the high-diversity regions were already more culturally diverse in the past.

To control for the spatial sorting of workers, we exploit the panel dimension of our micro data and add an auxiliary estimation step to the analysis where the unobserved abilities are accounted for by an individual fixed effect μ_i. Additionally, a municipality fixed effect κ_k captures all of the remaining wage variation between regions. The main idea is that the estimated coefficients $\hat{\kappa}_k$ serve as the dependent variable in a subsequent estimation step to identify the wage effect of regional ancestral diversity. This multi-step analysis of the role of spatial sorting is developed by Combes *et al.* (2008). An advantage is that the estimation can be easily combined with an instrumental variable approach without stressing the computational power of a combined model (Bakens *et al.* 2013). Second, our long-lagged instruments have only cross-sectional variation, so we cannot fully exploit them in the panel estimation. By including and analyzing the estimated municipality wage effects κ_k, we get right back to a cross-section estimation at the municipality level. The first step of our procedure becomes

$$w_{it} = \beta_{0a} + \cdot X_{it}\delta + \varsigma_s + \tau_t + \kappa_k + \mu_i + \varepsilon_{it} \tag{3}$$

where, in contrast to eq. (2), the estimation also includes year effects τ_t and the new vector of individual control variables X_{it} only contains the time-varying characteristics age, tenure and their respective squares. Other time-invariant characteristics like race or education are already accounted for by the fixed effect μ_i.

The subsequent estimation steps exploits only the wage differentials between regions net of the spatial sorting and individual wage determinants to identify the pure effect of ancestral diversity

$$\hat{\kappa}_k = \beta_{0b} + \theta \cdot diversity_k + \xi_k \tag{4}$$

Again, we also use the share of Iberian descendants and the diversity between people of non-Iberian origin in place of the naive aggregate diversity measure. Equation (4) will be estimated by 2SLS using the same instruments introduced above, that is, in total we perform a three-step estimation procedure. Because the dependent variable in eq. (4) is itself an estimate from eq. (3), bootstrapped standard errors are calculated here.

6 Results

6.1 Basic OLS results

Table 2 contains the results from the estimation of eq. (2) by OLS using all individuals in the prepared RAIS data set from the most recent available year 2013. All individual controls are significant at the 1% level and have the expected sign and magnitude in each of the regressions. For brevity we omit these coefficients as well as the sector dummies and report only the regional diversity estimates. The estimated coefficient of the naive diversity measure in the first column is positive but close to zero and insignificant. Recall that this diversity index – by construction – contains two effects which may potentially offset each other: (a) the effect of the share of workers with Iberian ancestry (by far the largest group in the population) and (b) the diversity of the remaining ancestry groups. These variables are used in separate regressions and reported in columns two and three. At this first glance it seems that the diversity between non-Iberian descendants has a positive effect on wages in the region, while the effect of the share of Iberian descendants itself is also insignificant. Cautioned by the descriptive statistics of these two variables, the last column reports their behavior in a joint regression. Both variables are now highly significant. The coefficient of the diversity of non-Iberian workers almost doubled and the IBR share has changed the sign. First of all, these abrupt changes in column (4) confirm our multicollinearity concern. Taken at face value ,the latter suggests that a higher participation of non-Iberian immigrant descendants has a positive impact on the wage level of the region. Hence, a more diverse but balanced cultural background would be potentially beneficial for all of the workers nearby.

Table 2: OLS wage regressions – RAIS 2013

	Dependent variable: log(wage)			
	(1)	(2)	(3)	(4)
diversity	0.081			
	(0.092)			
IBR share		0.117		−0.345***
		(0.117)		(0.133)
div. non-IBR			0.363**	0.691***
			(0.163)	(0.204)
R^2	0.476	0.476	0.478	0.479

Notes: Control variables used in each regression but not shown in the table are age, age squared, tenure, tenure squared and dummies for gender, sector, race, education and whether or not the person is illiterate or has a physical deficiency. Standard errors in brackets are clustered at the municipality-level. The number of observations in each regression is equal to 1,293,127. * denotes significance at ten, ** at five and *** at one percent level.

For comparison of these preliminary results with the previous literature, some comments are in order. All but one of the previous studies on cultural diversity cited throughout this paper assess the effect of diversity among foreigners (or foreign born) on the wages on natives. The correspondent of natives in our case are workers with an Iberian ancestry.

Vice versa, the groups of foreigners correspond to the groups of non-Iberian descendants. Despite this difference caused by the substitution of actual immigrants by the offspring of immigrants, the target group of the effect in our case are all citizens and not only native workers. Bakens *et al.* (2013) also use the average wages of all workers as their dependent variable and construct the diversity index including native and all different cultures, without making the distinction between the share and the pure diversity effect. Ottaviano and Peri (2006: 25) report a high and significant coefficient of foreign diversity (above 0.9) on the local wages of natives. The results are closer to those in table 2 when the authors run separate regressions on the two components of diversity – the share of foreign born and (0.57**) the diversity index among foreign born (0.14*). However, the coefficients of all diversity variables diminish and become insignificant when regional dummies are added to the regression. In a study on Germany, Südekum *et al.* (2014) find quite similar results. The large reactions in the estimations after the inclusion of region fixed effects is an indicator for the existence of an omitted variable bias.

6.2 Accounting for the endogeneity bias

To tackle the endogeneity of the diversity variables we estimate 2SLS regressions according to eq. (2). Table 3 shows that the results are essentially different from the OLS estimates. In column one, the composed diversity index is instrumented by the values of the diversity index in 1920 and by the mean distance to the nearest of each of the colony types. The naive diversity index is now positively and significantly associated with the local wage level. When only one of the two instrumental variables is used, the outcomes are quite similar and thus omitted for brevity. Besides the abbreviation of the instrumental variables in use, the bottom of the table contains further details of the first-stage estimations. Column one indicates that the two instruments explain 32% of the variation of the ancestral diversity in 2013. The F-value in this regression indicates that the instruments cannot be classified as weak and the Hansen-Sargan statistic clearly confirms the exogeneity of both IVs.

In columns two to four, the share of Iberian workers is instrumented. Its coefficient is now clearly negative but we only obtain significance when the mean distance to the nearest colony is used as IV. For the overidentified estimation with both instruments the coefficient is again significant at the 1% level and its size is between the two previous estimates. The insignificance of the coefficient in column two may be caused by a weak relation to the historic instrument. The first-stage statistics reveal that the explanatory power of the instrument is larger for the distance to colony IV, but even for the local share of Brazilians in 1920, the F-value is well above the rule-of-thumb threshold of 10. For reasons explained above, we are more confident about the exogeneity of the installation of the official colonies and therefore to the second instrument in column three. Nevertheless, the Hansen-Sargan test statistics indicates that both IVs are exogenous as well. The result in the last column where both the non-IBR diversity and the IBR-share are instrumented again shows that in our case the use of both variables at the same time is not advisable. Compared to the estimates in the previous columns, one notices a switch in the sign of the non-IBR

diversity and a jump in the strength of both variables, all of which are indicators for multicollinearity. In this case, the overidentification test statistic casts some doubt on the street name instrument since both of the other IVs are already classified as exogenous in columns one and four. Consequently, the result in column (3) is our best estimate for the effect of diversity on wages, thus far.

Table 3: 2SLS wage regressions – RAIS 2013

	Dependent variable: log(wage)					
	(1)	(2)	(3)	(4)	(5)	(6)
diversity	0.940***					
	(0.315)					
IBR share		-0.633	-0.969***	-0.896***		-1.353**
		(0.466)	(0.228)	(0.242)		(0.575)
div. non-IBR					-0.144	1.011
					(0.438)	(0.745)
	1.-stage statistics					
IVs:	div. (1920), colony dist.	BRA-sh. (1920)	colony dist.	BRA-sh. (1920), colony dist.	street div.	street div., BRA-sh. (1920), colony dist.
1. R^2-part.	0.323	0.280	0.175	0.317	0.368	0.386
1. F-stat	36.59	17.95	37.49	53.57	41.32	9.572
2. R^2-part.						0.456
2. F-stat						14.32
Hansen J	0.382			1.504		4.397
Hansen J-p	0.536			0.220		0.036

Notes: All regressions include the following control variables: age, age squared, tenure, tenure squared and dummies for gender, sector, race, education and whether or not the person is illiterate or has a physical deficiency. Standard errors in brackets are clustered at the AMC-level or the municipality-level, depending on the instruments used. Whenever one of the IVs from 1920 is used, AMC clustering is required. The number of observations is equal to 1,293,127 in all regressions. * denotes significance at ten, ** at five and *** at one percent level.

Hence, our main finding in table 3 is that an increase in the share of workers with a non-Iberian cultural background by 10 percentage points is associated with a 9.7% wage increase in the local labor market. In contrast, the balance, i.e., the diversity between the workers with non-Iberian ancestry does not seem to have any importance. The difference between the estimated coefficients in columns four and five and the respective OLS estimates suggest that there is an substantial endogeneity problem that results in an upward bias in the coefficient of both the Iberian share and the non-Iberian diversity variable, as well as a downward bias in the naive diversity index. According to the theoretical considerations in section 5, these biases suggest that either (1) people with Iberian ancestry are more or at least equally mobile nowaday than non-Iberian descendants and at the same time attracted by productive regions or (2) the non-Iberian immigrants in the 19^{th} and 20^{th} century were rather attracted to rural regions which are nowadays less productive.

Originally we performed the same analysis using micro-data from the Brazilian Census in 2010. Only in order to implement the last part of our analysis, which requires a panel data set, we abandoned the Census data and performed the entire analysis based on the RAIS data from 2008–2013. Nevertheless, our previous calculations can be seen as a robustness check of the reported results and we thus find it worthwhile to mention some comparisons. First, the Census includes all individuals, which is interesting because unlike

the RAIS, it also includes the remuneration of informal workers, self-employed individuals and employers. Second, the composition of workers is also different in both data sets because the RAIS contains the universe of formally employed workers, whereas the Census micro-data is a 10% random sample of the population.[16] It is reassuring that, despite the different composition, both data sets yield similar results. For example, for the coefficient of the Iberian share of workers reported in column 4 in table 3, we obtain a point estimate of -0.956 from the Census 2010 data. When the Census sample is restricted to formally employed workers to maximize comparability, the estimated coefficient is equal to -0.949. Both estimates are also significant at the 1% level. This comparison shows that our main result is largely unaffected if another similar data set is used.

6.3 Accounting for intrinsic abilities and spatial sorting

For the next part of the study we build a panel around those individuals that are analyzed thus far. The purpose of this extension is to control for unobserved individual abilities that distort the previous estimates if they are correlated with the diversity measures. In particular, there is reason to be worried about the spatial sorting of skilled workers based on this source of heterogeneity. Table 4 contains the results of the estimations according to eq. (3) and eq. (4). Apart from the individual fixed effect, these estimations include sector effects, those worker-specific controls variables that are time-variant and municipality fixed effects, whose estimates represent net wage differentials between municipalities. These estimates are analyzed subsequentially in a separate regression. Since the purpose of this first step regression is to account for the spatial sorting of workers but without solving the endogeneity problem, we report only the second-step instrumental variable regressions that use the estimated municipality fixed effects as the dependent variable.

The coefficients in table 4 essentially show the same pattern of the previous results in table 3 except for the fact that the magnitude of the effects is much lower. Despite the lower coefficient, their significance even tends to increase. The naive diversity measure and especially the share of non-Iberian workers have a positive effect on local wages. These estimates tell us that an increase by one standard deviation or 8,4% in diversity is associated with wage increments of 2.6%, ceteris paribus. On the other hand, an increase by one standard deviation or 11 percentage points in the non-Iberian share should increase wages by almost the same size. In other words, if the average municipality with a share of 58% of Iberian descendants experienced an increase in workers with non-Iberian ancestry such that the Iberian share drops to 47%, average wages should benefit from the increased cultural diversity by 2.5%. Only the diversity index calculated from the non-Iberian descendants continues to be without statistical significance. As before, the instruments show the desired statistical properties and work quite well.

[16] The preparation of the RAIS and Census data is as similar as possible and as described in section 3. Yet the classification of education and sectors in the RAIS is slightly different than the ones used in the Census regressions and the regression also differs because there is no information about job tenure in the Census.

Table 4: 3-step wage regressions – Baseline – RAIS 2008–2013

	Dependent variable: estimated municipality fixed effect					
	(1)	(2)	(3)	(4)	(5)	(6)
diversity	0.311***					
	(0.066)					
IBR share		-0.149***	-0.223***	-0.177***		-0.276**
		(0.039)	(0.056)	(0.035)		(0.129)
div. non-IBR					-0.053	0.168
					(0.062)	(0.169)
	1.-stage statistics from the 2SLS estimation					
IVs:	div. (1920), colony dist.	BRA-sh. (1920)	colony dist.	BRA-sh. (1920), colony dist.	street div.	street div., BRA-sh. (1920), colony dist.
1. R^2-part.	0.220	0.203	0.156	0.253	0.221	0.298
1. F-stat	70.20	149.3	121.3	116.6	158.5	94.33
2. R^2-part.						0.308
2. F-stat						106.1
Hansen J	1.864			2.607		3.996
Hansen J-p	0.172			0.106		0.046

Notes: The table reports the coefficients from an IV estimation that uses the estimated municipality fixed effect coefficient as the dependent variable, which itself was obtained from a first-step wage regression with individual, time and sector fixed effects as well as age, tenure and their respective square products. Therefore, standard errors in brackets are obtained from 100 bootstrap repetitions. The number of observation is equal to the number of municipalities (496) in all regressions. * denotes significance at ten, ** at five and *** at one percent level.

The substantial reduction of the estimates confirms that workers sort according to unobservable productive characteristics. In particular, those regions with a larger share of non-Iberian descendants, i.e., the more diverse regions attract workers with higher abilities. Despite this correction of the diversity effect's magnitude, our primary finding remains robust. Cultural diversity is persistent and continues to impose a positive externality on wages. Moreover, this effect is essentially driven by the share of those individuals who are not part of the majority group, i.e., non-Iberians. The last part also revealed that labor markets with a more diverse cultural background are more productive because the individuals posses better abilities. Thus, as an extension to the results of Combes *et al.* (2008), we find that the size of the region is not the only characteristics on which spatial sorting is based on.

6.4 Robustness and extensions

In this section, we build on our main results in tables 3 and 4 and extend both the 2SLS and the 3-step regression procedure in several dimensions.

The 25 largest cities. First, we are interested how the strength of the diversity effect changes according to the size of the municipality. Bakens *et al.* (2013) suggest that there is a heterogeneity of the effect in this dimension. Specifically, the authors find that – contrary to common expectation – the diversity effect is absent in the 25 largest cities in The Netherlands. Obviously, this extension reduces the variation of our main variables considerably (to only 25 degrees of freedom). Since our main estimations are based on

instrumental variable regressions and since some of them only have variation at the level of AMCs, i.e., a combination of municipalities, it is not possible to use them in this extension. For the diversity in 1920 and the share of Brazilians in 1920 there are less AMCs then covariates and thus the regression is underidentified. Consequentially, there is only one IV left for each endogenous variable and the analysis is now restricted to the two distinct diversity components. As in the previous estimations, the diversity between non-Iberian ancestry groups is insignificant. Yet the positive wage externality of the share of non-Iberian descendants is much higher than before and continues to be highly relevant in both specifications. The increase to about twice of the size of the estimates from the full sample of regions is certainly also due to the reduced imprecision in the small sample. Nevertheless, the results in table 5 are highly suggestive of a stronger cultural diversity effect in the largest cities.

Table 5: 2SLS and 3-step wage regressions – the 25 largest municipalities

Dep. var.:	log(wage)		estimated municipality fixed effect	
	(1)	(2)	(3)	(4)
IBR share	-1.492***		-0.637***	
	(0.397)		(0.219)	
div. non-IBR		-0.659		-0.413
		(0.442)		(1.020)
Obs.	760,248	760,248	25	25
			1.-stage statistics	
IVs:	colony dist.	street div.	colony dist.	street div.
1. R^2-part.	0.238	0.548	0.337	0.573
1. F-stat	11.25	43.83	18.98	16.41

Notes: The individual-level 2SLS wage regressions in columns (1) and (2) control for age, age squared, tenure, tenure squared and dummies for gender, sector, race, education and whether or not the person is illiterate or has a physical deficiency, cf. table 3. Columns (3) and (4) report the 3-step regressions where the dependent variable is the estimated municipality fixed effect coefficient from a first step wage regression with individual, time and sector fixed effects as well as age, tenure and their respective square products, cf. table 4. Standard errors in brackets are clustered at the municipality-level in columns (1) and (2), and obtained from 100 bootstrap repetitions in columns (3) and (4). * denotes significance at ten, ** at five and *** at one percent level.

Excluding internal migrants. Second, we exploit the panel dimension of of individual data once again to identify those people who have worked in a different Federal State within the last 5 years, i.e., at some date between 2008 and 2012. Recall that our estimations all refer to the year 2013. Independent of their place of birth or whether those individuals have worked in Rio Grande do Sul (our state of interest) before, they are classified as internal migrants. Similar to foreign immigrants, internal migrants also may bring different attitudes, skills, etc. with them and thus contribute to the cultural diversity in Rio Grande do Sul. In the present extension, those internal migrants are excluded from the sample. We then re-calculate the diversity measures and repeat the main estimations in order

to isolate the pure effect of cultural diversity that stems from having different ancestral backgrounds. 92,441 individuals corresponding to roughly 7% of the original sample are internal migrants.

Table 6: 2SLS and 3-step wage regressions – No internal migration

Dep. var.:	log(wage)		estimated municipality fixed effect	
	(1)	(2)	(3)	(4)
share IBR	-0.698***		-0.233***	
	(0.194)		(0.055)	
div. non-IBR		-0.145		-0.069
		(0.277)		(0.059)
Obs.	1,200,686	1,200,686	496	496
			1.-stage statistics	
IVs:	colony dist.	street div.	colony dist.	street div.
1. F-stat	33.79	26.7	??	
1. R2-part.	0.150	0.348		??

Notes: Regressions where the dependent variable is the estimated municipality fixed effect coefficient from a first stage wage regression with individual, time and sector fixed effects as well as age, tenure and their respective square products. Therefore, standard errors in brackets are obtained from 100 bootstrap repetitions. The number of observation is equal to the number of municipalities (496) in all regressions. * denotes significance at ten, ** at five and *** at one percent level.

Again, we repeat only the most interesting estimations using the new sample without internal migrants. Table 6 shows that the 2SLS regressions, without accounting for the individual fixed effects, are qualitatively similar to those in table 3. The diversity within non-Iberian descendants still has an insignificant effect on wages. On the other hand, a higher share of non-Iberian workers continues to have a positive impact on the local wage level but it is now substantially lower. Hence, individuals who have worked in a different Federal State seem to be intrinsically different from workers without such employment record. At least two interpretations for this finding are possible. Either the work experience in a different state represents an asset that is comparable to having a foreign cultural background. Or internal migrants posses particularly valuable unobservable skills that have biased the estimates upward in the full sample. Those unobservable skills are controlled for in the 3-step procedure. Columns (3) and (4) show precisely that including internal migrants or not, makes hardly any difference for the diversity estimates. This finding leads to the conclusion that migrants have higher abilities than the average individual and that work experience in a different Federal State is not comparable to having inherited a different cultural background. Moreover, table 6 also confirms that the 3-step estimation strategy is indeed capable of eliminating the effect of unobservable human capital and sorting from the pure effect of inherited cultural diversity.

The diffusion of technology. The following robustness check serves to discover more about the mechanism at work. According to Spolaore and Wacziarg (2009), relative genetic distance created a barrier to the diffusion of knowledge that explains the ability of countries to catch-up to the technological frontier which has been moving outward massively since the Industrial Revolution. Because the Industrial Revolution began in England, it became the technological leader and thus genetic distance is measured relative to the Englishmen. In the present context, their finding implies that Germans, other European immigrants and their descendants found and still find it easier to communicate, exchange, etc. with people in industrialized countries like England, the USA, or their countries of origin in Europe. Due to this facilitated diffusion of knowledge, regions that host a large share of workers with European ancestry may be more productive.

To bring this idea to the data, we calculate the aggregate value of imports of firms in each municipality in Rio Grande do Sul. Imports may permit access to superior technology, especially when the imports are from high-wage countries and when machines or intermediate products are imported. However, imports increase as a function of the local wage level and the demand faced by local firms. To avoid endogeneity between the imports and the regional wage level, we consider the imports from 2003 to 2006. Since imported technology becomes a durable asset of the firm and since it takes some time for investments to become fully productive, the temporal lag and the consideration of machines and intermediate products should thus avoid the reverse causality without diminishing much the true effect of the diffusion of technology.

The Brazilian Ministry of Development, Industry and International Trade (MDIC) provides data on the imports of Brazilian establishment at the country-product level. For further details about the data and the preparation procedure, see Ehrl (2014). It is well known that importers as well exporting firms are only a small percentage of the total number of firms. Consequently, only 205 of the 496 municipalities in our sample have a positive import value, even over the 4 year period. In line, imports from high-wage countries are even more rare and appear in 160 municipalities only. Nevertheless, the imports in municipalities have a mean equal to 40 million Reais (equivalent to about 18 million USD in 2006) and a standard deviation of 347 million. The region with the highest imports spent 6.6 billion on foreign goods.

Table 7 show the results once we additionally control for the aggregate value of several distinct import measures. To save space, only the results from the preferred 3-step procedure are reported here. Independent of the import measure, the coefficient of diversity between non-Iberian descendants remains insignificant. The remainder columns of the table refer to the share of Iberian workers. Column (1) indicates that the total value of imports in municipalities between 2003 and 2006 as no significant relation to the regional wage level, once all individual and sectoral characteristics are accounted for. If only imports from high-wage countries are considered instead, the coefficient becomes positive and significant.[17] Its value even increases when we restrict the imports to machinery from high-wage

[17] Brazil's main high-wage country trading partners are: Canada, the European Union, Japan and the US.

Table 7: Second step wage regressions by IV – Imports – RAIS 2008–2013

	Dependent variable: estimated municipality fixed effect				
	(1)	(2)	(3)	(4)	(5)
import measure:	all	from high-wage countries (hwc)	machines from hwc	intermediates from hwc	from hwc
IBR share	−0.221***	−0.211***	−0.213***	−0.215***	
	(0.051)	(0.053)	(0.059)	(0.048)	
div. non-IBR					−0.084
					(0.059)
imports	0.000	0.001***	0.002*	0.001***	0.001***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)

Notes: The dependent variable is the estimated municipality fixed effect coefficient from a first stage wage regression with individual, time and sector fixed effects as well as age, tenure and their respective square products. The instruments in each regression are the same as in table 4. This extension includes the municipality aggregate values of several import measures from the period 2003–2006, as indicated below the column numbering. Standard errors in brackets are obtained from 100 bootstrap repetitions. The number of observation is equal to the number of municipalities (496) in all regressions. * denotes significance at ten, ** at five and *** at one percent level.

countries. These estimations indicate that our proxy for imported technology works as expected and that it shows a positive relation to wages. Most important, the effect of inherited cultural diversity, as captured by the share of non-Iberian descendants remain highly stable and significant in all of these specifications. We conclude that even though European descendants may have better access to foreign technology this is not the driving channel of our externality.

General agglomeration economies. After all, there is also reason to suspect that some of the well known Marshallian agglomeration economies are driving or interfering with the identified cultural diversity effect. It is common to capture these effects by the population size of the region (Rosenthal and Strange 2004). Again, we maintain the first-step regression (eq. 3) and include the log population size as an additional control variables in the subsequent 2SLS regressions where the dependent variable is the estimated coefficient of the municipality fixed effect that reflects the remaining wage differences between regions.

Table 8 shows that the log population is indeed a significant factor in the residual wage regressions. Its estimate in column (1) indicates that a doubling of the local population induces wage gains of 4%. The share of Iberian descendants continues to be significant wage determinant and its magnitude is even slightly higher than in the previous estimations. As before, the diversity between the non-Iberian ancestry groups has no significant effect and in the remainder of the paper, these results are omitted for brevity. By the way, we obtain the same significance levels for the estimates in table 8 when standard errors are not calculated by bootstrapping but rather from the usual White-Huber robust errors.

Geography and institutions. Furthermore, we would like to make sure that the location of settlements were not based on first-nature advantages, such as climatic conditions,

Table 8: 3-step wage regressions – population size, geography, institutions

	Dependent variable: estimated municipality fixed effect			
	(1)	(2)	(3)	(4)
IBR share	-0.304***	-0.167**	-0.308***	-0.203**
	(0.068)	(0.074)	(0.076)	(0.083)
log(population)	0.039***			0.024***
	(0.005)			(0.005)
latitude		-0.024***		-0.018**
		(0.007)		(0.007)
longitude		0.021***		0.022***
		(0.003)		(0.003)
governance			-0.004	-0.005
			(0.008)	(0.007)
access to justice			0.033***	0.002
			(0.008)	(0.005)
Observations	496	496	467	467
	1.-stage statistics			
1. F-stat	105.99	40.08	92.12	36.85
1.R^2-part.	0.144	0.082	0.125	0.078

Notes: The dependent variable is the estimated municipality fixed effect coefficient from a first stage wage regression with individual, time and sector fixed effects as well as age, tenure and their respective square products. The instrument used in each regression is the mean distance to the official colonies. Standard errors in brackets are obtained from 100 bootstrap repetitions. * denotes significance at ten, ** at five and *** at one percent level.

access to the sea, etc. The best we can do at this stage is to control for the geographical latitude and longitude of municipalities Including an interaction term of these two variables does not alter the results significantly. Even though Spolaore and Wacziarg (2013) conclude that in the recent literature on cross-country income differences "[t]he bottom line, however, is that human traits are important to account for comparative development patterns, quite apart from the effects of geographic and institutional factors", we prefer to test both of these factors in our current sample. Despite the focus on a single Federal State, some differences in institutions may remain, and this variation may have been caused partly by immigrants. We use data from Naritomi *et al.* (2012) for the access to justice and an indicator on the municipal governance.

Column (2) in table 8 indicates that the geographical control variables are indeed highly significant and that they reduce the coefficient of inherited cultural diversity. Access to justice is also positively related to the current wage level, whereas local governance seems to be less important. Column (4) combines the prior control variables. This horse race specification indicates that institutions are least important. The geographical control variables remain significant and almost unaffected. The elasticity of population size also remains significant but is reduced by half, indicating a 2% wage elasticity. Finally, the impact of the share of Iberians is almost the same as without these control variables, cf. 4. All of these specification clearly demonstrate that the effect of inherited continues to be relevant

over and above the effect of institutions, geography and other agglomeration economies.

Measurement of the diversity indices. Finally, we also vary the definition of the diversity index in order to verify that the choice of one specific definition among several possible does not fundamentally affect the results. Especial interest is in the diversity between the non-Iberian descendants because we have found somewhat different results from the prior literature and because we already know that the naive diversity index is mainly driven by the share of the largest group. In regional science, the question whether industrial diversity is responsible for agglomeration economies has also produced several diversity indices. Following Henderson (2003), we calculate the difference between the share of each ancestry group in each municipality and the share of this ancestry group in the entire state, and then calculate the sum the squares of these differences across all ancestry groups. In fact, this is an inverse diversity index but in contrast to the fractionalization index, it is a relative measure that takes the size of each ancestry group explicitly into consideration. Ehrl (2013) argues that the absolute value of the differences (instead of using the squares) makes the index less vulnerable to outliers. Both of the new versions of the diversity index between non-Iberian descendants are tested in our preferred three-step regression approach analog to those in table 4 and in simple 2SLS estimations analog to table 3. All four estimations essentially replicates our prior findings: The composition of the non-Iberian ancestry groups does not have an independent effect on wages.

7 Conclusion

The present paper revealed that ancestry is an important dimension of cultural diversity. In particular, the composition of workers' cultural backgrounds in localized labor markets is an important determinant of wage levels. Those benefits from the diversity of the workforce are especially high in the largest cities. A decomposition of the frequently used fractionalization index of cultural diversity shows that the main effect operates via the share of the workers with a non-Iberian ancestry, i.e., those workers with a cultural background different from the majority population group. The size of this positive externality is of economic relevance, and these conclusions hold in a variety of different specifications that address reverse causality, omitted variable bias, spatial self-selection of workers, different agglomeration economies, first-nature advantages and issues related to the measurement of the diversity index.

Empirical literature on immigration shows that newcomers provide a different set of skills and this has economic value. Our results go one step further. We have found evidence that the impact of immigration can last for generations. People in locations with more diverse surname based ancestry earn higher wages, with full controls, despite the fact that arrivals from non-Iberian countries have ceased almost a century ago. In other words, our results provide support to the thesis that immigration can create intertemporal externalities, and not only local, contemporary effects. Those positive impacts are usually not taken into account in the public debate on immigration policies.

References

Ager, P. and Brückner, M. (2013). Cultural diversity and economic growth: Evidence from the US during the age of mass migration. *European Economic Review*, **64**, 76–97.

Alesina, A. and Ferrara, E. L. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature*, **43** (3), 762–800.

Amstad, T. (1999). *Cem Anos de Germanidade no Rio Grande do Sul–1824/1924*. São Leopoldo: Unisinos.

Andrews, G. R. (1988). Black and White Workers: Sao Paulo, Brazil, 1888-1928. *The Hispanic American Historical Review*, **68** (3), 491–524.

Bakens, J., Mulder, P. and Nijkamp, P. (2013). Economic impacts of cultural diversity in the Netherlands: Productivity, utility, and sorting. *Journal of Regional Science*, **53** (1), 8–36.

Bartel, A. P. (1989). Where do the new US immigrants live? *Journal of Labor Economics*, **7** (4), 371–391.

Bisin, A. and Verdier, T. (2000). "beyond the melting pot": cultural transmission, marriage, and the evolution of ethnic and religious traits. *Quarterly Journal of Economics*, **115** (3), 955–988.

Borjas, G. J. (1995). Ethnicity, neighborhoods, and human-capital externalities. *American Economic Review*, **85** (3), 365–90.

Boyd, R. and Richerson, P. J. (1988). *Culture and the evolutionary process*. University of Chicago Press.

Carneiro, P., Lee, S., Reis, H. and others (2015). *Please call me John: name choice and the assimilation of immigrants in the United States, 1900-1930*. Tech. rep., Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Cavnar, W. B. and Trenkle, J. M. (1994). N-gram-based text categorization. *Ann Arbor MI*, **48113** (2), 161–175.

Clark, G. (2015). *The Son Also Rises: Surnames and the History of Social Mobility*. Princeton University Press, reprint edition edn.

—, Cummins, N., Hao, Y. and Vidal, D. D. (2015). Surnames: A new source for the history of social mobility. *Explorations in Economic History*, **55**, 3–24.

Combes, P., Duranton, G. and Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, **63** (2), 723–742.

Dancygier, R. M. (2014). Electoral Rules or Electoral Leverage? Explaining Muslim Representation in England. *World Politics*, **66** (02), 229–263.

de Carvalho Filho, I. and Colistete, R. P. (2010). Education Performance: Was It All Determined 100 Years Ago? Evidence From São Paulo, Brazil.

— and Monasterio, L. M. (2012). Immigration and the origins of regional inequality: Government-sponsored European migration to Southern Brazil before World War I. *Regional Science and Urban Economics*, **42** (5), 794–807.

Dohmen, T., Falk, A., Huffman, D. and Sunde, U. (2012). The intergenerational transmission of risk and trust attitudes. *Review of Economic Studies*, **79** (2), 645–677.

Ehrl, P. (2013). Agglomeration economies with consistent productivity estimates. *Regional Science and Urban Economics*, **43** (5), 751–763.

— (2014). Task trade and the employment pattern: the offshoring and onshoring of brazilian firms. *BGPE Discussion Paper*, **151**.

— (2017). Minimum Comparable Areas for the period 1872–2010: An aggregation of Brazilian municipalities. *Estudos Econômicos*, **47** (1), 215–229.

— and Monasterio, L. M. (2016). Historical trades, skills and agglomeration economies. *MPRA Discussion Paper*, **69829**.

Gertz, R. E. (1991). *O perigo alemão*, vol. 5. Editora da Universidade, Universidade Federal do Rio Grande do Sul.

Güell, M., Mora, J. V. R. and Telmer, C. I. (2014). The Informational Content of Surnames, the Evolution of Intergenerational Mobility and Assortative Mating. *The Review of Economic Studies*, p. rdu041.

Henderson, J. V. (2003). Marshall's scale economies. *Journal of Urban Economics*, **53** (1), 1–28.

IBGE (1990). *Estatísticas históricas do Brasil: séries econômicas, demográficas e sociais*. RIo de Janeiro: IBGE.

IBGE (2003). *Estatísticas do século XX*, vol. 1. Ibge.

Iotti, L. H. (2001). *Imigração e colonização: legislação de 1747 a 1915*. Caxias do Sul: EDUCS.

Kemeny, T. (2012). Cultural diversity, institutions, and urban economic performance. *Environment and Planning A*, **44** (9), 2134–2152.

— (2014). Immigrant diversity and economic performance in cities. *International Regional Science Review*, **forthcoming**.

Lago, L. A. C. d. (2014). *Da escravidão ao trabalho livre: Brasil, 1550-1900*. Companhia das Letras, google-Books-ID: DIeNBQAAQBAJ.

Lakha, F., Gorman, D. R. and Mateos, P. (2011). Name analysis to classify populations by ethnicity in public health: validation of Onomap in Scotland. *Public health*, **125** (10), 688–696.

Levy, M. S. (1974). O papel da migração internacional na evolução da população brasileira (1872 a 1972). *Rev. Saúde Pública*, **8** (suppl), 49–90.

Ministério do Trabalho e Emprego (2016). Relação Anual de Informações Sociais de 2013.

Monasterio, L. (2017). Surnames and ancestry in brazil. *PloS ONE*, **12** (5), e0176890.

Musacchio, A., Fritscher, A. M. and Viarengo, M. (2014). Colonial Institutions, Trade Shocks, and the Diffusion of Elementary Education in Brazil, 1889–1930. *The Journal of Economic History*, **74** (3), 730–766.

Naritomi, J., Soares, R. R. and Assunção, J. J. (2012). Institutional development and colonial heritage within brazil. *Journal of Economic History*, **72** (02), 393–422.

Niebuhr, A. (2010). Migration and innovation: does cultural diversity matter for regional R&D activity? *Papers in Regional Science*, **89** (3), 563–585.

Nogueira, O. (1985). *Tanto preto quanto branco: estudos de relações raciais*, vol. 9. Ta Queiroz.

Osório, R. G. (2004). O sistema classificatório de cor ou raça do IBGE. In Joaze Bernardino and Daniela Galdino (eds.), *Levando a raça a sério: ação afirmativa e a universidade*, Rio de Janeiro: Dp&A, LPP, UERJ,, pp. 85–135.

Ottaviano, G. I. P. and Peri, G. (2005). Cities and cultures. *Journal of Urban Economics*, **58** (2), 304–337.

— and — (2006). The economic value of cultural diversity: evidence from US cities. *Journal of Economic Geography*, **6** (1), 9–44.

Petersen, J., Longley, P., Gibin, M., Mateos, P. and Atkinson, P. (2011). Names-based classification of accident and emergency department users. *Health & place*, **17** (5), 1162–1169.

Piza, E. and Rosemberg, F. (1999). Cor nos censos brasileiros. *Revista USP*, **40**, 122–137.

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of Political Economy*, **90** (6), 1257–1278.

Rocha, R., Ferraz, C. and Soares, R. R. (forthcoming). Human capital persistence and development. *American Economic Journal: Applied Economics*.

Roche, J. (1954). Un exemple d'instabilité de la population rurale dans un pays neuf: Les migrations rurales dans le Rio grande do Sul. *Annales. Histoire, Sciences Sociales*, **9** (4), 481–504.

— (1969). *A colonização alemã e o Rio Grande do Sul*. Porto Ale: Editôra Globo, 1st edn.

Rodríguez-Díaz, R., Manni, F. and Blanco-Villegas, M. J. (2015). Footprints of Middle Ages Kingdoms Are Still Visible in the Contemporary Surname Structure of Spain. *PLOS ONE*, **10** (4), e0121472.

Rosenthal, S. S. and Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. In V. Henderson and J. F. Thisse (eds.), *Handbook of Regional and Urban Economics*, vol. 4, Elsevier – North Holland, pp. 2119–2171.

Seyferth, G. (1997). A assimilação dos imigrantes como questão nacional. *Mana*, **3** (1), 95–131.

— (2000). As identidades dos imigrantes e o melting pot nacional. *Horizontes Antropológicos*, **6** (14), 143–176.

— (2002). Colonização, imigração e a questão racial no Brasil. *Revista USP*, **53**.

Spolaore, E. and Wacziarg, R. (2009). The diffusion of development. *Quarterly Journal of Economics*, **124** (2), 469–529.

— and — (2013). How deep are the roots of economic development? *Journal of Economic Literature*, **51** (2), 325–369.

Stolz, Y., Baten, J. and Botelho, T. (2013). Growth effects of nineteenth-century mass migrations: "Fome Zero" for Brazil? *European Review of Economic History*, **17** (1), 95–121.

Südekum, J., Wolf, K. and Blien, U. (2014). Cultural diversity and local labour markets. *Regional Studies*, **48** (1), 173–191.

Susewind, R. (2015). What's in a name? Probabilistic inference of religious community from South Asian names. *Field Methods*, **27** (4), 319–332.

Trax, M., Brunow, S. and Südekum, J. (2015). Cultural diversity and plant-level productivity. *Regional Science and Urban Economics*, **53**, 85–96.

Waibel, L. (1950). European Colonization in Southern Brazil. *Geographical Review*, **40** (4), 529–547.