Migration-prone and migration-averse places.
Path dependence in long-term migration to the US

by

Andrés Rodríguez-Pose and Viola von Berlepsch

Department of Geography and Environment

London School of Economics

e-mail: a.rodriguez-pose@lse.ac.uk; v.k.von-berlepsch@lse.ac.uk

Abstract

Does past migration beget future migration? Do migrants from different backgrounds, origins and ethnicities, and separated by several generations always settle – in a path dependent way – in the same places? Is there a permanent separation between migration-prone and migration-averse areas? This paper examines whether that is the case by looking at the settlement patterns of two very different migration waves to the United States (US), that of Europeans at the end of the 19th and early 20th centuries and that of Latin Americans between the 1960s and the early 21st century. Using Census data aggregated at county level, we track the settlement pattern of migrants and assess the extent to which the first mass migration wave has determined the later settlement pattern of Latin American migrants. The analysis, conducted using ordinary least squares, instrumental variable and panel data estimation techniques, shows that past US migration patterns create a path dependence that has conditioned the geography of future migration waves. Recent Latin American migrants have flocked, once other factors are controlled for, to the same migration prone US counties where their European predecessors settled, in spite of the very different nature of both migration waves and a time gap of three to five generations.

Keywords: Migration, migration waves, long-term, Latin America, Europe, counties, US

JEL Codes: F22, J15, O15, R23

Introduction

The beginning of the 21st century marked a very important shift in the population composition of the United States (US). African Americans were replaced by Hispanics (or Latinos)[1] as the largest ethnic minority. Today, Latinos make 18.3 percent of the US population and their economic, cultural, and political clout is growing rapidly. By 2050, Latinos are expected to represent around 30 percent of the total US population (Bergad and Klein, 2010; U.S. Census Bureau, 2016). Despite a sizable Spanish-speaking community traditionally established in what were former Mexican territories in the South West, the massive migration of Hispanics is a relatively new phenomenon in most parts of the US. It was not until the post-second World War years that Hispanics became the largest migrant group. In earlier migration waves they were only a small fraction of the incoming population. Migration to the US in the 19th and early 20th centuries was dominated by Europeans.

There is no shortage of literature aiming to understand the reasons why people migrate and the factors determining not only the volume but also the direction of migration flows and the final settlement patterns. Much of this literature has revolved around socio-demographic, climatic, and cultural aspects influencing individual migration decisions (i.e. Pissarides and McMaster, 1990; Massey *et al.*, 2005; Partridge and Rickman, 2008; Rodríguez-Pose *et al.*, 2015).

One migration-pull factor that has attracted considerable scientific attention is the presence of kinship and migrant networks at destination. Migrants follow the path previously 'beaten' by their relatives, friends, and co-nationals (Hoover, 1971; DaVanzo, 1983). The presence of a migrant network at destination decreases the costs of moving and raises the potential net gains. But do migrant networks expand beyond this 'beaten-path' dominated by kinship and ethnic and national origins? Do migrants flock to places that have, over time, become 'migration-prone'? And, more importantly, do migrants of different national origins and generations end up in the same places that welcomed historical migration once the diverse factors that make a place attractive to different migrant generations are controlled for? These questions have attracted little attention in migration research, which has predominantly focused on micro-personal ties and networks rather than long-term factors that make particular places migration-prone or migration-averse.

This paper aims to answer these questions, using a dataset comprising county level data of foreign-born population shares during two major migration waves – the Age of Mass Migration at the turn of the 20th century and the period of Latin American immigration to the US from 1950/60 onwards. The objective of this research is to ascertain if completely unrelated migration waves follow similar geographical patterns once the socio-economic factors affecting migration are controlled for. We posit that large numbers of migrants settling in a region at a given point in time leave an imprint on the territory – a 'migration buzz' – which acts as pull factor for ensuing migrant waves, irrespective of ethnicity or origin. It will be argued that the migrants' character, in particular their willingness to take risks, their entrepreneurialism, and their dynamism becomes engraved in particular territories,

---

[1] There are intense debates regarding whether it is more appropriate to use the term Hispanic or Latino (Taylor *et al.*, 2012). In this paper, following Sáenz and Morales (2015), the two terms are used indistinctively to refer to the population originating from Mexico, Spanish-speaking Caribbean islands, and Central and South America.

contributing to the creation of a long-lasting division between migration-prone and migration-averse places.

The paper adopts the following structure. We first summarise the main facts and present descriptive data on the two migration waves in question. Second, we review the relevant literature before presenting a discussion of the model and data. This is followed by the examination of the regression results, using a variety of different estimation techniques. The final section concludes and presents some preliminary policy implications.

US migration patterns – a brief sketch

The history of the US has been shaped by huge inflows of people seeking a better life for themselves and their families. Millions of people from all over the world have moved to the US, transforming the country into a 'nation of immigrants'. According to Spickard (2007: 4), "More than 99 percent of the current US population can at least theoretically trace its ancestry back to people who came […] from somewhere else".

Figure 1. US migrant population (absolute and share of total population), 1850-2010
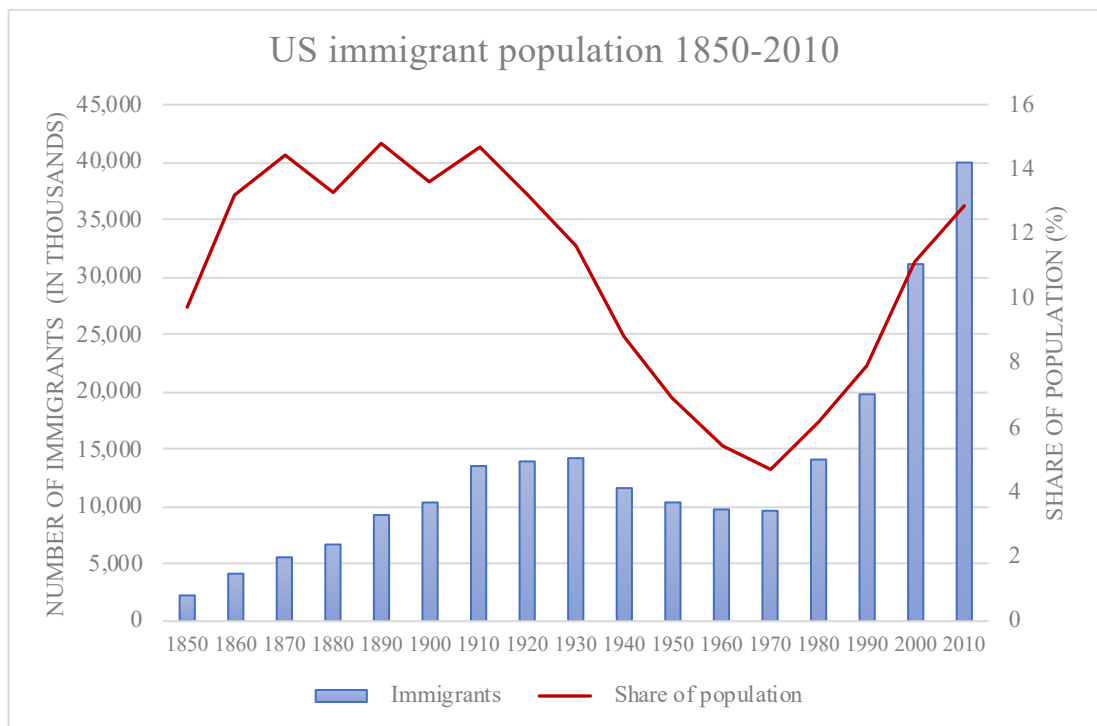
Data source: Migration Policy Institute (2017); own elaboration

Two episodes in US immigration history stand out: the Age of Mass Migration around the turn of the 20th century and the period between 1970 and 2010. During these periods immigrant numbers rose rapidly year-on-year just as immigrant population growth rates exceeded formerly known levels. Between 1850 and 1920, the absolute number of immigrants – defined as those people having a non-US birthplace but residing in the US – increased from 2 to 14 million (Figure 1). Between 1970 and 2010, the foreign population in the US rose from nearly 10 million to 40 million. Such numbers represented close to 15 percent of the total

population in historical and 13 percent in modern times (Figure 1). Both migration waves differ, however, in composition and settlement pattern. Figure 2 depicts the shift in major sending regions between 1880 and 1999.

The first two columns of Figure 2 depict the US migrant population during the two peaks of the Age of Mass Migration. Between 1880 and 1889 close to 90 percent of the foreign-born population in the US was of European stock (Bertocchi and Strozzi, 2006). Early migrant contingents involved Northern and Western Europeans, primarily from England, Ireland, Germany, or the Scandinavian countries. Later migrants had Southern and Eastern European roots: Italians, Poles, Russians, Greeks, or Portuguese (Alexander, 2007; Rodríguez-Pose and von Berlepsch, 2015).
Migrants tended to be male, young, single, relatively poor, and unskilled, with limited knowledge of English. After entering the US mostly via the major ports of the eastern seaboard, migrants quickly followed in the footsteps of friends and relatives, forming a distinct migrant settlement pattern and creating marked migrant communities across the country (Bruhn, 2005). Most migrants settled in what were then sparsely populated regions of the north and west of the country and avoided southern states (cf. Rodríguez-Pose and von Berlepsch, 2014).

Figure 2. Persons obtaining legal permanent resident status, 1880–1999
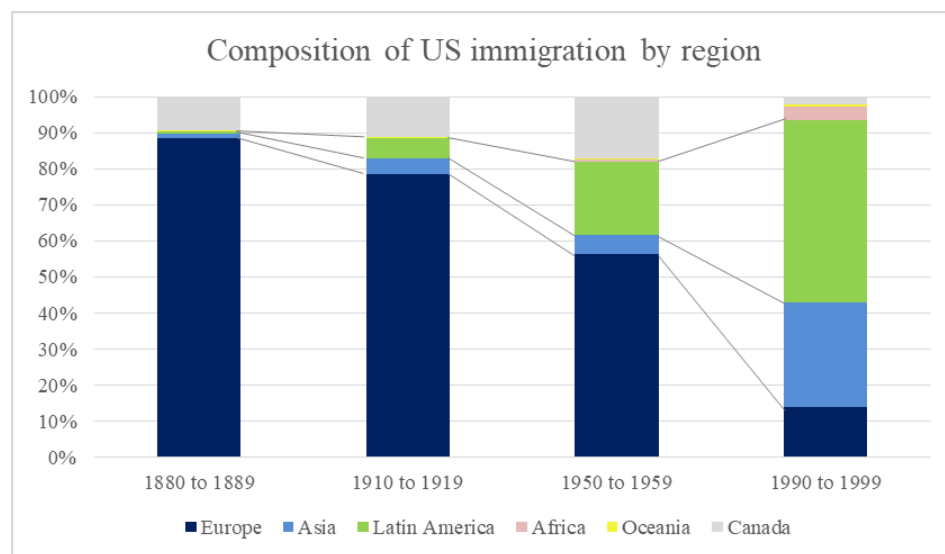
Data source: US Department of Homeland Security (2011); own elaboration

The post-1950 columns in Figure 2 display a clear shift in demographic makeup of the foreign-born population in the US. Over the course of the second half of the 20th century, the proportion of Europeans drastically shrank and two 'new' major sending regions emerged – Asia and, to an even greater degree, Latin America. By 2000, migrants from Mexico, the Spanish-speaking Caribbean, and Central and South America made up more than 50 percent of the foreign-born population, Asians represented around one third, while the European share had decreased to just 14 percent (Figure 2).

The foreign-born of the late 20th century adopted a somewhat different settlement pattern than their predecessors one hundred years earlier (see Figure A1 in

Appendix). Late 20th and early 21st century migrants fundamentally resided in western states, such as Washington, Oregon, parts of Idaho, California, or Nevada and along the border with Mexico in Arizona, New Mexico, and Texas. Historically attractive states around the Great Lakes and the Mid-West appealed to limited numbers of migrants. Additional concentrations were found in Florida and the north east. The states of the 'Old South' remained, at least until 1990, relatively migration free.

Table 1. Latin Americans obtaining legal permanent resident status, 1960-69, 2000-09 (by country of last residence)

Region and country of last residence	1960 to 1969		2000 to 2009	
	Total number	Share	Total number	Share
Latin America	1.241.044	100%	4.205.877	100%
Mexico	441.824	36%	1.704.166	41%
Caribbean	427.235	34%	1.053.969	25%
Cuba	202.030	16%	271.742	6%
Dominican Republic	83.552	7%	291.492	7%
Haiti	28.992	2%	203.827	5%
Jamaica	62.218	5%	172.523	4%
Other Caribbean	50.443	4%	114.385	3%
Central America	98.560	8%	591.130	14%
Belize	4.185	0%	9.682	0%
Costa Rica	17.975	1%	21.571	1%
El Salvador	14.405	1%	251.237	6%
Guatemala	14.357	1%	156.992	4%
Honduras	15.078	1%	63.513	2%
Nicaragua	10.383	1%	70.015	2%
Panama [1]	22.177	2%	18.120	0%
Other Central America	-	-	-	-
South America	250.754	20%	856.593	20%
Argentina	49.384	4%	47.955	1%
Bolivia	6.205	0%	21.921	1%
Brazil	29.238	2%	115.404	3%
Chile	12.384	1%	19.792	0%
Colombia	68.371	6%	236.570	6%
Ecuador	34.107	3%	107.977	3%
Guyana	4.546	0%	70.373	2%
Paraguay	1.249	0%	4.623	0%
Peru	19.783	2%	137.614	3%
Suriname	612	0%	2.363	0%
Uruguay	4.089	0%	9.827	0%
Venezuela	20.758	2%	82.087	2%
Other South America	28	0%	87	0%
Other America	22.671	2%	19	0%

Source: own elaboration using data from the US Department of Homeland Security (2011)

These migration patterns have been mainly shaped by the largest migrant group: the Latinos (see Figure 3). Although migration from the 'new world' is by no means a new phenomenon, migrant flows from Latin America rapidly increased after WWII. The influx of Latinos was boosted by initiatives targeting labour shortages, such as the *bracero* program with Mexico (Daniels, 1990); by changing regulations in immigration law – replacing a system of national origin quotas by one giving

preference to skills, occupations, refugee status, or facilitating family reunifications (Sáenz and Morales, 2015); and by political unrest in many Latin American countries, such as Castro's ascent to power in Cuba or the later *Mariel Crisis* (Daniels, 1990; Bergad and Klein, 2010).

While in the early 1960s, the incoming Latino immigrants were overwhelmingly from Mexico, with relatively large Cuban and Puerto Rican[2] contingents, by 2000, the Latino community grew massively in diversity (see Table 1).[3] Caribbeans – both Spanish- as well as French- and English-speaking migrants – and Central and South Americans – mainly from El Salvador, Honduras, Guatemala, Nicaragua, Colombia, Peru, and Ecuador – accounted for almost 40 percent of immigration from the 'new world'. Despite the differences in countries of origin and in traditions, backgrounds, and customs, the late 20th century Latin American migrant did not greatly differ from the late 19th century European migrant: young, poor, unskilled, pious, and from rural background. There was, however, a more balanced gender-ratio compared to their European precursors. Women immigrants, especially from South and Central America, were strongly represented in this migration wave (Bergad and Klein, 2010).
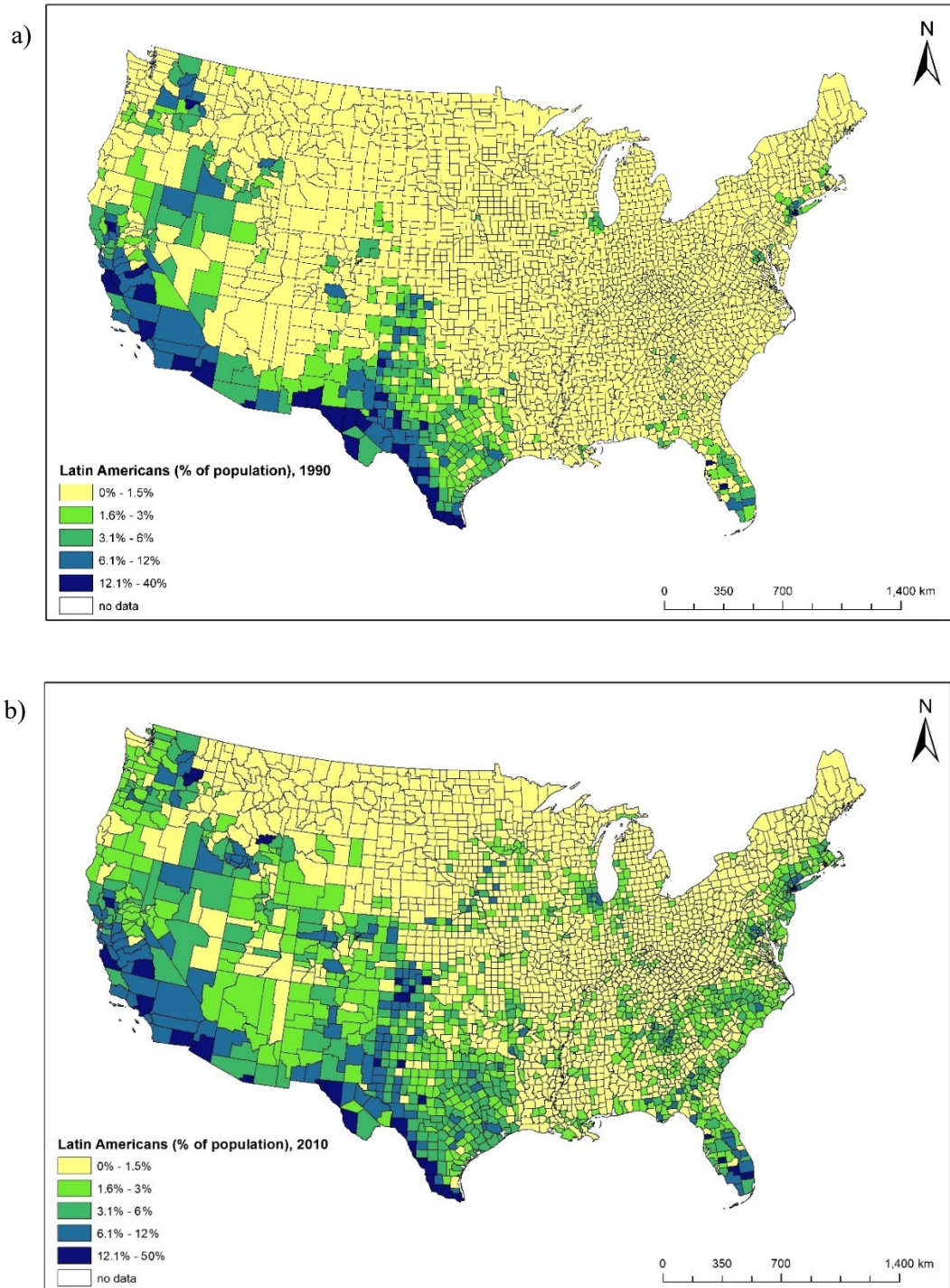
Once in the US, Latino immigrants generally occupied the lowest levels of the social and economic scale. They tended to work in low-paying jobs within the "Latino immigrant occupational niche" (Sáenz and Morales, 2015: 109), such as "agriculture; meat, poultry, and seafood processing; construction; waiters/waitresses; cooks; maids and housekeeping cleaners; and janitors and building cleaners […]" (Douglas and Sáenz, 2008: 169). Although still overrepresented in agriculture, by 1990 over 90 percent of Latinos lived in metropolitan areas. New York, Los Angeles, and Miami are the hubs of Latino culture in the US (Cafferty and Engstrom, 2000).

Prior to 2000, Latinos were concentrated in three states: Florida, California, and Texas. New York, New Mexico, Arizona, and parts of Nevada and Washington also had sizeable Latino populations (Figure 3a). From the 2000s onwards and following processes of industrial restructuring, variations in the geography of labour demand, and changes to immigration legislation and border policies, Latino immigrants fanned out beyond their traditional destinations (Donato *et al.*, 2008). As depicted in Figure 3b, states such as Colorado, Utah, and Wyoming began to receive large Hispanic contingents. Areas without a history of immigration of any sort, such as North Carolina, Virginia, or Georgia, as well as some states in the Midwest lured growing numbers of Latino migrants, (i.e. Bump, 2005; Hansen, 2005). Figure 3 displays the geographic dispersion of Hispanic immigrants, which transformed Latin American population flows "from a regional to a national phenomenon" (Massey and Capoferro, 2008: 47).

---

[2] The Puerto Ricans are not shown in Table 1, as they are considered American-born since the Jones Act in 1917 (Bergad and Klein, 2010).

[3] Table 1 only captures the official extent of legal immigration. Undocumented migrants to the US are not included in these numbers, even though they became increasingly relevant from the 1960s onwards.

Figure 3. Latino foreign-born population as share of total population, 1990 and 2010

a)

Latin Americans (% of population), 1990
- 0% - 1.5%
- 1.6% - 3%
- 3.1% - 6%
- 6.1% - 12%
- 12.1% - 40%
- no data

0   350   700   1,400 km

b)

Latin Americans (% of population), 2010
- 0% - 1.5%
- 1.6% - 3%
- 3.1% - 6%
- 6.1% - 12%
- 12.1% - 50%
- no data

0   350   700   1,400 km

Data source: Minnesota Population Center (2016); own elaboration

Why do migrants end up in particular places and not in others?

Many studies have aimed to understand why people migrate and what determines the volume of migration and settlement patterns. Often migration is considered a logical consequence of factor price differentials across geographic units generated by

varying endowments in the supply and demand of labour. Migrants are expected to move from low- to high-wage regions, with the aim of maximising the individual income (i.e. Hicks, 1932; Ödberg, 1997). Where a migrant decides to settle hinges on the maximisation of individual productivity – i.e. where the highest expected net income level can be attained relative to the acquired skill set and expected time horizon (Sjaastad, 1962). From this perspective, regions attracting large number of migrants are generally high productivity areas offering high wages (Greenwood, 1997; Ödberg, 1997).

The pure income maximization perspective has, however, been challenged in recent decades. Greater attention has been paid to how labour markets, globalisation, history, households, quality of life, and social capital influence migration (see overviews by Massey *et al.*, 1993, 2005; Bodvarsson *et al.*, 2014). Two strands have become particularly prominent – one analysing the regional characteristics behind a territory's attractiveness to potential immigrants and another evaluating individual factors influencing the probability to migrate. A combination of both goes far in explaining not only the magnitude of migration but also the composition, direction, and final settlement pattern (Massey *et al.*, 1993).

Key push and pull factors influencing population movements have been identified. On top of high wages, job availability, the possibility of job progress, and low unemployment (Pissarides and McMaster, 1990; Faggian and McCann, 2008; Biagi *et al.*, 2011), the size and composition of the economy are also a draw for immigrants (Partridge and Rickman, 1996). Regional market potential linked to agglomeration patterns increases a place's attractiveness (Ottaviano and Puga, 1998). Other relevant push factors are quasi-financial forms of income, such as social welfare spending and public amenities, as well as re-distributional transfer mechanisms, providing insurance against income losses, and the availability of public goods and regional institutional quality (Day, 1992; Ketterer and Rodríguez-Pose, 2015). In short, economically dynamic regions with high levels of regional development, adequate financial incentives, and widely available employment opportunities generally act as magnets for immigrants.

Beyond the purely economic factors, a raft of social aspects also captured the attention of researchers. A good education system and favourable human capital endowments at destination have been shown to facilitate increases in migrant productivity as well as easing transitions into the job market. Highly educated individuals are also more likely to migrate (Greenwood, 1997; Zimmermann, 2005), as "higher education reduces the risks of migration through a higher ability to collect and process information" (Zimmermann, 2005: 429). The demographic composition of the population in the regions of origin and of destination further shapes migration flows. A predominantly young population, for example, is more likely to lead to larger population out-flows, once labour market conditions tighten. Expected gains in lifetime income are significantly lower for the elderly than for the young, making the latter group more mobile (i.e. Zimmermann, 2005). Urban and natural amenities, such as a region's cultural life, its history, climate and natural environment, and quality of life, have also attracted the attention of migration research. Boosting a region's amenities is increasingly regarded as a basic pull factor for incoming skills and talent (Ferguson *et al.*, 2007; Partridge, 2010; Rodríguez-Pose and Ketterer, 2012; Liu *et al.*, 2018).

Lastly, the presence of social relationships and networks – defined in most cases by kinship, ethnicity, or friendship – is key for the attraction of specific groups of migrants to a certain territory. The propensity of migrants to move to places where family, relatives, and friends have previously settled is normally referred to as repeat or 'beaten-path' migration (Hoover, 1971; DaVanzo, 1983). The 'beaten-path' refers to community and social ties stretching from place of origin to host region (i.e. Epstein, 2002; Epstein and Gang, 2006; McGovern, 2007; Radu, 2008; Jewell and Molina, 2009; Sharma, 2012; Bodvarsson *et al.*, 2014; Biagi and Dotzel, 2018). Migration decisions therefore depend not only on the individual's own actions and characteristics, but also on the specific migration choices of the individual's peer group. A large community of friends and family at destination (a so-called kinship network) reduces relocation costs and increases the expected return associated with migration. Having relatives in a foreign destination generates a self-perpetuating element as, "each act of migration creates additional social ties for future migrants, who in turn extend the range of social capital for further migrants" (McGovern, 2007: 220). Similarly, large and well established groups of earlier migrants, sharing a common origin (known as a migrant network), generates an equally positive externality (Bodvarsson *et al.*, 2014). Once the kinship and migrant network in a territory reaches a critical mass, it affects new migrant settlement patterns by attracting more and more population from the shared region of origin. The result is a geographical clustering of immigrants from specific local, national, or ethnic backgrounds (Bauer *et al.*, 2007, 2009; Jaeger, 2007).

Access to these networks lowers the psychological and information costs for the arriving individual and hence significantly affects the volume of migrants and their choice of destination (i.e. Yap, 1977; Massey and España, 1987; Radu, 2008). A large stock of migrants from the same origin creates social capital which reduces employment and housing search costs, lowers language barriers, offers protection from crime and income loss, provides temporary credit and lodging, and eases the individual's settlement process. It therefore facilitates integration into what is initially an alien environment, administration, culture, and society (Daniels, 1990; Carrington *et al.*, 1996; Massey *et al.*, 2005; Bodvarsson *et al.*, 2014). This phenomenon implies that total moving costs are endogenous to the volume of previous migrants. "Once started, migration develops momentum, as current migration reduces the cost of future migration [...] [it] continues or even accelerates [...] [and] is channelled in that migratory paths emerge" (Carrington *et al.*, 1996: 910). Many examples can be found to support these findings. Massey *et al.* (1987), for example, reported that nearly 40 percent of Mexicans found employment in the US via friends or relatives. Munshi (2003) calculated that established community ties led to significantly higher employment rates and income levels among Mexican immigrants in the US. 'Migration clubs' managed by kith and kin among black southerners in the north, as well as soccer clubs involving Mexican immigrants in Los Angeles, provided the linguistic and cultural cushion for the new arrivals (Massey *et al.*, 1987; Carrington et al., 1996). Similarly, European migrants nearly a century earlier sent letters to family and friends with information on employment and the housing market in the US, often asking to join them and attaching money to pay for transatlantic trips (Daniels, 1990; Joly, 2000).

Repeat migration can trigger path dependence or 'herd behaviour' (Bikhchandani *et al.*, 1998; Epstein, 2002) between place of origin and destination, affecting the volume and direction of migration flows. "Once the number of network connections

in an origin area reaches a critical threshold, migration becomes self-perpetuating because each act of migration itself creates the social structure needed to sustain it" (Massey *et al.*, 1993: 449). In short, a large, pre-established migrant network impacts self-selection, decreases the costs of moving, raises the potential net gain for the would-be migrant, stimulates mobility, and steers migration flows into migration-prone regions, following ethnic, village, or even family ties (Shah and Menon, 1999; Clark *et al.*, 2007; McKenzie and Rapoport, 2010; Simpson and Sparber, 2013).

One factor of particular importance in the migration network literature is the common birthplace. A shared geographical origin, ethnic bond, or common background is crucial for migrants to reap the benefits of networks and in determining a region's appeal. The literature assumes that a common origin, a shared community-belonging, or family ties are the sources for the establishment of migrant networks and the development of path dependence. Membership of a shared community therefore acts as a pull factor for incoming migrants, determining the final migrant settlement pattern. However, one important question has not been explored: what if the migrant networks are not formed on the basis of a common ethnicity or birthplace, but simply on the presence of previous migration groups, regardless of origin? What if a large community of migrants in a given place creates a favourable environment for migrants which endures over time? Could the institutional setting established by previous generations of migrants transform a locality or region into a migration-prone area, welcoming to migrants for generations, regardless of origin? What if past migration begets future migration? In their work on the long-term impact of migration and its implications for regional development in the US, Rodríguez-Pose and von Berlepsch (2014, 2015) assume that the character of the migrant – more risk-seeking, entrepreneurial, and dynamic – becomes engraved in the territories where migrants settle. They posit that the presence of large groups of immigrants can generate a migration vibe in the receiving areas which transforms the institutional framework in ways that sends signals to future migrant generations (Rodríguez-Pose and von Berlepsch, 2019).

There is, however, limited empirical evidence to support or refute the claim that past migration patterns transform territories into migration-prone or migration-averse areas over long timeframes. This is precisely what this paper aims to demonstrate, that past migration begets future migration and that this connection takes place beyond sharing common ties and a network. Past migration creates migration-prone territories that attract new migrants, irrespective of their national origin. Two basic hypotheses are, therefore, tested:

H_1: Past migration patterns determine future migration;

and

H_2: This path-dependency in migration patterns does not necessarily rely on kinship, national, and/or ethnic networks (the 'beaten-path' hypothesis), but on the transformation of territories by past migrants into migration-prone places

Econometric specification

The two hypotheses presented above are tested by comparing the settlement patterns of two migration waves of a very different nature – Europeans in the late 19th and early 20th centuries with Latinos at the end of the 20th and the beginning of the 21st centuries. The econometric model estimated to assess whether past migration waves shape the settlement pattern of completely unrelated later waves of migrants takes the following form:

$$Latinos_{i,t} = \alpha + \beta\, Migration_{i,t_0} + \partial X_{i,t-10} + \theta\, Z_{i,t_0} + \mu\, distMX_i + \vartheta\, state + \varepsilon_{it}$$

where *Latinos* is the share of total population born in a Latin American country[4] in county i in period t (t=1970, 1980….2010); *Migration*, the variable of interest, represents the share of foreign-born population in county i at t_0 (t_0=1880 or 1910); X is a vector of economic and socio-demographic characteristics of county i which are assumed to serve as pull factors to recent migrants (at time *t-10*; that is, 10 years before the migrant account is taken) and Z corresponds to a vector at t_0 which includes the same county characteristics as vector X that would have determined the appeal of the county to migrants more than a century ago. *DistMX* stands for the distance of any given county i to the Mexican-American border, while *state* represents state-specific fixed-effects controlling for potential spatial correlation between counties within a given state. ε describes robust standard errors.

The model is run consecutively per decade between 1970 and 2010 for the 3,109 continental counties of the US in 2010, covering a 40-year timeframe and shifting the dependent variable in each regression by 10 years. As any analysis involving economic and migration data is prone to endogeneity issues, introducing the dependent (*Latinos*) and explanatory variables (vector X) within the same time structure is highly problematic (Treyz *et al.*, 1993; Özgen *et al.*, 2011). Hence, all explanatory variables determining current migration flows are lagged by 10 years. This way, the risk of reverse causality is mitigated with, on the one hand, migration flows impacting regional economic features and, on the other, regional characteristics simultaneously directing migratory settlement patterns.

Further endogeneity issues connected to omitted variable bias or spatial sorting are treated by means of three different robustness checks. First, the consecutive estimations are rerun, shifting the base migration year by 30 years, from 1880 to 1910. Both 1880 and 1910 depict peak years within the period of mass migration to the US: 1880 represents the peak of the first wave of migrants, while 1910 that of the second wave. Two different migrant compositions are thus considered. As the main source of migration shifted around the turn of the century, the 1880 specification comprises mostly northern and western Europeans, predominantly from the British Isles, Germany, or Scandinavia. Using 1910 as base year, an entirely different composition of the foreign-born population is represented: mostly southern and eastern Europeans from Italy, Poland, the former Soviet Union, Portugal, or Greece (Hatton and Williamson, 1998). The shift in the composition of migrant groups

---

[4] The focus of the analysis lies only on the foreign-born Latino population. Other indicators recorded by the census, such as a Spanish last name, the usage of Spanish at home, or Hispanic origin were not consistently included in the questionnaires across the years and are therefore discarded.

between 1880 and 1910 minimises the risk of the results being driven by omitted variables connected to the character of the specific type of migrant.

Second, an instrumental variable estimation is used in order to extract the underlying effect of historical on current migration waves. By capturing the exogenous variation in the migration variable of 1880, the potential endogeneity bias in the least-squares estimates is reduced.

Last, a third robustness check consists of reshaping the dataset into a quasi-panel structure. As the use of traditional FE-models is impossible, given that the analysis is built around a time-invariant variable of interest (*Migration*) and the probability of a high correlation between some of the time-varying variables and region-specific fixed-effects, an alternative Hausman and Taylor (HT) (1981) instrumental variable estimator is used. This estimator allows for the inclusion of time-invariant variables such as *Z* or *distMX* in a panel setting using the within transformation of time-varying variables to estimate consistent coefficients. It also accepts the potentiality of a correlation between some of the independent variables and individual specific effects. Individual means of the uncorrelated regressors, on the other hand, are employed to instrument for the endogenous variables (Baltagi *et al.*, 2003; Baltagi, 2008). Vector *X* contains the endogenous variables in our model.

Data

The migration data employed to construct the independent variable of interest *Migration* was extracted from the Integrated Public Use Microdata Series USA database (IPUMS) Version 6.0 (Ruggles *et al.*, 2015). This database provides US microdata covering censuses and American Community Surveys between 1790 and 2010.[5] Using a weighted sample of the US population including 5,791,531 individuals (11 percent of the total population in 1880) and 923,153 individuals (1 percent of the population in 1910), individual data points are matched to the respective county of residence at the time of the census and aggregated at county level. Following changes in county numbers, size, and boundaries between 1880 and 2010, cartographic boundary files of the 48 continental states (excluding Alaska and Hawaii) provided by the US Census are used for every decade of analysis. In order to normalise borders to 2010 county boundaries, averages weighted by population density at the time of the boundary change were calculated for each individual county in historical years. With the exception of Oklahoma in 1880,[6] this method allowed to match historical county features to their 2010 equivalent.[7]

The dependent variable, *Latinos*, was constructed using the National Historical Geographic Information System (NHGIS) Version 11.0 (Minnesota Population Center, 2016), which consists of data on US geographic units covering the population, housing, agriculture, and the economy from 1790 until today. The data

---

[5] The American Community Survey was only initiated in 2005.

[6] Oklahoma only became an organized territory in 1890.

[7] The US territory comprised 2,875 counties or equivalent territorial units in the 48 contiguous states in 1880; 3,123 in 1910; and 3,109 in 2010.

on the Latino population at county level was extracted for the years between 1970 and 2010. As not all Latin American countries are listed independently within the NHGIS database, the share of Latin American population refers to Cubans, Mexicans, and all other people born in the Americas, excluding the US and Canada.

Three different data sources were used to construct the county level database of control variables included in vector *X*. All county data for the years 1960 to 2000 was extracted from the US Bureau of Economic Analysis (BEA) database, the Social Explorer data collection, and the Inter-University Consortium for Political and Social Research (ICPSR) historical databases. Vector *X* is included in the model to control for recent county level economic and socio-demographic characteristics, covering the relevant pull factors for migrant volume and settlement discussed in the literature. These are assumed to directly affect the settlement pattern of Latino migrants and the share of Latinos in the population of any given US county. Control factors include income per capita (as natural log), the share of population living in urban areas, the education level measured as the share of adults with tertiary education, the unemployment rate, the share of employment in agriculture as proxy for industry structure, female participation rates in the labour force, and the share of the black and female populations. All income variables were adjusted for inflation and converted to 2010 dollars using the CPI inflation converter of the Bureau of Labor Statistics (BLS).

The second vector of control variables, vector *Z*, dates from the time of the historical migration: 1880 and 1910 respectively. Vector *Z* comprises the historical equivalent of all variables included in vector *X*. It is used to control for historical economic and socio-demographic county features which may have influenced not only economic development levels but also the settlement pattern of migrants in 1880 and 1910 respectively. These historic variables are either extracted from the ICPSR database or constructed using the IPUMS USA weighted population samples. In the latter case, the same method as with the independent variable of interest is repeated, aggregating all individual data at county level. As income per capita data were not collected at the turn of the century, income proxies for 1880 and 1910 are constructed using the median total income per occupation category in 1950 dollars. These occupational income equivalents were allocated to an individual's occupation in 1880 and 1910 and then aggregated at county level.

The distance to Mexico (as natural log), included as further control, was constructed using GIS software, calculating the point-distance matrix between a county's centroid and the nearest point on the continental border with Mexico. An exact description of all variables, including sources, is presented in Table B1 in Appendix.

Instrumentation strategy

Potential endogeneity bias in the least-squares estimates is tackled by means of instrumental variable (IV) estimations. Two instruments are used to assess the direction of causation and certify the exogenous variation of the variable of interest, *Migration*. The first is population density in 1880, extracted from the ICPSR database. The second is a measure of the endowment in public services, proxied by share of employment in the health and education sector as well as in public

administration in any given county i in 1880 (or 1910). Data for the endowment in public services stem from the IPUMS USA individual data points, aggregated at the county level.

The justification for the choice of both instruments is related to their role in determining migrant settlement patterns during the Age of Mass migration. Density played a role as the newly arrived generally moved to places where their predecessors had already settled and where basic forms of infrastructure were already established. Public and often cultural organisations, such as churches, schools, newspapers, and meeting houses, became especially important in the settlement process, as they offered channels for information and knowledge exchange. Following the 'beaten-path' theory, German meeting houses, Swedish schools, or Irish churches facilitated settling down in previously unknown places (Daniels, 1990). Moreover, migrants valued access to markets and nearby settlements for basic supplies as well as the presence of community networks, schooling for their children, and the possibility of sustaining connections to their home countries. The presence of a post office was crucial in maintaining ties both with family and friends in the country of origin and with other migrant settlements across the US. Finally, high-density and an efficient public service increased security and reduced the perception of danger. Hence, population density and public services influenced migrant settlement decisions. As population density patterns and employment features drastically changed between 1880 (1910) and 2010, the exogeneity condition separating out the uncorrelated component of our endogenous variable, *Migration*, in the first stage regression is fulfilled. There is no significant correlation between 1880 population density and public service endowments and the location pattern of Latinos more than 90 years later. This reinforces the exogeneity of the chosen instruments.

From an econometric perspective, the essential relevance property for IV analysis is satisfied as the combination of both instruments retrieves sufficiently large first-stage F-statistics based on the Staiger and Stock (1997) test for weak instruments in combination with the Stock and Yogo (2005) critical values. A further validation test of the quality of our chosen instruments is undertaken when testing for overidentifying restrictions employing the Anderson-Rubin test. This partial test of instrument exogeneity confirms, yet again, the quality of the instruments.

Results

Table 2 reports the results of estimating the model for the settlement pattern of Latinos in the US between 1970 and 2010, employing a cross-sectional structure. Specifications (1) through (5) present the coefficients and significance levels, shifting the dependent variable (share of Latinos in a given county i) by 10 years in each regression.

In line with H_1, historical migration is positively associated with the dependent variable, *Latinos*, across all five decades of analysis ranging from 1970 to 2010. A large share of migrants in 1880 is strongly and positively connected to the share of foreign-born Latino migrants in any given US county 90 to 130 years later. Hence,

once other factors are controlled for, counties that attracted large inflows of European settlers at the end of the 19th century remain significantly more appealing to incoming Latino migrants over a century later than those which were largely bypassed by European migration, also confirming H2. Despite differences in time periods, backgrounds, and cultures, both migration waves are connected by place of settlement at significance levels of 5% and 1% respectively. Hence, the results of the analysis uncover a strong territorial dependency in US migration over the last century (Table 2).

The coefficients of the lagged recent (*t*-10) control variables indicate that, for much of the late 20th century, Latino migration to the US did not follow traditional pull factors. Prior to 2000, economic factors mattered little for Latino migration. Latinos generally settled in areas of the US not particularly well off in terms of income or education levels. These pull factors do not display any significant relation to the dependent variable of interest. Even more surprisingly, local unemployment levels were positively associated with the share of Latino immigrants in a county. Hispanic settlement areas had, on average, higher levels of unemployment than the rest of the country. Socio-demographic factors, such as ethnic and gender composition, also display non-significant coefficients. Before 2000, Latinos mainly settled in two types of counties: either in highly urbanised areas in and around Los Angeles, the Bay Area, El Paso, San Antonio, Miami, or Houston, or in rural areas with large employment in agriculture. A large participation of women in the labour force also lured Latino migrants. These findings are in line with the relevant literature emphasising the 'beaten-path', the appeal of the cities with largest Latino concentrations, the rising contribution of Latinos to agricultural employment, and the gender balance in migration flows (Sáenz and Morales, 2015).

In more recent decades (2000 and 2010), Latino migrant settlement increasingly responds to more traditional factors. Regional income levels are, in line with the relevant migration literature, positively and significantly associated with Latino population shares. This implies that the location of Latinos in the US has become more geographically diversified in recent decades (see Figure 3), moving towards more prosperous areas without eroding the relevance of distance to the Mexican border as a key marker for the location of Latino immigrants (Table 2).

As could be expected, most 1880 controls are completely irrelevant to Latino settlement patterns 90 to 130 years later. There are two exceptions: regional income levels in 1880 are positively correlated to present day shares of Latino population, albeit with fluctuating significance levels. A higher female participation in the labour force more than a century ago is, by contrast, associated with lower levels of Latino presence today.

As the OLS results presented above could be prone to potential endogeneity issues caused by omitted variable bias, spatial sorting, or reversed causality, a range of robustness checks is conducted. These imply, first, shifting the base year, second, resorting to IV estimation techniques and, last, using the dataset as a quasi-panel structure and estimating the model by means of a HT-estimator.

Table C1 in Appendix displays the results for the regressions shifting the base year to 1910. Apart from temporal changes to the control variables (vector *Z*) and the *Migration* variable, all factors remain as before. The results are almost a carbon copy of those presented in Table 2. *Migration 1910* is positively and significantly

associated with the presence of Latino immigrants 60 to 100 years later. Constant significance levels of below 1% underline the robustness of this link. This means that the settlement pattern of southern and eastern European migrants in 1910 is connected, in the same way as that of their northern and western European predecessors, to Hispanic migration to the US. Latino immigrants have followed in the footsteps of past migration waves, drawn into seemingly migrant-prone regions and bypassing areas of the country that developed a historical aversion to migration.

Table 2. The impact of historical migration on Latino settlement patterns, OLS 1880

Dep. Var: % of Latinos in county	(1) 1970	(2) 1980	(3) 1990	(4) 2000	(5) 2010
Migration 1880	0.0109***	0.0165***	0.0198***	0.0213**	0.0146**
	(0.00315)	(0.00524)	(0.00726)	(0.00944)	(0.00696)
Income pc (ln)~	0.000664	0.00418	0.00272	0.0409***	0.0462***
	(0.00134)	(0.00273)	(0.00374)	(0.00807)	(0.00487)
Urban share~	0.00411***	0.00658***	0.0118***	0.0298***	0.0449***
	(0.00128)	(0.00179)	(0.00231)	(0.00338)	(0.00272)
Education~	-0.00294	-0.00263	-0.00629	-0.0604***	-0.122***
	(0.00282)	(0.00571)	(0.00726)	(0.0112)	(0.00983)
Unemployment~	0.0293**	0.0356**	0.0102	0.105***	0.196***
	(0.0131)	(0.0140)	(0.0157)	(0.0330)	(0.0295)
Empl. in agri~	0.0102***	0.0172***	0.0231***	0.0800***	0.156***
	(0.00230)	(0.00396)	(0.00718)	(0.0150)	(0.0135)
Fem. labforce~	0.00808*	0.0141**	0.0233**	-0.000482	0.0779***
	(0.00424)	(0.00692)	(0.0101)	(0.00827)	(0.0145)
Female pop~	-0.0144	0.00253	-0.0354	-0.185***	-0.229***
	(0.0125)	(0.0135)	(0.0260)	(0.0424)	(0.0320)
Black pop~	0.000216	0.00472	0.0117**	0.0124	-0.0119
	(0.00193)	(0.00342)	(0.00553)	(0.00837)	(0.00802)
Distance to MX (ln)	-0.0251***	-0.0386***	-0.0440***	-0.0465***	-0.0377***
	(0.00263)	(0.00403)	(0.00438)	(0.00504)	(0.00215)
Income (ln) 1880	0.00525**	0.00811**	0.00859*	0.00461	0.0115***
	(0.00225)	(0.00373)	(0.00513)	(0.00500)	(0.00398)
Urban share 1880	0.00382	0.00467	0.00657	0.0117*	0.0115**
	(0.00257)	(0.00420)	(0.00561)	(0.00666)	(0.00480)
Education 1880	-0.00107	-0.00395	0.00185	0.00634	0.00698
	(0.00293)	(0.00477)	(0.00616)	(0.00744)	(0.00614)
Unempl. 1880	-0.000149	0.00185	0.00162	0.00174	-0.000253
	(0.00109)	(0.00190)	(0.00195)	(0.00301)	(0.00202)
Empl. in agri 1880	0.00341**	0.00497*	0.00567	0.00627	0.00653
	(0.00161)	(0.00283)	(0.00389)	(0.00542)	(0.00457)
Fem. labforce 1880	-0.0182**	-0.0355***	-0.0376**	-0.0414**	-0.0737***
	(0.00849)	(0.0137)	(0.0189)	(0.0173)	(0.0193)
Female pop 1880	0.0172**	0.0275**	0.0260*	0.0301	0.0407**
	(0.00717)	(0.0122)	(0.0158)	(0.0217)	(0.0164)
Black pop 1880	0.000327	-0.000676	-0.00228	-0.00772	-0.00136
	(0.00282)	(0.00448)	(0.00561)	(0.00822)	(0.00782)
States	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	2,846	2,851	2,851	2,851	2,850
R-squared	0.640	0.625	0.583	0.515	0.527

~ Controls from 10 years prior to dependent variable
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

The large majority of control variables reproduce the results of the regressions in Table 2. A high income per capita, a large urban share, a strong agricultural sector, a large participation of women in the labour force, and close proximity to the Mexican border are, once again, key for Latino immigrant settlement (Table C1 in Appendix). The coefficients of traditional pull factors, such as high general education or low regional unemployment rates, yet again counter expectations.

The control variables that affected migration waves in historical times are mostly insignificant. The only regional characteristic dating back to the Age of Mass Migration that is significantly connected to the settlement decisions of Latino migrants since the 1970s is the share of foreign-born population in 1910.

In the second robustness check, IV estimation techniques are employed to extract the underlying effect of historical migration on current Latino migration waves. As mentioned earlier, *Migration* is instrumented using the combination of historical population density and public service endowment. Table 3 displays the results referring to both base years 1880 and 1910.

The use of IV estimation techniques – which minimise the potential influence of omitted variable bias on the regression coefficients – validates the results of the previous OLS regressions in Table 2. Nine out of ten regression outcomes, incorporating shifts between both decades of recent migration inflow and historical base years, confirm the territorial connection between past European and recent Latino migration flows. With significance levels mostly below 5% for 1880 and below 1% for 1910, counties that attracted migrants in 1880 have again become the chosen areas of settlement for Latino immigrants 60 to 130 later. The geography of US migration is therefore shaped by a strong path dependency, with different generations of migrants, regardless of national and/or ethnic origin, discriminating between areas that are more welcoming (migration-prone) and those that are more inimical to migrants (migration-averse). The estimated sign and significance of coefficients for both sets of controls are in line with those reported for the previous analysis (Table 3).

After reshaping the database to a panel structure, the model is run again using a HT-estimation technique. Exploiting the panel data structure while using time and regional fixed effects allows controlling for omitted factors that do not change over time but could potentially cause biased regression coefficients. The results are presented in Table 4.

For the third time, historical migration is confirmed as fundamental pull factor for ensuing migration waves. There is strong evidence, as indicated by coefficients significant at the 1 percent level, that past US migration patterns shape the geography of future migration in a path dependent way irrespective of origin or ethnicity. The result holds for both base years in question, 1880 and 1910 respectively (Table 4). Regardless of whether the focus falls on western and northern European migrants, forming the majority of the first migration wave, or on eastern and southern European migration inflows culminating in 1910, Latin American immigrants settling in the US in the second half of the 20th century and early 21st century were drawn – once other factors influencing recent migration are controlled for – to the exact same regions where their European predecessors settled over a century   earlier.

Table 3. The impact of historical migration on Latino settlement patterns, IV 1880, 1910

Dep. Var: % of Latinos in county	1880					1910				
	1970	1980	1990	2000	2010	1970	1980	1990	2000	2010
Migration #	0.0501**	0.0903**	0.0993**	0.109*	0.0562	0.127***	0.231***	0.382***	0.470***	0.451***
	(0.0237)	(0.0377)	(0.0488)	(0.0571)	(0.0606)	(0.0299)	(0.0536)	(0.0708)	(0.109)	(0.129)
Income pc (ln)~	-0.000423	0.00273	-0.000217	0.0353***	0.0436***	-0.000842	0.000631	-0.00623	0.0122	0.0181
	(0.00148)	(0.00272)	(0.00390)	(0.00902)	(0.00786)	(0.00188)	(0.00370)	(0.00502)	(0.0141)	(0.0146)
Urban share~	0.00369***	0.00704***	0.0122***	0.0302***	0.0455***	0.00282*	0.00598**	0.0150***	0.0378***	0.0539***
	(0.00137)	(0.00199)	(0.00246)	(0.00353)	(0.00365)	(0.00169)	(0.00263)	(0.00387)	(0.00534)	(0.00533)
Education~	-0.000232	0.00106	-0.00269	-0.0512***	-0.118***	0.00910*	0.0112	0.00837	-0.0266	-0.113***
	(0.00365)	(0.00679)	(0.00812)	(0.0124)	(0.0122)	(0.00475)	(0.00889)	(0.0116)	(0.0180)	(0.0167)
Unemployment~	0.0314**	0.0355**	0.00229	0.101***	0.191***	0.0458***	0.0724***	0.0502*	0.152***	0.219***
	(0.0135)	(0.0164)	(0.0188)	(0.0343)	(0.0385)	(0.0152)	(0.0222)	(0.0286)	(0.0424)	(0.0530)
Empl. in agri~	0.00870***	0.0161***	0.0192***	0.0816***	0.156***	0.0117***	0.0244***	0.0375***	0.117***	0.175***
	(0.00240)	(0.00437)	(0.00764)	(0.0151)	(0.0201)	(0.00296)	(0.00590)	(0.0108)	(0.0203)	(0.0265)
Fem. labforce~	0.0159**	0.0173**	0.0240**	0.00172	0.0768***	0.0126*	0.0187*	0.0192	0.00566	0.102***
	(0.00668)	(0.00774)	(0.0108)	(0.00855)	(0.0149)	(0.00643)	(0.00996)	(0.0166)	(0.0123)	(0.0215)
Female pop~	-9.44e-05	0.0176	-0.00761	-0.153***	-0.220***	-0.000259	0.0370*	0.0215	-0.165***	-0.209***
	(0.0154)	(0.0170)	(0.0327)	(0.0446)	(0.0600)	(0.0135)	(0.0211)	(0.0388)	(0.0568)	(0.0594)
Black pop~	-0.000105	0.00426	0.0115**	0.0138*	-0.00998	0.000972	0.00350	0.00813	0.00536	-0.0140*
	(0.00189)	(0.00334)	(0.00535)	(0.00828)	(0.00879)	(0.00227)	(0.00368)	(0.00533)	(0.00747)	(0.00835)
Dist. to MX (ln)	-0.0215***	-0.0316***	-0.0364***	-0.0379***	-0.0336***	-0.0183***	-0.0264***	-0.0228***	-0.0192**	-0.0122
	(0.00367)	(0.00573)	(0.00665)	(0.00762)	(0.00777)	(0.00329)	(0.00561)	(0.00683)	(0.00949)	(0.0100)
Income (ln) #	0.00558**	0.00892**	0.00929*	0.00552	0.0120*	-0.00127	-0.00405	-0.00280	-0.000364	0.000301
	(0.00233)	(0.00397)	(0.00527)	(0.00511)	(0.00698)	(0.00143)	(0.00273)	(0.00432)	(0.00563)	(0.00603)
Urban share #	-0.000467	-0.00310	-0.00184	0.00353	0.00785	-0.00667***	-0.0116**	-0.0233***	-0.0290***	-0.0319***
	(0.00387)	(0.00618)	(0.00822)	(0.00897)	(0.00977)	(0.00305)	(0.00505)	(0.00703)	(0.0101)	(0.0113)
Education #	0.00247	0.00345	0.00935	0.0151	0.0111	0.0180**	0.0433***	0.0812***	0.109***	0.127***
	(0.00346)	(0.00576)	(0.00678)	(0.00955)	(0.00925)	(0.00780)	(0.0141)	(0.0184)	(0.0276)	(0.0318)
Unemployment #	0.000384	0.00275	0.00276	0.00276	0.000195	-0.00854	-0.00791	-0.0134	-0.0260	-0.0391
	(0.00120)	(0.00226)	(0.00236)	(0.00332)	(0.00303)	(0.00837)	(0.0136)	(0.0203)	(0.0257)	(0.0299)
Empl. in agri #	0.00857***	0.0145**	0.0155**	0.0173*	0.0117	0.00260	0.00787***	0.0195***	0.0280***	0.0258***
	(0.00346)	(0.00572)	(0.00693)	(0.00896)	(0.00873)	(0.00271)	(0.00478)	(0.00735)	(0.00937)	(0.0102)
Fem. labforce #	-0.0196**	-0.0367***	-0.0385**	-0.0432**	-0.0740***	-0.00557	-0.00406	-0.00709	-0.0246	-0.0119
	(0.00877)	(0.0141)	(0.0191)	(0.0174)	(0.0251)	(0.00444)	(0.00845)	(0.0129)	(0.0165)	(0.0175)
Female pop #	0.0279**	0.0495***	0.0481**	0.0571**	0.0536*	0.124***	0.209***	0.372***	0.503***	0.450***
	(0.0111)	(0.0192)	(0.0235)	(0.0279)	(0.0315)	(0.0307)	(0.0591)	(0.0862)	(0.127)	(0.148)
Black pop #	0.00279	0.00455	0.00266	-0.00310	0.000262	0.00947***	0.0150***	0.0218***	0.0149	0.0231*
	(0.00295)	(0.00476)	(0.00590)	(0.00900)	(0.00891)	(0.00218)	(0.00434)	(0.00715)	(0.0118)	(0.0139)
States	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,844	2,849	2,849	2,849	2,848	3,118	3,120	3,120	3,120	3,119
First stage F-stat	18.99	17.14	16.04	12.21	11.69	25.59	33.39	35.75	25.82	18.16

~ Controls from 10 years prior to dependent variable | # respective base year 1880 or 1910 | Robust standard errors in parentheses | *** p<0.01, ** p<0.05, * p<0.1

The control variables largely reproduce the results of the OLS and IV regressions. The only differences relate to income per capita which, once the characteristics of the panel structure controlling for regional and time fixed effects are exploited, turns out to be negatively associated to the presence of a large Latino community in any given US county from 1970 onwards. In contrast to the importance of historical migration settlement patterns, income and employment opportunities cannot be considered a key driver for the recent settlement of Hispanic populations in the US.

Table 4. HT estimation exploiting quasi-panel structure, 1880 and 1910

Dep. Var. Share of Latino population (1970-2010)	(1) Migrants 1880	(2) Migrants 1910
Migration #	0.0202***	0.0341***
	(0.00626)	(0.00879)
Income per capita (t-10)	-0.0136***	-0.0122***
	(0.00214)	(0.00189)
Urban share (t-10)	0.0258***	0.0264***
	(0.00283)	(0.00277)
Education (t-10)	0.0105**	-0.000224
	(0.00529)	(0.00553)
Unemployment (t-10)	0.0787***	0.0770***
	(0.0111)	(0.0104)
Employment in agriculture (t-10)	0.0257***	0.0277***
	(0.00424)	(0.00400)
Fem. part. in the lab.force (t-10)	0.00250	0.00212
	(0.00385)	(0.00368)
Female population (t-10)	-0.188***	-0.190***
	(0.0361)	(0.0347)
Black population (t-10)	-0.00139	-0.00238
	(0.0110)	(0.0108)
Distance to Mexico (ln)	-0.0365***	-0.0359***
	(0.00378)	(0.00359)
Income (ln) (t_0)	0.00910**	0.00334
	(0.00422)	(0.00224)
Urban share (t_0)	0.00642	0.00114
	(0.00500)	(0.00342)
Education (t_0)	0.0114**	0.0240***
	(0.00495)	(0.00662)
Unemployment (t_0)	0.00146	-0.0176
	(0.00203)	(0.0119)
Employment in agriculture (t_0)	0.00776**	0.00806***
	(0.00349)	(0.00306)
Fem. part in the lab. force (t_0)	-0.0423***	-0.0158*
	(0.0153)	(0.00931)
Female population (t_0)	0.0292**	0.0296
	(0.0147)	(0.0258)
Black population (t_0)	0.00356	0.00274
	(0.00781)	(0.00813)
States	*Yes*	*Yes*
Years	*Yes*	*Yes*
Observations	14,249	15,597
Number of cnty2	2,851	3,120

Robust standard errors in parentheses | *** p<0.01, ** p<0.05, * p<0.1 |
respective base year 1880 or 1910

In short, all three robustness checks have confirmed the initial OLS results and H_1 and H_2. Irrespective of the method, there is a consistent, positive, and strongly significant geographical link between past and current migrant stocks. Settlement patterns of predominantly European migrants around the turn of the 19th to the 20th century have had a strong influence on those of Latin American migrants a century later. The transformation of territories by past migration has turned these regions into magnets for subsequent migrant waves, serving as a long-lasting pull factor – a more important one than any other historical county characteristic. Entirely different backgrounds, ethnicities, and origins, fundamental differences in customs, traditions, and migration incentives, and a time gap of three to five generations have not managed to break the invisible cord linking both migration waves. Past migration waves have left an imprint that makes particular areas of the US attractive to new generations of migrants. This legacy, generated more than 100 years ago, has become engraved in the receiving territories and represents a fundamental but often ignored pull factor for current migrant inflows, creating a long-lasting division between migration-prone and migration-averse areas.

Conclusion

Over the course of two and a half centuries, the US has attracted millions of migrants. Those migrants have come from every part of the globe with the aim of building a better life for themselves and their families. Ninety-nine percent of the current US population can trace their ancestry to former migrant stock (Spickard, 2007). Whilst settlement patterns of migrants in the US have been highly uneven, and despite differences from migration wave to migration wave, certain areas of the country have proven more attractive to the foreign-born population than others.

Why is this the case? Traditional migration theory has aimed to explain why migrants flock to particular areas using economic and socio-demographic characteristics as well as natural and cultural aspects as explanations. More recently, the focus has been on the presence of kinship and migrant networks at destination. The presence of a large migrant network, a shared geographical origin, ethnic connection, or common background between already established and newly arriving migrants in a particular place has been identified as a particularly crucial pull factor for new arrivals.

Nevertheless, there is virtually no research on whether past migration serves – once 'beaten-path' networks based on nationality, ethnic origin, village, or family ties have been accounted for – as significant pull factor for new migrants across centuries. This paper has aimed to cover this gap in the literature by assessing the presence of path dependency in migration flows in the US in absence of a common ethnicity or birthplace. The research intended to ascertain if completely unrelated migration waves to the US, involving very different groups, separated by several generations, have followed similar geographical patterns, creating a division between migration-prone and migration-averse areas. The research has involved putting together a dataset comprising county data on foreign-born individuals for two of the largest waves of migration to the US – the Age of Mass migration at the turn of the 19th to the 20th century and the large Latin American immigration to the US from the 1960s onwards.

The results of the analysis lead to the conclusion that the settlement pattern of the predominantly European migrants at the turn of the 19th to the 20th century strongly influenced the geography of Latino migration to the US a century later. Counties which received large population inflows of European settlers at the end of the 19th century remain significantly more appealing to incoming population from Latin America today than those which were largely bypassed by migrant streams in the 1900s. This result is robust to introducing a large number of controls and estimating the analysis with different econometric techniques. The differences in the composition of migrants between both waves and a considerable time gap has not altered what is a highly persistent geography of migration. Latino migrants have followed in their European predecessors' footsteps 60 to 130 years later, creating a path dependency in migration that is stronger than any historical factor that may have shaped the prosperity of a given county in the US. Historical migration stocks act as important pull factor for subsequent generations of migrants in a more consistent way than many of the standard regional pull factors. Past income levels, historic unemployment rates or levels of education have not left a trace shaping future migration; past migration has. A county's historical foreign-born migrant stock acts as a long-lasting magnet for future generations of migrants.

The results stress that past migration in itself, beyond national or kinship networks, represents an important and long-lasting pull factor for future migrants. Migrants rework the territory where they settle in large numbers. Late 19th century European migrants left a legacy in specific parts of the US which survives until today and still appeals to new generations of migrants. The migration buzz generated in historic times by Germans, Italians, Poles, or Irish settlers has become engraved in the receiving territories and persists well after the 'original' migrants are long gone and their descendants have been assimilated and become American.

The results of the analysis point to factors that have hitherto not been considered in migration policy and provide interesting food for thought for future migration and development policy design. Policies crafted to attract migrants into a specific region need to take into account not only the region's economic and socio-demographic characteristics but also the territory's migration history. Areas with a limited background in immigration history may lack an important pull factor for new migrants: 'the buzz of past migration'. A lack of 'migrant buzz' would therefore have to be overcome by other means – i.e. better local amenities or additional economic incentives – in order to attract a similar magnitude of incoming migration.

However, the mechanisms through which path dependent migration mechanisms are created remain a mystery. Further research is needed in this respect. In-depth analyses of migration-prone areas, such as New England, California, or the north west of the US, can provide crucial insights about the mechanisms that make regions consistently migration-prone. Similarly, studies of new destination areas without an important history of immigration, such as the South of the US, could lead to eye-opening results as to what factors can help overcome a lack of history in migration. Furthermore, there is a need to pinpoint what is exactly behind migration path dependence and a 'regional migration buzz'. In-depth, more anthropological case studies can play an important role in understanding how this buzz is created, how it reproduces itself, and how exactly it affects the attractiveness of a place to newly arriving migrants.

References

Alexander, J. (2007): *Daily life in immigrant America: 1870–1920*. London: Greenwood Press.

Baltagi, B. (2008): *Econometric analysis of panel data*. 4th edn. Chichester, UK: John Wiley & Sons.

Baltagi, B., Bresson, G. and Pirotte, A. (2003): Fixed effects, random effects or Hausman–Taylor? A pretest estimator, *Economics Letters*, *79*(3), 361–369.

Bauer, T., Epstein, G. and Gang, I. (2007): The influence of stocks and flows on migrants' location choices, *Research in Labor Economics*, *26*, 199–229.

Bauer, T., Epstein, G. and Gang, I. (2009): Measuring ethnic linkages among migrants, *International Journal of Manpower*, *30*(1/2), 56–69.

Bergad, L. and Klein, H. (2010): *Hispanics in the United States: a demographic, social, and economic history, 1980-2005*. Cambridge: Cambridge University Press.

Bertocchi, G. and Strozzi, C. (2006): *The age of mass migration: Economic and institutional determinants*, Discussion paper Series, IZA Institute for the Study of Labour, 2499. Bonn, Germany.

Biagi, B. and Dotzel, K. R. (2018). Theoretical advances on interregional migration models. In B. Biagi, A. Faggian, I. Rajbhandari, & V.A. Venhorst. *New Frontiers in Interregional Migration Research* (pp. 21-47). Heidelberg: Springer.

Biagi, B., Faggian, A. and McCann, P. (2011): Long and short distance migration in Italy: The role of economic, social and environmental characteristics, *Spatial Economic Analysis*, *6*(1), 111–131.

Bikhchandani, S., Hirshleifer, D. and Welch, I. (1998): Learning from the behaviour of others: Conformity, fads, and informational cascades, *The Journal of Economic Perspectives*, *12*(3), 151–170.

Bodvarsson, Ö., Simpson, N. and Sparber, C. (2014): Migration Theory, in Chiswick, B. and Miller, P. (eds) *Handbook of the Economics of International Migration,* 3–51. Burlington: Elsevier Science.

Bruhn, J. (2005): *The sociology of community connections*. Boston, MA: Springer US.

Bump, M. (2005): From temporary picking to permanent plucking: Hispanic newcomers, integration, and change in the Shenandoah Valley, in Goździak, E. M. and Martin, S. F. (eds) *Beyond the Gateway: Immigrants in a Changing America,* 137–177. Lanham: Lexington Books.

Cafferty, P. and Engstrom, D. (2000): *Hispanics in the United States: an agenda for the twenty-first century*. New Brunswick: Transaction Publishers.

Carrington, W., Detragiache, E. and Vishwanath, T. (1996): Migration with endogenous moving costs, *American Economic Review*, *86*(4), 909–930.

Clark, X., Hatton, T. and Williamson, J. (2007): Explaining U.S. immigration, 1971–1998, *The Review of Economics and Statistics*, *89*(2), 359–373.

Daniels, R. (1990): *Coming to America: A history of immigration and ethnicity in American life*. New York: Harper Perennial.

DaVanzo, J. (1983). Repeat migration in the United States: who moves back and who moves

on? *Review of Economics and Statistics*, 552-559.

Day, K. (1992): Interprovincial migration and local public goods, *Canadian Journal of Economics*, *25*(1), 123–144.

Donato, K., Tolbert, C., Nucci, A. and Kawano, Y. (2008): Changing faces, changing places: The emergence of new nonmetropolitan immigrant gateways, in Massey, D. S. (ed.) *New Faces in New Places,* 75–98. New York, NY: Russell Sage Foundation.

Douglas, K. and Sáenz, R. (2008): No phone, no vehicle, no English, and no citizenship: The vulnerability of Mexican immigrants in the United States, in Hattery, A., Embrick, D. G., and Smith, E. (eds) *Race, Human Rights and Inequality,* 161–180. Lanham: Rowman and Littlefield.

Epstein, G. and Gang, I. (2006): The influence of others on migration plans, *Review of Development Economics*, *10*(4), 652–666.

Faggian, A. and McCann, P. (2008): Human capital, graduate migration and innovation in British regions, *Cambridge Journal of Economics*, *33*(2), 317–333.

Ferguson, M., Ali, K., Olfert, M.R. and Partridge, M. (2007): Voting with their feet: Jobs versus amenities, *Growth and Change*, *38*(1), 77–110.

Greenwood, M. (1997): Chapter 12 Internal migration in developed countries, *Handbook of Population and Family Economics*, *1*, 647–720.

Hansen, A. (2005): Black and white and the other: International immigration and change in metropolitan Atlanta, in Goździak, E. M. and Martin, S. F. (eds) *Beyond the Gateway: Immigrants in a Changing America,* 87–110. Lanham: Lexington Books.

Hatton, T. and Williamson, J. (1998): *The Age of Mass Migration: Causes and economic impact*. New York: Oxford University Press.

Hicks, J. (1932): *The theory of wages*. London: Macmillan.

Hoover, E.M. (1971). *An Introduction to Regional Economics*, New York: Alfred A. Knopf.

Jaeger, D. (2007): Green Cards and the Location Choices of Immigrants in the United States, 1971–2000, *Research in Labor Economics*, *27*, 131–183.

Jewell, R. and Molina, D. (2009): Mexican migration to the US: A Comparison of income and network effects, *Eastern Economic Journal*, *35*(2), 144–159.

Joly, D. (2000): Some structural effects of migration on receiving and sending countries, *International Migration*, *38*(5), 25–40.

Ketterer, T. and Rodríguez-Pose, A. (2015): Local quality of government and voting with one's feet, *The Annals of Regional Science*, *55*(2), 501–532.

Liu, Y., Song, W. and Deng, X. (2016). Changes in crop type distribution in Zhangye City of the Heihe River Basin, China. *Applied Geography*, *76*, 22-36.

Massey, D., Alarcon, R., Durand, J., and Gonzalez, H. (1987): *Return to Aztlan: The social process of international migration from western Mexico*. Berkeley: University of California Press.

Massey, D., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., and Taylor, J. (1993): Theories of International Migration: A Review and Appraisal, *Population and Development Review*, *19*(3), 431–466.

Massey, D., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A. and Taylor, J. (2005):

Worlds in motion: Understanding international migration at the end of the millennium. Oxford: Oxford University Press.

Massey, D. and Capoferro, C. (2008): The geographic diversification of American immigration, in Massey, D. S. (ed.) *New Faces in New Places,* 25–50. New York: Russell Sage Foundation.

Massey, D. and España, F. (1987): The Social Process of International Migration, *Science*, *237*(4816), 733–738.

McGovern, P. (2007): Immigration, labour markets and employment relations: Problems and prospects. *British Journal of Industrial Relations*, *45*(2), 217–235.

McKenzie, D. and Rapoport, H. (2010): Self-Selection patterns in Mexico-U.S. migration: The Role of migration networks, *The Review of Economics and Statistics*, *92*(4), 811–821.

Migration Policy Institute (2017): *U.S. immigrant population and share over time, 1850-Present*. Available at: http://www.migrationpolicy.org/programs/data-hub/us-immigration-trends#history (Accessed: 28 June 2017).

Minnesota Population Center (2016): *National Historical Geographic Information System*, Version 11.0 [Database], Minneapolis: University of Minnesota.

Munshi, K. (2003): Networks in the modern economy: Mexican migrants in the US labor market, *The Quarterly Journal of Economics*, *118*(2), 549–599.

Ödberg, S. (1997): Theories in inter-regional migration: An overview, in Blotevogel, H. and Fielding, A. (eds) *People, jobs and mobility in the new Europe,* 3–22. Chichester: Wiley.

Ottaviano, G. and Puga, D. (1998): Agglomeration in the global economy: A survey of the "New Economic Geography", *World Economy*, *21*(6), 707–731.

Özgen, C., Nijkamp, P. and Poot, J. (2011): Immigration and innovation in European regions, *IZA Discussion Papers*, Institute for the Study of Labor (IZA), No. 5676. Bonn, Germany.

Partridge, M. (2010): The duelling models: NEG vs amenity migration in explaining US engines of growth, *Papers in Regional Science*, *89*(3), 513–536.

Partridge, M. and Rickman, D. (2008): Distance from urban agglomeration economies and rural poverty, *Journal of Regional Science, 48*(2), 285–310.

Partridge, M. and Rickman, D. (1996): The role of industry structure, costs, and economic spill-overs in determining state employment growth rates, *The Review of Regional Studies*, *26*(3), 235–264.

Pissarides, C.A. and McMaster, I. (1990): Regional migration, wages and unemployment: Empirical evidence and implications for policy, *Oxford Economic Papers*, *42*(4), 812–831.

Radu, D. (2008): Social Interactions in economic models of migration: A review and appraisal, *Journal of Ethnic & Migration Studies*, *34*(4), 531–548.

Rodríguez-Pose, A. and von Berlepsch, V. (2014): When migrants rule: The Legacy of mass migration on economic development in the United States, *Annals of the Association of American Geographers*, *104*(3), 628–651.

Rodríguez-Pose, A. and von Berlepsch, V. (2015): European migration, national origin and long-term economic development in the United States, *Economic Geography*, *91*(4), 393–424.

Rodríguez-Pose, A. and von Berlepsch, V. (2019). Does population diversity matter for economic development in the very long term? Historic migration, diversity and county wealth in the US. *European Journal of Population*, *35*(5), 873-911.

Rodríguez-Pose, A., Ketterer, T. and Castells-Quintana, D. (2015): Do we follow the money? The drivers of migration across regions in the EU, *Region*, *2*(2), 27–45.

Rodríguez-Pose, A. and Ketterer, T. (2012): Do local amenities affect the appeal of regions in Europe for migrants?, *Journal of Regional Science*, *52*(4), 535–561.

Ruggles, S., Genadek, K., Goeken, R., Grover, J. and Sobek, M. (2015): *Integrated Public Use Microdata Series: Version 6.0 [dataset]*. Minneapolis: University of Minnesota.

Sáenz, R. and Morales, M.C. (2015): *Latinos in the United States: Diversity and change*. Cambridge: Polity Press.

Shah, N. and Menon, I. (1999): Chain migration through the social network: Experience of labour migrants in Kuwait, *International Migration*, *37(2*), 361–382.

Sharma, A. (2012). Exploratory spatial data analysis of older adult migration: A case study of North Carolina. *Applied Geography*, *35*(1-2), 327-333.

Simpson, N. and Sparber, C. (2013): The short- and long-run determinants of less-educated immigrant flows into U.S. States, *Southern Economic Journal*, *80*(2), 414–438.

Sjaastad, L. (1962): The Costs and returns of human migration, *Journal of Political Economy*, *70*(5, Part 2), 80–93.

Spickard, P. (2007): *Almost all aliens: immigration, race, and colonialism in American history and identity*. New York: Routledge.

Staiger, D. and Stock, J. (1997): Instrumental Variables regression with weak instruments, *Econometrica*, *65*(3), 557–586.

Stock, J. and Yogo, M. (2005): Testing for weak instruments in linear IV regression, in Donald, A. and Stock, J. (eds) *Identification and Inference for Econometric Models*. New York: Cambridge University Press.

Taylor, P., Hugo Lopez, M., Hamar Martinez, J. and Velasco, G. (2012): *When labels don't fit: Hispanics and their views of identity*. Pew Research Center Hispanic Trend Project, Washington, D.C.

Treyz, G., Rickman, D., Hunt, G. and Greenwood, M. (1993): The dynamics of U.S. internal migration, *The Review of Economics and Statistics*, *75*(2), 209-214.

U.S. Census Bureau (2016): *Population Estimates Program (PEP)*. Population and Housing Unit Estimates. Washington, D.C.

U.S. Department of Homeland Security (2011): *Yearbook of Immigration Statistics: 2010*. Washington, D.C.

Yap, L. (1977): The attraction of cities: A review of the migration literature, *Journal of development economics*, *4*(3), 239–264.

Zimmermann, K. (2005): European labour mobility: Challenges and potentials, *De Economist*, *153*(4), 425–450.

Appendix A

Figure A1. Foreign-born population as share of total population, 1990

Foreign Stock (1990)
0% - 1.5%
1.6% - 3%
3.1% - 6%
6.1% - 12%
12.1% - 50%
no data

0   350   700   1,400 km
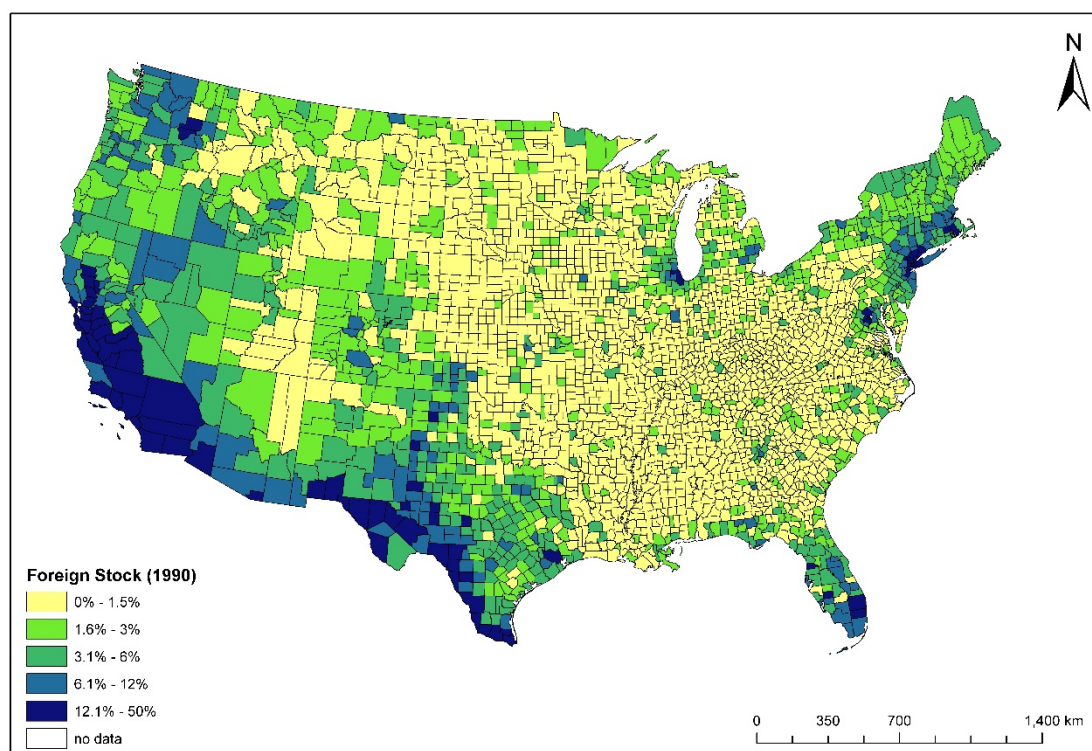
Data source: Minnesota Population Center (2016); own elaboration.

Appendix B

Table B1. Variables descriptions and sources

Variable	Description	Source
Main variables of interest		
Latinos	Share of foreign-born Latino migrants relative to total population in county i in year t^*. Latino is defined if originating from Mexico, the Caribbean, and Central and South American countries.	NHGIS
Migration 1880 / 1910	Share of foreign-born population relative to total population in county i in year t_0^*	IPUMS USA
Instruments		
Population density	Population per square mile in county i in year 1880	ICPSR
Public good endowment	Share of employment in health, education, and public administration relative to total employment in any given county i in 1880 (or 1910)	IPUMS USA
Control Variables		
Income	Measure for county income: • Historical years: Constructed mean income score on county level. Individual income levels assigned to occupational data on the basis of median total income per occupation in hundreds of 1950 dollars, as natural log in t_0 • 1960-2000: Natural log of average income per capita levels in 2010 US dollars on county level for county i (adjusted for inflation using the BLS converter)	IPUMS USA BEA and Social Explorer
Urban share	Share of urban population relative to total population in county i in year t_0 or $t-k$, respectively	ICPSR and Social Explorer
Education	Percentage of population of county i with college degree in $t-k$	Social Explorer

Variable	Description	Source
Literacy	Literacy rate in county i in t_0	IPUMS USA
Unemployment	Unemployment rate in county i in t_0 and t-k Proxy in 1880 (unemployment rate not available): Months unemployed last year	IPUMS USA and ICPSR
Agriculture	Percentage of labour force employed in agriculture in county i in t_0 and t-k	IPUMS USA and ICPSR
Female Participation	Female participation rate in the labour force in county i in t_0 and t-k	IPUMS USA and ICPSR
Female	Percentage of female population in county i in t_0 and t-k	Social Explorer
Black Population	Percentage of black population in county i in t_0 and t-k	Social Explorer
DistMX (ln)	Distance in km between county i's centroid and the nearest point on the continental border between the US and Mexico (as natural log)	Own construction
State	State dummies	Own construction

*t_0= 1880 or 1910 while t = 1970, 1980, 1990, 2000, 2010. k= 10

Appendix C

Table C1. The impact of historical migration on Latino settlement patterns, OLS 1910

Dep. Var: % of Latinos in county	(1) 1970	(2) 1980	(3) 1990	(4) 2000	(5) 2010
Migration 1910	0.0159***	0.0224***	0.0279***	0.0384***	0.0370***
	(0.00460)	(0.00778)	(0.00969)	(0.0122)	(0.00822)
Income pc (ln)~	0.00121	0.00597***	0.00340	0.0471***	0.0553***
	(0.00112)	(0.00229)	(0.00302)	(0.00690)	(0.00462)
Urban share~	0.00320***	0.00554***	0.0130***	0.0315***	0.0483***
	(0.00122)	(0.00181)	(0.00240)	(0.00382)	(0.00281)
Education~	-0.000553	-0.00174	-0.00409	-0.0679***	-0.137***
	(0.00296)	(0.00542)	(0.00687)	(0.0103)	(0.00942)
Unemployment~	0.0290**	0.0465***	0.0214	0.125***	0.225***
	(0.0122)	(0.0138)	(0.0136)	(0.0299)	(0.0290)
Empl. in agri~	0.0134***	0.0210***	0.0310***	0.0760***	0.144***
	(0.00221)	(0.00362)	(0.00577)	(0.0125)	(0.0119)
Fem. labforce~	0.00500	0.0111*	0.0238**	0.00323	0.0730***
	(0.00394)	(0.00625)	(0.00958)	(0.00807)	(0.0140)
Female pop~	-0.0171	0.00528	-0.0491**	-0.222***	-0.248***
	(0.0106)	(0.0122)	(0.0240)	(0.0401)	(0.0312)
Black pop~	-0.00282	0.000354	0.00579	0.00645	-0.0159**
	(0.00176)	(0.00292)	(0.00385)	(0.00622)	(0.00676)
Distance to MX (ln)	-0.0241***	-0.0380***	-0.0431***	-0.0467***	-0.0378***
	(0.00229)	(0.00365)	(0.00411)	(0.00483)	(0.00204)
Income (ln) 1910	-7.00e-05	-0.00250	1.51e-05	0.00301	0.00282
	(0.000899)	(0.00174)	(0.00224)	(0.00341)	(0.00328)
Urban share 1910	0.00192	0.00374*	0.00327	0.00289	-0.000822
	(0.00138)	(0.00223)	(0.00305)	(0.00488)	(0.00395)
Education 1910	-0.00453	-0.000543	0.00847	0.0157*	0.0385***
	(0.00334)	(0.00560)	(0.00716)	(0.00889)	(0.00795)
Unemployment1910	-0.00304	-0.00178	-0.00476	-0.0158	-0.0273
	(0.00623)	(0.00949)	(0.0123)	(0.0168)	(0.0186)
Empl. in agri 1910	-0.00312**	-0.00195	0.00313	0.00963**	0.00905*
	(0.00135)	(0.00246)	(0.00312)	(0.00465)	(0.00465)
Fem. labforce 1910	-0.00422	-0.00538	-0.0125*	-0.0310***	-0.0214*
	(0.00301)	(0.00527)	(0.00710)	(0.0104)	(0.0115)
Female pop 1910	0.0168	0.000381	0.0145	0.0566	0.0143
	(0.0115)	(0.0213)	(0.0270)	(0.0367)	(0.0325)
Black pop 1910	0.00583***	0.00476*	0.00321	-0.0123	0.000612
	(0.00174)	(0.00257)	(0.00407)	(0.00770)	(0.0123)
States	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	3,118	3,120	3,120	3,120	3,119
R-squared	0.639	0.614	0.573	0.514	0.523

~ Controls from 10 years prior to dependent variable

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1