



CENTRO STUDI LUCA D'AGLIANO

WWW.DAGLIANO.UNIMI.IT

CENTRO STUDI LUCA D'AGLIANO
DEVELOPMENT STUDIES WORKING PAPERS

N. 445

February 2019

**From Immigrants to Americans: Race and Assimilation during
the Great Migration**

*Vasiliki Fouka**
*Soumyajit Mazumder***
*Marco Tabellini****

* Stanford University
** Harvard University
*** Harvard Business School

ISSN 2282-5452

From Immigrants to Americans: Race and Assimilation during the Great Migration

Vasiliki Fouka[†] Soumyajit Mazumder[‡] Marco Tabellini[§]

January 2019

Abstract

How does the appearance of a new out-group affect the economic, social, and cultural integration of previous outsiders? We study this question in the context of the first Great Migration (1915-1930), when 1.5 million African Americans moved from the US South to urban centers in the North, where 30 million Europeans had arrived since 1850. We test the hypothesis that black inflows led to the establishment of a binary black-white racial classification, and facilitated the incorporation of – previously racially ambiguous – European immigrants into the white majority. We exploit variation induced by the interaction between 1900 settlements of southern-born blacks in northern cities and state-level outmigration from the US South after 1910. Black arrivals increased both the effort exerted by immigrants to assimilate and their eventual Americanization. These average effects mask substantial heterogeneity: while initially less integrated groups (i.e. Southern and Eastern Europeans) exerted more assimilation effort, assimilation success was larger for those that were culturally closer to native whites (i.e. Western and Northern Europeans). These patterns are consistent with a framework in which changing perceptions of out-group distance among native whites lower the barriers to the assimilation of white immigrants.

JEL Codes: J11, J15, N32.

Keywords: Immigration, assimilation, Great Migration, race, group identity.

*We gratefully acknowledge financial support from the Russell Sage Foundation. We thank Nicola Genaioli, David Laitin, Salma Mousa, Agustina Paglayan, Imran Rasul, Ken Scheve, Alain Schlaepfer, Guido Tabellini, Hans-Joachim Voth and participants at the Zurich Workshop on the Origins and Consequences of Group Identities, the LSE Historical Political Economy Conference, the UC Irvine Workshop on Identity, Cooperation and Conflict, the 2018 NBER Summer Institute Political Economy Meeting, the Barcelona GSE Summer Forum Workshop on Migration, as well as seminar participants at LSE, UCL, Durham, Yale, Harvard, MIT and Stanford for helpful comments and suggestions. Valentin Figueroa provided excellent research assistance.

[†]Stanford University, Department of Political Science. Email: vfouka@stanford.edu

[‡]Harvard University, Department of Government. Email: smazumder@g.harvard.edu

[§]Harvard Business School. Email: mtabellini@hbs.edu

1 Introduction

Our tendency to categorize ourselves into social groups is perhaps one of the most fundamental characteristics of human societies (Tajfel, 1974, 1979). In-group identity is an important driver of cooperation, successful collective action and the maintenance of smaller and larger organized units, such as teams, firms or states (Goette, Huffman and Meier, 2006; Chen and Li, 2009; Hjort, 2014). At the same time, identification with one's group can lead to parochialism, discrimination, and violence (Bernhard, Fischbacher and Fehr, 2006; Choi and Bowles, 2007; Hargreaves Heap and Zizzo, 2009; Sambanis and Shayo, 2013). Though group categories are fundamental to social life, we know relatively little about the ways in which group boundaries are determined and change over time (Gaertner and Dovidio, 2000). Under what conditions do individuals who belong to out-groups at the bottom of a social hierarchy make their way up to the top and into the in-group?

This paper addresses this question in the context of US history. In the late 19th and early 20th century, the US attracted close to 30 million European immigrants. Immigration and immigrant integration were as much a concern and topic of political debate at the time as they are today (Vigdor, 2010). Nativism and anti-immigrant attitudes were widespread, especially towards Eastern and Southern Europeans, who were religiously and culturally different from native Anglo-Saxons. Yet 20th century immigrants assimilated economically and culturally at a fast pace (Abramitzky, Boustan and Eriksson, 2014, 2018) and were soon fully integrated into American society, fueling the myth of the American melting pot. In this study, we test the idea that an important catalyst in this rapid assimilation and redrawing of in-group boundaries was the arrival of another group, perceived as even more distant by the native whites – African Americans. From 1915 to 1930, approximately 1.6 million blacks migrated for the first time from the southern United States to the industrializing urban centers of the North and West, in a movement that was termed the First Great Migration. In areas receiving these new migrants, the changing racial profile of the population changed the perceptions of the local native whites. Skin color replaced ethnicity as the most salient social category and previously discriminated European immigrants, by virtue of their whiteness, were allowed entry into the native majority group.

To demonstrate this empirically, we use information on the universe of foreign-born individuals in the US living in non-southern metropolitan statistical areas (MSAs) between 1910 and 1930. These areas, collectively, received almost the entire population of African Americans who migrated from the South to the North during this period. We examine two types of indicators: proxies of increased assimilation efforts on the part of immigrants, such as naturalization rates and naming patterns, and equilibrium outcomes, such as intermarriage rates, that depend both on immigrants' efforts to fit in and on the barriers to assimilation erected by the natives. Figure 1 presents our main findings. Accounting for time-invariant MSA characteristics and region-specific shocks, and relying on plausibly exogenous variation for black

in-migration discussed in detail below, the figure shows that black inflows were associated with higher naturalization rates and a higher likelihood of marriage to a native-born spouse of native parents. These effects are quantitatively large. An inflow of black migrants such as that experienced by Detroit increased the share of naturalized immigrants by 5 percentage points, or 10 percent relative to the 1910 mean, and raised the probability of intermarriage between immigrants and natives by 1.7 percentage points, or 24 percent relative to the 1910 mean. Black migration also induced foreign-born parents to choose more American-sounding names for their children. An inflow of blacks amounting to half that received by Chicago led to a name like Luciano being abandoned in favor of one like Mike, and a name like Stanislav to be replaced by Max. Alongside social assimilation, immigrants in MSAs receiving many black migrants were also more likely to leave the manufacturing sector and experience occupational upgrading – patterns consistent with economic assimilation.

The key econometric challenge to our analysis is that both black and foreign-born migrants might have been attracted by similar MSA characteristics that may have directly favored (or hindered) assimilation. To address this and similar concerns, we construct a “shift-share” instrument (Card, 2001) that assigns estimated black outflows from southern states to northern MSAs based on settlement patterns of African Americans in 1900, more than 15 years before the onset of the Great Migration. These predicted migration flows strongly correlate with actual black migration to the North, but are more plausibly orthogonal to any omitted variables that may drive both immigration and assimilation patterns among the foreign-born.

The validity of the instrument relies on one, crucial identifying assumption: MSAs that received more blacks (from each southern state) before 1900 should not have been on differential trends in terms of the evolution of economic, political, and social conditions that could have also affected immigrant assimilation after 1910. There are at least three threats to identification.

First, the (fixed) characteristics of MSAs that attracted early blacks might have had persistent, confounding effects on migration patterns as well as on changes in the outcomes of interest. It is possible, for instance, that larger urban centers attracted more African Americans already in the nineteenth century, and that these areas kept growing more also in subsequent decades, introducing a spurious correlation between, e.g. economic activity and the Great Migration. We address this first concern by performing a large number of checks, such as testing for pre-trends, interacting year dummies with several 1900 MSA characteristics, and documenting that the instrument is uncorrelated with the pre-period change in European immigration.

Second, one may worry that the location decisions of African Americans were correlated with changes in immigration patterns induced by the immigration quotas of the 1920s. To deal with this possibility, we show that the instrument is orthogonal to the local exposure to the immigration quotas, as predicted by the distribution of pre-existing immigrant enclaves across MSAs (Ager and Hansen, 2017). We also document that the instrument is uncorrelated with changes in the average number of years spent by immigrants in the US and has no direct effect on either the number or the national composition of immigrants.

Third, the identifying assumption would be violated if outmigration from each southern state were correlated with cross-MSA pull factors systematically related to 1900 settlers' state of origin. We address this potential concern in two ways. First, we show that neither the strength of the instrument nor our 2SLS results are affected when interacting year dummies with the share of blacks born in each southern state. Second, similar to [Boustan \(2010\)](#), we replicate our results using a modified version of the instrument, which exploits only variation across southern push factors to predict black outflows across sending states.

The second part of the paper explores the channels through which the Great Migration affected immigrant assimilation. We formalize the idea that the appearance of a more "distant" out-group, such as African Americans were for white natives living in the US North in the early 1920s, can facilitate the integration of other out-groups (i.e. European immigrants) by making them seem closer, in relative terms, to the white majority. Our simple model delivers two testable predictions that we take to the data. First, it predicts an inverted U-shaped relationship between the effect of black inflows on assimilation and measures of distance between natives and immigrants. Second, it predicts that the effect of black inflows on assimilation effort is largest for groups that are relatively distant to the natives, while the effect on successful assimilation is largest for the least distant groups.

We find empirical support for these predictions, and the social mechanism that generates them. When considering continuous measures of immigrants' distance from natives (such as linguistic or genetic distance), and when extending the analysis to non-European immigrants, we observe that both assimilation effort and, to a lesser extent, actual assimilation exhibit an inverted U-shaped pattern: very distant or non-white groups either did not respond to or did not benefit from black arrivals. When examining the heterogeneity of the effects by national origin, we find that naturalization rates increased the most among immigrants from new source regions, such as Southern and Eastern Europe. This is consistent with assimilation efforts responding more for groups that previously experienced discrimination and faced increased chances of assimilating once black arrivals shifted natives' perceptions of relative distance. At the same time, however, assimilation efforts did not mechanically translate into successful assimilation. Intermarriage rates with natives increased the most for the immigrant groups that were closest to native whites, such as the English, the Western and the Northern Europeans.

We provide additional evidence in support of the main assumption of our model. Using the frequency of terms expressing anti-immigrant sentiment in historical newspapers, we show that MSAs that experienced larger black inflows exhibited larger reductions in concerns about immigration. More importantly, we find evidence of a reduction in the stereotyping of large immigrant groups, such as the Irish and the Italians. Not only did these nationalities become less associated with negative stereotypes, such as criminality or alcohol abuse, but they also became less likely to be perceived as Catholics. This set of results is again supportive of our interpretation that the Great Migration reduced the importance of features such as religion, which differentiated immigrants from natives, and in turn also reduced prejudice against European immigrants.

The predictions of the model differ from those generated by other possible mechanisms, thus allowing us to arbitrate between different alternative explanations for observed assimilation patterns. We find some support for higher labor market competition between immigrants and blacks driving immigrants' increased assimilation effort. However, this mechanism is not consistent with the observation that immigrants exerting more efforts were less likely to assimilate. In particular, the effect of black inflows on naturalization rates is larger for immigrant groups that were more likely to be employed in manufacture and in unskilled occupations in 1900. These groups were arguably the most exposed to black competition, and thus had higher incentives to signal Americanization. However, successful assimilation was not higher for immigrant groups exerting more effort, suggesting that our main result cannot be entirely explained by competition.

We provide evidence against the possibility that labor market complementarities between immigrants and African Americans led to the economic and social advancement of the former. If this were to be the case, we would expect the effect of the Great Migration on economic outcomes to be most pronounced for groups exhibiting the highest complementarity with blacks (Peri and Sparber, 2009; Foged and Peri, 2016). On the contrary, we do not find evidence of heterogeneity in economic outcomes, such as employment in manufacturing or the native-immigrant gap in occupational income scores.

Finally, we consider the possibility that our results are due to political channels. We find no evidence of job patronage, with black inflows having no effect on the employment of immigrants in the public sector. This weighs against the possibility that politicians facilitated the incorporation of immigrants in order to enlarge white majorities in areas receiving more African Americans. Similarly, using data on local labor unions, we find little evidence of increased political mobilization of immigrants in response to competition from African Americans.

The historical context we focus on offers a number of advantages. The First Great Migration – though smaller in magnitude than the migration of African Americans that took place after World War II – constituted a massive change in racial composition, and the first instance of such a change for the northern United States. Furthermore, it happened at a time when immigration from Europe was a highly salient topic in the political arena. Concerns about immigration and immigrant assimilation substantially subsided after the introduction of the 1924 quotas (Spiro, 2008). This period thus allows us to examine the behavior and the outcomes of the large number of immigrants who arrived to the US prior to the immigration restrictions of the 1920s, at a time when nativist sentiment was likely a hindrance to those immigrants' integration. This time frame also allows us to use rich historical data, such as the newly digitized full count of the US census, not available in the post-WWII period.

Our study builds on ideas from social psychology, and especially self-categorization theory (Turner et al., 1987), which studies how individuals classify themselves and others in groups. A central tenet of the theory is that classification follows the meta contrast principle, whereby objects are classified as part of a group if differences between them are smaller than the difference between the group of objects and other groups (Turner et al., 1994; McGarty, 1999).

This principle explains why native whites might have reclassified European immigrants as in-group members once the number of African Americans increased. A related strand of research, the common in-group identity model (Gaertner and Dovidio, 2000), predicts that the salience of a superordinate group identity – in this case being “white” – can reduce intergroup bias. We show that re-categorization and prejudice reduction can have important implications for the social and economic outcomes of the groups being re-categorized.

A set of theoretical papers in economics study the cognitive process of group classification. Fryer and Jackson (2008) provide a model of categorical thinking in which classification of objects into clusters follows the rule of within-cluster variance minimization. Re-classification of immigrants as “whites” is consistent with such a rule. Gennaioli and Tabellini (2018) rely on the meta contrast principle to construct a model of political identities. Bordalo et al. (2016) show that group stereotypes are context dependent. When the reference group changes, stereotypes are more likely to be defined on the dimension that displays the largest difference across groups. In our case, skin color replaces religion and language as a relative dimension once African Americans appear as part of the out-group.

More broadly, our study contributes to a large literature on in-group and out-group biases starting with Tajfel et al. (1971), and in particular to a smaller strand of this literature that examines spillovers of biases across multiple groups. In a recent paper, McConnell and Rasul (2018) show that increased animosity towards Muslims spurred by 9/11 had a negative impact on attitudes towards other minority groups, such as Hispanics, as evidenced by decisions in the Federal Criminal Justice system. Their paper provides evidence of *contagious animosity* – hostility against an out-group leading to increased hostility against other out-groups.¹ The results of our study are instead more supportive of *parochial animosity*, with hostility towards an out-group (blacks) increasing altruistic preferences towards other out-groups (immigrants).² However, compared to related studies, our setup is novel. Instead of examining how a shock to preferences towards one out-group affects preferences towards other groups, we exploit the appearance of a new out-group to examine how that affects both the preferences of the in-group, and the resulting outcomes for out-group members

Our study also contributes to a growing literature in economics that formalizes ideas from sociology, and models the formation and transmission of cultural and ethnic identities (Akerlof and Kranton, 2000; Bisin and Verdier, 2001; Shayo, 2009). While many of these studies consider a binary division of society into a majority and a minority group, we show that the presence of multiple minorities and the composition of the minority group can play a crucial role in the formation of in-group identity.

The existing literature on immigrant assimilation is vast and has identified a number of

¹Similarly, a set of studies based on public goods games provide evidence of contagious altruism (Fowler and Christakis, 2010; Suri and Watts, 2011; Jordan et al., 2013).

²Empirical studies that have found evidence of parochial altruism, show that out-group hostility increases in-group identity, but not how it affects preferences towards other out-groups (Bauer et al., 2016).

determinants of integration and its speed (Watkins, 1994), including immigrant group size (Shertzer, 2016), ethnic networks (Edin, Fredriksson and Åslund, 2003), as well as education and other government policies (Lleras-Muney and Shertzer, 2015; Fouka, 2018; Bandiera et al., forthcoming; Mazumder, 2018). Yet, to our knowledge, there has been no comprehensive quantitative study of the causal effect of race and its salience on immigrant outcomes. While a substantial scholarship has examined the interaction between different ethnic and racial groups, it has mostly focused on competition as a driver of inter-group conflict and prejudiced attitudes (McAdams, 1995; Olzak and Shanahan, 2003; Olzak, 2013). Our paper shows how the interaction of minorities can change the white majority’s perceptions about in-group boundaries and be a driver of assimilation.

Finally, a large literature in economic history examines the economic effects of the Great Migration, primarily focusing on white flight, black and white economic outcomes, and, more recently, city finances (Boustan, 2010; Collins and Wanamaker, 2014; Boustan, 2016; Shertzer and Walsh, 2016; Tabellini, 2018b). Our study borrows methodological techniques from this literature to extend the analysis of the impact of the migration of southern blacks to social and cultural outcomes. Moreover, we are the first to examine the effects of the Great Migration on European immigrants, a group that was as large as 25% of the population of several northern cities during the period of reference. We show that by inducing lower status groups to assimilate, the Great Migration had effects beyond those on native whites, and that the assimilation of immigrants in response to black arrivals may have been an additional factor, beyond racial segregation, that reinforced racial stratification.

The paper proceeds as follows. Section 2 provides a historical overview of the Great Migration, immigrant assimilation and the interaction of race and ethnicity in the first quarter of the 20th century. Section 3 presents our data and empirical strategy. Section 4 shows that black inflows encouraged Americanization effort and assimilation for immigrants. Section 5 turns to the mechanisms driving this effect. It provides a simple model relating the appearance of a new out-group to the assimilation of other out-groups, tests the main implications of the model empirically, and discusses alternative explanations for the main results. Section 6 concludes.

2 Historical background

2.1 The first Great Migration

Outmigration of African-Americans from the US South started during World War I, largely triggered by the war-induced increase in industrial production and demand for industrial labor in northern urban centers. Between 1915 and 1919, more than 2 million jobs – most of them requiring minimal levels of skills – were created in northern cities, thereby increasing labor market opportunities for blacks (Boustan, 2016). These pull factors were not unrelated to European immigration. The 1921 and 1924 immigration quotas restricted the pool of available low-skilled industrial workers, especially Southern and Eastern Europeans, and allowed African Americans to substitute for the foreign-born in the industrial sector. Alongside pull

factors in the North, a number of push factors in southern states drove black outmigration during this period. Natural disasters such as the 1927 Mississippi flood (Boustan, Kahn and Rhode, 2012; Hornbeck and Naidu, 2014), and shocks to agricultural production such as the Boll Weevil infestations that destroyed cotton crops in the late 19th century (Lange, Olmstead and Rhode, 2009), negatively impacted the demand for labor in the agricultural sector, where most blacks were employed. Added to these economic factors, racism and violence in the South provided an additional migration incentive to the black population (Tolnay and Beck, 1990).

Taking advantage of the newly constructed railroad network, close to 1.5 million blacks moved from the South to the North between 1915 and 1930 (Black et al., 2015; Boustan, 2010), with the fraction of blacks living in the North rising from 10% to 25% in the same period. The unprecedented inflow of African Americans and the induced change in the racial landscape of northern cities triggered mounting hostile reactions by white residents. As described in Massey and Denton (1993), during the early 1920s, whites used to coordinate to racially segregate blacks and to bomb their houses. Boustan (2010) and Shertzer and Walsh (2016) show, respectively for the second and for the first wave of the Great Migration, that uncoordinated actions, such as the “white flight”, were as important as formal and coordinated ones for the rise of the American ghetto. Specifically, whites often reacted to black inflows by leaving racially segregated cities and neighborhoods. Historical accounts suggest that such hostile reactions against African Americans tended to be stronger among recently arrived immigrants, especially Eastern and Southern Europeans, possibly because these groups were more directly exposed to the competition, both in the labor and in the housing market, brought about by blacks (see Rieder, 1987; Cho, 1993, among others).

2.2 Immigrant assimilation during the era of Mass Migration

Between 1850 and 1920, during the Age of Mass Migration, no restrictions to European immigration to the US existed, and approximately 30 million immigrants – two thirds of the total migration out of Europe – moved to the US, increasing the share of the foreign-born from 10% in 1850 to 14% in 1920 (Abramitzky and Boustan, 2017). The composition of those immigrant inflows underwent large changes during the period. In 1870 almost 90% of the foreign born came from the British Isles, Germany, and Scandinavia. By 1920, in contrast, the share of migrant stock from Southern and Eastern Europe had climbed to 40%.

Europeans from new regions were culturally more distant from natives, and significantly less skilled than those from old sending regions (Hatton and Williamson, 1998). They were also younger, more likely to be male, and less likely to permanently settle in the US. This typical immigrant profile suggests that immigrants from new sending regions likely had lower incentives and faced higher barriers to assimilation. Indeed, return migration prior to 1920 is estimated to have been 30% or higher (Bandiera, Rasul and Viarengo, 2013), and fell only after the imposition of the 1924 quotas, which induced a dramatic change in the composition

of the foreign born, in favor of old sending regions³

Until recently, immigrants in the early 20th century US were thought to face substantial occupational earnings penalties upon arrival, but to rapidly converge with the natives after 15-20 years in the country (Chiswick, 1978). Works by Borjas (1987) and, more recently, by Abramitzky, Boustan and Eriksson (2014) show instead that, when accounting for compositional changes associated with new arrivals, immigrants did not experience substantial labor market assimilation, and their gap from natives persisted in the second generation. While the immigrant-native gap was not large for immigrants with similar skills as natives, there was wide heterogeneity by country of origin. Abramitzky, Boustan and Eriksson (2018) also show substantial, though far from complete, cultural assimilation, with immigrants choosing more American-sounding names for their children over the course of their stay in the US (similar patterns were observed for intermarriage and citizenship outcomes).

A potential explanation of why immigrants failed to narrow the gap with natives over time despite efforts to assimilate culturally could be the fact that they faced substantial barriers to assimilation in the form of prejudice and discrimination. These were most often directed toward new immigrant arrivals. Though the Irish, Italians, and Eastern Europeans were phenotypically white, their social status was in many respects that of an inferior race (Guglielmo, 2003). The Irish were often derided as “white niggers” – a sign that natives viewed these individuals with similar disdain as African Americans (Ignatiev, 1995). Discrimination against immigrants was not confined only to the private sphere. The state played an important role in cultivating prejudice against foreigners. As Hochschild and Powell (2008) and Guglielmo (2003) point out, the US government made formal distinctions between different immigrant groups, in turn allowing both local governments and native citizens to engage in discrimination against immigrants on the basis of pseudo-scientific evidence.

2.3 Race and Americanization in the early 20th century

An extensive historical literature suggests that the Great Migration catalyzed the assimilation of immigrants and substantially contributed to their Americanization. One factor emphasized throughout this literature is the role of changing perceptions of native whites toward racial boundaries in propelling this assimilation (Ignatiev, 1995; Guterl, 2001). With the arrival of African Americans, skin color became a salient determinant of racial distinctions, and this eased the path to inclusion into the native majority for white Europeans.

The early decades of the 20th century were dominated by academic theories about race and eugenics that emphasized fine grained racial distinctions among the various European groups. Madison Grant, the author of the opus magnum of scientific racism, *The Passing of*

³With the 1924 National Origins Act, the total number of immigrants that could be admitted in a given year was capped at 150,000. In 1921, quotas were specified to reflect the 1910 composition of immigrants. However, they were rapidly changed to reflect that of 1890 in order to limit immigration from new sending countries even further (Goldin, 1994).

the Great Race, and one of the intellectuals behind the design of the immigration restrictions of the 1920s, proclaimed Americans to be “Nordics”, the race of “the Homo Europaeus, the white man par excellence” (Spiro, 2008). Below the Nordic man, in the hierarchy of races, followed the Alpines and Mediterraneans – the color of the former being described as “fair to dark”, and that of the latter as “swarthy”. Grant and his followers were worried that the Nordic type in the US was being “elbowed out of his own home” and “literally driven off the streets of New York City by the swarms of Polish Jews.” Grant’s views were highly influential among academic circles, but also beyond them. Hiram Wesley Evans, the imperial wizard and emperor of the Knights of the Ku Klux Klan, cited Grant to warn that the Jews, like the Catholics and the Negroes, were “an absolutely unblendable element...alien and unassimilable” (Spiro, 2008).

The Great Migration changed these dynamics and shifted the focus, both of academics and of society at large, from ethnic differences to color as a racial group identifier. Lothrop Stoddard, another prominent eugenicist and Klansman, and author of the best-seller *The Rising Tide of Color against White World Supremacy*, emphasized how color-coding race would lead to assimilation and unification of ethnic and cultural differences in the US. At the same time, race riots in northern cities contributed to the framing of blacks as the primary social threat – the emerging “Negro problem” (Guterl, 2001). In place of the pre-existing multitude of racial and ethnic classifications, the northwards migration of African Americans redrew racial distinctions along the black-white line, which now included Southern, Eastern, and other European ethnic groups as part of the white majority (Jacobson, 1999). The salience of race reduced ethnic prejudice, and made it easier for immigrants to assimilate into the US society.

3 Data and empirical strategy

3.1 Data

To examine how the Great Migration affected immigrant assimilation, we use data from the full count of the US census for the period 1900 to 1930. We restrict our analysis to the 108 metropolitan statistical areas (MSAs) outside the US South with a positive number of European-born residents in 1900 – a requirement imposed by the construction of the instrument, as explained in detail in Section 3.3 (see Figure 2 and Table A.1 for the complete list of MSAs in our sample). We focus on MSAs, rather than on counties or cities, for two reasons. First, the majority of black migrants settled in urban areas during the first Great Migration. Second, black inflows had a quantitatively large impact on the residential decision of native whites, often triggering a “white flight” from central cities to suburban rings. As demonstrated by Shertzer and Walsh (2016) for the first, and Boustan (2010) for the second Great Migration, whites largely relocated to suburbs within the same MSA, so that, while city population and its composition changed substantially in response to black migration, such changes were substantially less pronounced at the MSA level.

We distinguish between assimilation effort provided by immigrants and actual assimilation.

The latter is an equilibrium outcome, which depends on the actions of both immigrants and natives. Our main proxy for social assimilation is intermarriage, arguably a good measure of both effort and acceptance, defined by [Gordon \(1964\)](#) as “the final stage of assimilation”. We measure intermarriage as an indicator for an individual married with a native-born spouse of native parentage.⁴ In the Appendix we also study the effects of black inflows on economic assimilation, measured as employment in the manufacturing sector. As noted above, most African Americans were employed in occupations with minimum skill requirements and were concentrated in manufacturing. These were precisely the types of jobs held by many European immigrants, especially from new sending regions. Instead, native whites were significantly more likely than the foreign born to work in the trade sector and in occupations with higher skill requirements.⁵ For this reason, we interpret a reduction in the share of immigrants working in manufacturing as (economic) assimilation.

Our main outcome aimed at capturing assimilation effort is naturalization rates. In 1906, the path to citizenship for immigrants was standardized by the Bureau of Immigration and Naturalization, and most naturalization cases were handled by federal courts. Immigrants would usually file a Declaration of Intention (known as “first papers”) upon arrival or shortly thereafter. Within five years, they were eligible to file a Petition for Naturalization (“second papers”), which was the last step required before the court finalized the naturalization process. While rates of naturalization reflect both the decision of immigrants to obtain citizenship and the decision of the courts to grant it to them, rejection rates of petitions were very low in practice.⁶

In the Appendix we consider an auxiliary measure of effort: the decision of immigrants to give a foreign-sounding name to their children. Since it involves their offspring and not immigrants themselves, this is a less direct signal of Americanization than an application for citizenship. However, unlike intermarriage or other equilibrium measures of assimilation, the naming choice is fully under the control of the parents. Furthermore, to the degree that parents are attached to their culture, choosing a non-ethnic name for one’s children is a costly signal of assimilation. Several studies show that there is a labor market penalty associated with foreign-sounding names ([Biavaschi, Giuliotti and Siddique, 2017](#); [Algan, Mayer and Thoenig, 2013](#)). If immigrant parents are aware of this – and extensive name Americanization among immigrants to the US with the aim of reaping economic benefits indicates that they are ([Biavaschi, Giuliotti and Siddique, 2017](#)) – then this penalty can proxy for the monetary value they assign to their children having a name indicative of their ethnic origin.

To capture the ethnic content of names, we compute an index of name distinctiveness that

⁴In the Appendix we also consider endogamous marriages, measured as an indicator for a spouse, or a spouse’s parents, born in the same country of origin as the immigrant.

⁵See [Tabellini \(2018a\)](#) for a more detailed discussion of the difference in occupations held by natives and immigrants in US cities between 1900 and 1930.

⁶In a sample of approximately 3,300 naturalization petitions filed in New York City in 1930, [Biavaschi, Giuliotti and Siddique \(2017\)](#) find that only 2.6% were rejected.

was first used by Fryer and Levitt (2004), and more recently by Abramitzky, Boustan and Eriksson (2018) and Fouka (2018, forthcoming). The index measures the frequency of a name within an ethnic group relative to its frequency in the population at large. It is computed as follows:

$$FNI_{name,c} = \frac{Pr(name|Nationality_c)}{Pr(name|Nationality_c) + Pr(name|Other_c)} * 100$$

A value of zero implies that a name is never found among individuals belonging to a national group, while a value of 100 implies instead that a name is never encountered among other nationalities (including natives). We construct this index for US-born children of a foreign-born father using the full count data for years between 1910 and 1930. The subscript c denotes a birth cohort. For each year of birth, the information used for the computation of the index comes only from people born before that year. The aim is to capture what parents perceived as a foreign or American-sounding name when they made their naming decisions, without contamination from changes in naming patterns in later generations. Figure A.1 in the Appendix plots mean values of the Foreign Name Index in 1910 for second-generation immigrants in our sample by father’s country of origin. There is substantial variation in the ethnic distinctiveness of names, with countries like Italy, Romania, and Russia having some of the most distinctive first names.

Table 1 presents summary statistics for MSA-level characteristics (e.g. total, black, and immigrant population), and for our two main outcome variables described above aggregated at the MSA level. Table A.2 in the Appendix presents individual-level summary statistics for main and auxiliary outcomes. As it appears, only 7.6 percent of immigrant men were married with a native of native parentage, while almost 78 percent of them had a spouse born in the same country of origin. Table A.2 also reveals that one in four immigrants worked in the manufacturing sector, and at least 41 percent of them were unskilled. These numbers are significantly higher than for natives: only 15 (resp. 30) percent of natives were working in the manufacturing (resp. unskilled) sector. This pattern suggests that both the social and economic assimilation of immigrants were far from complete. At the same time, 52 percent of immigrants in our sample were naturalized, which may indicate instead a high desire to assimilate, potentially curbed by barriers to assimilation erected by the natives.⁷

⁷To put this number into perspective, 44 percent of immigrants in the US today are naturalized. If one considers undocumented immigrants, this is likely an upper bound of the actual share of foreign born who have obtained the US citizenship today. See <http://www.pewresearch.org/fact-tank/2018/01/18/naturalization-rate-among-u-s-immigrants-up-since-2005-with-india-among-the-biggest-gainers/>.

3.2 Difference-in-differences

3.2.1 Repeated Cross-Sections

Our basic research design is based on repeated cross-sections of individuals living in the 108 non-southern MSAs listed in Table [A.1](#) in the three Census years between 1910 and 1930. In particular, we examine how immigrant assimilation responds to changes in the black population, accounting for time-invariant MSA characteristics and for common time-variant shocks. Formally, stacking the data for the three decades between 1910 and 1930, we estimate the following equation:

$$Y_{int} = \alpha_n + t_t + \beta_1 B_{nt} + \beta_2 Pop_{nt} + \mathbf{X}'_{int} + u_{int} \quad (1)$$

where Y_{int} is the outcome for foreign born individual i living in non-southern MSA n in Census year t , and B_{nt} is the number of blacks living in MSA n in year t . We always include MSA and year fixed effects (α_n and t_t), and in our preferred specification, we also control for interactions between year dummies and region dummies as well as for a number of individual level controls (such as geographic region of origin, age, and years in the US) collected in the vector \mathbf{X}_{int} in [\(1\)](#).⁸ Notably, controlling for MSA and region by year fixed effects implies that β_1 is estimated from changes in the number of blacks within the same MSA over time, as compared to other MSAs in the same region in a given year. Finally, following [Boustan \(2010\)](#), since growing areas might attract both African Americans and European immigrants, we control for total MSA population, Pop_{nt} . Standard errors are clustered at the MSA level.

3.2.2 Linked Panel Dataset

Throughout our analysis, we also report results obtained from a panel of immigrants linked across census years. It is possible that any effect of the Great Migration on immigrant assimilation found in the repeated cross-sections may be due to compositional changes in the immigrant population. Previous work has demonstrated the effect of black migration on the “white flight” ([Boustan, 2010](#); [Shertzer and Walsh, 2016](#)). Black inflows could have similarly led to a selective out-migration of more (or less) assimilated immigrants from the MSAs in our sample. A linked panel dataset deals with this problem because it allows us to track the same individuals over time, identifying their assimilation trajectory.

In addition to dealing with possible compositional effects, focusing on comparing the same immigrant over time is also desirable when considering outcomes such as marriage or

⁸When defining regions, we follow the Census Division classification. We classify immigrants into eleven countries or country groupings. These are: Northern Europe (Denmark, Finland, Iceland, Norway and Sweden), UK (England, Scotland and Wales), Ireland, Western Europe (Belgium, France, the Netherlands and Switzerland), Southern Europe (Albania, Greece, Italy, Portugal and Spain), Central and Eastern Europe (Austria, Bulgaria, Czechoslovakia, Hungary, Poland, Romania and Yugoslavia), Germany, Russian Empire (Russia, Estonia, Latvia and Lithuania), Mexico, China and Canada.

naturalization which are “absorbing states”. Once an individual obtains citizenship, he does not go back to non-citizen status, and a similar argument holds for intermarriage, with divorce rates before 1930 being lower than 1 percent. A panel dataset allows us to restrict attention to those immigrants actually likely to respond to the Great Migration, i.e. those who were not already naturalized or married at the time of black arrivals.

To construct the linked panel dataset, we start with the universe of foreign-born men in 1910 who are unique by first and last name, birthplace and year of birth. Following standard automated census-linking procedures used in the economic history literature (Ferrie, 1996; Abramitzky, Boustan and Eriksson, 2014), we match men to a target census year using the phonetic equivalent of their first and last name, their birthplace and their year of birth (allowing for a two-year band around the recorded year) and we discard any records with multiple matches.

Our final dataset consists of individuals linked across all three decades from 1910 to 1930 and who were always observed to reside in one of our 108 sample MSAs. Importantly for the internal validity of results derived using the linked sample, the match rate is uncorrelated to the change in our instrument between 1910 and 1930, as shown in Figure A.2.⁹ While a selected sample of all male immigrants, this linked dataset of non-movers has the advantage of allowing us to observe the dynamics of assimilation across the entire period of focus. Table A.3 in the Appendix compares the demographic and socioeconomic characteristics of records that were and were not linked across census years. While significant, differences between the two groups are small. Immigrants in the linked panel have lived in the US one year longer on average, are more likely to be naturalized, and more likely to be employed in manufacture. Intermarriage rates in the two groups are comparable and differences in endogamy rates are small. Finally, the English and Western Europeans are over-represented in the linked panel compared to other immigrant origin regions. The last three columns of Table A.2 present summary statistics of our main and auxiliary outcomes for the linked dataset.

3.3 Instrument for black population

The northwards movement of African Americans was largely dictated by economic conditions in northern cities. Those same conditions also likely affected the location choices of foreign migrants as well as their assimilation patterns. A priori, it is not clear whether these omitted factors introduce a positive or negative bias. On the one hand, blacks may have been attracted to areas with better job opportunities, or with more appealing tax-public spending bundles, that also favored the social and economic assimilation of the foreign-born. Similarly, OLS estimates may be biased upwards if blacks moved to cities where European immigrants were better able to mobilize, and where their political clout was stronger. On the other hand, both African Americans and European immigrants may have settled in otherwise declining MSAs,

⁹The t-statistic of the underlying regression is -1.14 and drops to 0.61 when the outlier observation of Oklahoma City, OK is excluded.

where house prices were lower and prospects for integration less bright.

Additionally, around the time of our study, the introduction of the literacy test (1917) and, more importantly, of immigration quotas (1921 and 1924) drastically reduced immigration flows to the US (see [Goldin, 1994](#)). It is conceivable that more African Americans moved to parts of the US North where the impact of the quotas was larger, to cover the needs in the low-skilled workforce created by the immigration restrictions ([Collins, 1997](#)). If the reduction in the number of incoming migrants facilitated the assimilation of immigrants already in the US, there could be a spurious correlation between the effect of the quotas and black migration.

To isolate the causal effect of the Great Migration on immigrant assimilation, we construct an instrument for the location decision of black migrants using a version of the “shift-share” instrument commonly adopted in the immigration literature ([Card, 2001](#)). This instrument exploits two sources of variation: first, cross-sectional variation in the distribution of settlements of African Americans living outside the South in 1900, who were born in different southern states. Second, time-series variation in the number of blacks who left the South from each state over time.

Because data on internal migration do not exist before 1940, we estimated migration rates from each southern state in each decade using the forward survival method ([Gregory, 2005](#))¹⁰. Specifically, using data for the United States as a whole, we first computed survival ratios for each age-sex-race group, and then, relying on the latter, we estimated net migration from each southern state. Next, we predicted the number of blacks received by each northern MSA in any given year by interacting the estimated number of migrants with the share of southern born African Americans from each state living in each MSA in 1900. Formally, the predicted number of blacks moving to MSA n in year t is given by

$$Z_{nt}^{MOV} = \sum_{j \in South} \alpha_{jn}^{1900} O_{jt} \quad (2)$$

where α_{jn}^{1900} is the share of blacks born in southern state j residing in the non-South who were living in MSA n in 1900, and O_{jt} is the number of African Americans born in state j who left the South between $t - 1$ and t . Since we are interested in instrumenting a stock, i.e. the total number of blacks in the MSA, B_{nt} , we recursively sum Z_{nt}^{MOV} (i.e. a flow):

$$Z_{nt} = \sum_{s=1}^t Z_{ns}^{MOV} \quad (3)$$

where Z_{nt}^{MOV} is given by [\(2\)](#).

As for other works in the literature ([Card, 2001](#); [Boustan, 2010](#)), the instrument constructed above is based on the empirical regularity that settlement patterns of blacks were

¹⁰For robustness, we compared our measure of estimated outmigration with that computed in [Lee et al. \(1957\)](#). The correlation between the two measures is 0.93.

highly persistent over time. As discussed in [Black et al. \(2015\)](#) and [Boustan \(2010\)](#) among others, the railroad network was key in determining the location decision of early migrants. The stability of settlement patterns was further reinforced by chain migration: more recent migrants tended to move where other migrants from the same county (or state) had moved in the past (see [Wilkerson, 2010](#), among others). In addition to stability over time, settlement patterns of migrants also exhibited wide variation across both receiving MSAs and sending states. That is, even before 1900, several MSAs had received blacks from many southern states, and each southern state had sent migrants to a number of different MSAs.

Figure [3](#) plots the share of southern born blacks from selected states living in a number of northern MSAs in 1900, and confirms visually that there was indeed variation across both sending and receiving places.

3.3.1 Identifying Assumptions and Instrument Validity

The key assumption behind the instrument is that MSAs receiving more blacks (from each southern state) before 1900 must not follow differential trends in the evolution of political, economic, and social conditions that could affect immigrant assimilation after 1910. This assumption can be violated for at least three reasons. First, if the (fixed) characteristics of MSAs that attracted early blacks had persistent, confounding effects on migration patterns as well as on changes in the outcomes of interest. It is possible, for instance, that larger urban centers attracted more African Americans already in the nineteenth century, and that these areas kept growing more also in subsequent decades, introducing a spurious correlation between, e.g. economic activity and the Great Migration. Similarly, one may be worried that the industry mix of MSAs affected both the location decision of early settlers and subsequent changes in economic and political conditions.

To deal with these and similar issues, we perform two sets of robustness checks, which we describe in detail when presenting our results. First, we show that the pre-period (1900-1910) change in the outcomes of interest is uncorrelated with subsequent changes in black population predicted by our instrument. Second, we augment our baseline specification by including interactions between year dummies and several 1900 MSA characteristics, such as the fraction of blacks, the fraction of European immigrants, and value added by manufacturing. Controlling for the interaction between 1900 fraction of blacks and year dummies is particularly important because it implies that the effects of black in-migration are identified exploiting variation only in the (southern state) composition of African Americans' enclaves across MSAs, holding constant the size of their black populations.

Second, one may be concerned that the instrument is spuriously correlated with changes in the immigration regime induced by the Immigration Acts of the 1920s. While the quotas were introduced at the national level, they likely had a differential effect across MSAs depending on pre-existing ethnic composition ([Collins, 1997](#); [Ager and Hansen, 2017](#)). We directly tackle this issue by checking that our instrument is not correlated with local exposure to the quotas, as predicted by the distribution of pre-existing immigrant enclaves across MSAs. Moreover, we provide evidence that our instrument is uncorrelated with changes in the average length

of stay of immigrants in the US, and has no direct effect on either the number or the national composition of immigrants.

Third, the identifying assumption would be violated if outmigration from each southern state were not independent of cross-MSA pull factors systematically related to 1900 settlers' state of origin. We address this concern, recently formalized in [Goldsmith-Pinkham, Sorkin and Swift \(2018\)](#) and [Borusyak, Hull and Jaravel \(2018\)](#), by examining the degree to which the instrument depends on variation coming from flows from a specific state to specific MSAs. Reassuringly, both the strength of the instrument and our main results are unchanged when separately interacting year dummies with the share of blacks born in each southern state, i.e. α_{jn}^{1900} in [\(2\)](#). Note that, even if early settlements were as good as randomly assigned, one remaining concern is that black outflows from each southern state, O_{jt} , might be differentially affected by specific, time-varying shocks in northern destinations. To deal with this potential threat, as in [Boustan \(2010\)](#), we construct a modified version of the instrument in [\(3\)](#) by replacing O_{jt} with predicted (rather than actual) outmigration. We describe this strategy in detail in the Appendix Section [B](#), and only briefly review the main steps in the next paragraph.

First, we predict outmigration from southern counties by exploiting only local demographic and agricultural conditions at the beginning of each decade. Next, we aggregate these flows to the state level to obtain the predicted number of blacks leaving each southern state j in each decade, \hat{O}_{jt} . Finally, we replace O_{jt} with \hat{O}_{jt} in [\(2\)](#) to derive the (push-factors induced) predicted number of blacks moving to city c in year t . By construction, this (predicted) measure of black outmigration from the South is orthogonal to any specific shock occurring in the North. Moreover, by exploiting southern shocks to agricultural conditions, this alternative instrument is less likely to suffer from the problem of high serial correlation in migration patterns between sending and receiving areas – a possible concern for standard shift-share instruments ([Jaeger, Ruist and Stuhler, 2018](#)). Below, we show that both first stage and 2SLS results are robust to using this, instead of our baseline version of the instrument.

3.4 First Stage

Table [2](#) reports first stage results for the relationship between the actual number of blacks and the instrument constructed in [\(3\)](#). Column 1 controls for total MSA population, and includes MSA and year fixed effects. Column 2 presents our preferred specification, and augments the set of controls included in column 1 by interacting year dummies with region dummies. As it appears, there is a strong and positive relationship between the instrument and the number of blacks, and the F-stat is above conventional levels. A coefficient as in column 2 implies that every predicted new black arrival in the MSA is associated with 1.1 more actual black residents. These estimates are very similar to those reported in [Shertzer and Walsh \(2016\)](#) and in [Tabellini \(2018b\)](#) for the same historical period, for neighborhoods and cities, respectively. They are instead an order of magnitude smaller than those obtained in [Boustan \(2010\)](#), who focuses on the second wave of the Great Migration. Figure [4](#) plots the graphical analogue of the regression estimated in column 2, and verifies the strong relationship between the actual and the predicted number of blacks.

Subsequent columns of Table 2 explore the robustness of results reported in column 2. First, we show that our estimates are not sensitive to running unweighted regressions (column 3).¹¹ Next, we augment our baseline specification by interacting year dummies with, respectively, the 1900 fraction of blacks, the 1900 fraction of immigrants, the log of 1900 output in manufacturing, and the fraction of men aged 15–64 employed in manufacturing (columns 4 to 7). In all cases, the coefficient remains quantitatively close to that reported in column 2, and its statistical significance is not affected.

In Appendix Figure C.3 we plot the first stage coefficient for regressions that include, respectively, interactions between year dummies and the share of blacks born in each southern state, i.e. α_{jn}^{1900} in equation (2). Reassuringly, the point estimate always remains highly significant and very similar to that obtained from our baseline specification (Table 2, column 2), which is the first point estimate on the left in Figure C.3.

Finally, Table A.4 replicates the exercise in Table 2 using the version of the instrument that relies on southern push factors to predict net black migration rates. While smaller in magnitude, the relationship between actual and predicted number of blacks is always strong and statistically significant, and the F-stat remains high.

Overall, the pattern presented in this section suggests that there is a strong relationship between the actual and the predicted number of blacks, which is robust to the inclusion of several controls and the use of alternative specifications.

4 Results

In this section we present our main results for the effects of the Great Migration on immigrant assimilation. First, the inflow of blacks increased successful assimilation, measured as intermarriage between immigrants and natives and occupational upgrading (Section 4.1). Second, black in-migration raised the share of immigrants who were naturalized citizens, and induced foreign born parents to give more American sounding names to their children, suggesting that immigrants responded to black arrivals by increasing their assimilation efforts (Section 4.2). We conclude by performing several robustness checks (Section 4.3).

4.1 Social and economic assimilation

In Panel A of Table 3, we study the effects of the Great Migration on equilibrium measures of assimilation. Our most preferred proxy for assimilation success is intermarriage between immigrants and natives, which captures not only immigrants’ desire to Americanize, but also natives’ willingness to accept immigrants as part of their group. Restricting attention to immigrant men, we define the dependent variable as a dummy equal to one for being married with a native of native parentage. We start by estimating equation (1) with OLS: column

¹¹In our main analysis we estimate individual level regressions. This is equivalent to running MSA-level regressions, weighted by the number of immigrants.

1 only includes MSA and year fixed effects, and controls for a battery of individual level characteristics (age, origin region, and years in the U.S. fixed effects), whereas column 2 also includes year by region fixed effects. In both cases, the point estimate is positive, but close to zero and not statistically significant.

From column 3 onwards, we present 2SLS results. Column 3 replicates column 2 instrumenting the number of blacks with the shift-share instrument introduced in Section 3.3 above. The coefficient is now larger in magnitude and statistically significant at the 5% level. The downward bias in the OLS estimates indicates that black migrants may have selected into MSAs in which the prospects for immigrant assimilation were not that bright. The point estimate in column 3 implies that one standard deviation increase in the number of blacks (approximately 45,000 people) increases intermarriage rates by 0.54 percentage points, or 7.5% of the 1910 mean. For a large recipient city like Chicago, that received close to 230,000 blacks during the period, this effect amounts to 2.74 percentage points, or 57.1% of the 1910 mean.

In columns 4 and 5, we gradually add a more stringent set of controls – respectively, MSA by region of origin, and year by region of origin fixed effects – but, reassuringly, both the magnitude and the precision of the coefficient are left unchanged. Finally, in column 6, we present results for the linked sample of immigrants who always stayed in the same MSA between 1910 and 1930. Results remain qualitatively in line with those reported in columns 3 to 5, but become larger in magnitude and less precisely estimated¹²

Our main results focus on social assimilation. In the Appendix, as an additional proxy for the successful integration of immigrants, we examine economic status. In Panel A of Table A.5, the dependent variable is a dummy equal to one for working in manufacturing, which was the “immigrant-intensive” sector at the time. Both OLS and 2SLS estimates are strongly negative and statistically significant in all our specifications and samples. Unlike the case of intermarriage, OLS coefficients are less negative than the corresponding 2SLS ones. That the direction of the OLS bias is not the same for social and economic assimilation is not entirely surprising – blacks were most likely attracted by economically booming MSAs, in which immigrants were able to experience occupational upgrading. At the same time, it was those same MSAs that attracted more new migrants and potentially saw the formation of migrant enclaves (Eriksson and Ward, 2018), factors that could have hindered the social integration of immigrants.

Immigrants were almost twice as likely as natives to be employed in manufacturing during the period under study. We thus interpret these findings as reflecting economic assimilation, with foreign born men moving out of the “immigrant sector”. The effects of the Great Migration are once again economically relevant: the coefficient in column 3 of Table A.5

¹²In the Appendix, we also show that a similar pattern, albeit noisier, arises for endogamous marriage (Table D.1, column 1): the inflow of African Americans is associated with a decline in the probability of marrying a spouse born in or originating from the same country.

(Panel A) implies that one standard deviation increase in the number of African Americans lowers the share of immigrants working in the manufacturing sector by 2.5 percentage points, or 10% relative to the 1910 mean.

In Appendix [D](#) we consider a number of additional outcomes capturing economic and social assimilation. Confirming the results on manufacturing employment, we find a reduction in the share of immigrants who were unskilled, again suggesting that immigrants were able to improve their socio-economic status because of the Great Migration (Column 2 of Table [D.1](#)). The inflow of blacks did not substantially impact immigrants' employment probabilities, suggesting that, if there was labor market competition between immigrants and blacks, its effects were muted, perhaps as a result of immigrants' socio-economic advancement (Column 3 of Table [D.1](#)).¹³ Black inflows reduced ethnic residential segregation, but these effects are small and never precisely estimated.

4.2 Assimilation Effort

We next analyze immigrant effort. In Panel B of Table [3](#), the dependent variable is a dummy equal to one for being a naturalized citizen, our main proxy for immigrant assimilation effort. As in Panel A, columns 1 and 2 report OLS results, while subsequent columns present 2SLS estimates from the repeated cross-sections (columns 3 to 5) and the linked (column 6) datasets. In all cases, the point estimate is positive and statistically significant, suggesting that black inflows increased the effort exerted by the foreign-born to become part of the society's main in-group, i.e. that of native whites. According to our most preferred specification, reported in column 5, one standard deviation increase in black population raises naturalization rates by approximately 1.5 percentage points, or by 3.5% relative to the 1910 mean. When using the linked sample (column 6), the effect becomes more precisely estimated and substantially larger, suggesting that, if anything, results obtained from the repeated cross-sections underestimate the effect of the Great Migration on immigrants assimilation.

As an alternative proxy for immigrant effort, in Panel B of Appendix Table [A.5](#), we consider the names chosen by immigrant parents for their children. We use the logarithm of the foreign name index as dependent variable to facilitate the interpretation of coefficients in percentage terms. In line with results reported in Table [3](#), there is a negative and statistically significant effect of black inflows on the foreign name index, which is robust to the inclusion of all sets of controls.¹⁴ The magnitude of the effect is substantial. It implies that the inflow of 100,000 blacks – or, less than half of those received by Chicago – led to a change in Italian names equivalent to that from Luciano to Mike, and a change in Russian names equivalent to

¹³It is also possible that black workers were substitutes for some immigrant groups but complements to others, and that the null effects reported in Table [D.1](#) were due to opposite effects which, on average, canceled each other out. In Section [5](#) we return to this point and explore the heterogeneity of the economic impact of black inflows across immigrant groups.

¹⁴We do not construct the index in the panel dataset, which consists of foreign-born individuals whose names were decided in their country of origin.

that from Stanislav to Morris or Max.

Overall, results in this section suggest that the arrival of a new out-group, i.e. African Americans, induced foreign born individuals to exert more effort to assimilate and become part of the majority in-group. Coupled with the findings of Section 4.1, our analysis indicates that the arrival of blacks fostered immigrant Americanization, either because their increased assimilation effort was by and large successful, or because black arrivals additionally lowered pre-existing barriers to immigrant integration, such as discrimination by native whites. As we show in Section 5 below, these average effects mask substantial heterogeneity, and suggest that the reduction in native whites’ prejudice against European immigrants played a major role in immigrant assimilation. Before discussing the mechanisms, in the next section, we summarize a number of checks that test the robustness of our main results.

4.3 Summary of Robustness Checks

We summarize here the robustness checks we conduct to address concerns regarding the validity of our identification strategy. A detailed description of these checks can be found in Section C of the Appendix.

To show that 1900 black settlements are unlikely to be correlated with time-varying characteristics of MSAs that could have affected assimilation patterns we perform three checks: (i) we show that the 1900 to 1910 change in European immigration is uncorrelated with predicted black inflows between 1910 and 1930 (Figure C.1), (ii) we formally demonstrate the absence of pre-trends for our outcome variables (Table C.1), and (iii) we show that our results are robust to interacting year dummies with a number of 1900 MSA characteristics, including the 1900 share of blacks at the MSA (Table C.2).

To specifically tackle the concern that blacks moved to northern MSAs more affected by the 1920s immigration quotas – and that this spurious correlation is not dealt with by our instrument – we construct a measure of “quota exposure” for each MSA, by interacting the immigration restrictions (for each immigrant group) with pre-existing settlements in the MSA (Ager and Hansen, 2017). Using this variable, we document that our instrument is uncorrelated with the number of “missing” immigrants that an MSA would have received had immigration restrictions not been introduced (Figure C.2). We also verify that black in-migration is uncorrelated with changes in immigrants’ average length of stay in the US (Table C.3).

We then turn to the possibility that our results might be driven by compositional changes in the immigrant population, triggered by black arrivals. The robustness of our results to the use of the linked sample already mitigates this concern. We perform three additional checks: we show that black inflows did not affect (i) the number of international immigrants in the MSA (Table C.4), (ii) the ethnic composition of immigrants, measured as shares of different origin regions over total MSA foreign population (Table C.5), or (iii) sex ratios within the immigrant group, either for younger or for older immigrants (Table C.6).

To address concerns related to the validity of Bartik instruments, as detailed in Goldsmith-Pinkham, Sorkin and Swift (2018), we show that both our first stage (Figure C.3) and 2SLS

results (Figure C.4) are robust to interacting year dummies with the share of southern born blacks from each state, i.e. the Bartik weights. Next, to show that time-specific shocks in northern MSAs are unlikely to have driven outmigration flows from the South, we replicate our results using a push version of the instrument, following Boustan (2010) (Table C.7). Finally, we show that neither first stage nor 2SLS results are sensitive to the exclusion of outliers (Figures C.5 and C.6).

5 Mechanisms

Section 4 showed that, along important social and economic dimensions, black in-migration fostered the assimilation of European immigrants. We find evidence both for increased immigrant efforts and for increased assimilation as observed in equilibrium outcomes. What mechanism generated these effects? Historians have suggested channels that are social in nature. The arrival of African Americans made European immigrants, who were previously viewed as members of a social and cultural out-group, appear “white” in the eyes of natives (Ignatiev, 1995; Jacobson, 1999). This, in turn, allowed their inclusion into the native majority.

In this section, we provide evidence for this mechanism. We do so in three steps. First, we formalize the intuition of historians in a simple model. Starting from the assumption that the Great Migration reduces the perceived distance between natives and immigrants, the model delivers testable predictions on the type of heterogeneity the assimilation patterns of different immigrant groups should be expected to display. In the second step, we take the predictions of the model to the data. Finally, we explore alternative mechanisms that could have generated our main empirical result of higher assimilation. We address, in particular, the possibility that higher assimilation was driven by economic competition and labor market complementarities between immigrants and African Americans. We show that these alternatives are inconsistent with the patterns we observe empirically.

5.1 A simple theoretical framework

We build a simple theory, relying on core findings of the social psychology literature on categorization (Turner et al., 1987), in order to formalize the effects that black inflows have on native attitudes and immigrant outcomes. Our goal is not to explain the cognitive processes that reduced native prejudice towards immigrants after the Great Migration. Instead, we assume that a cognitive mechanism led natives to re-categorize immigrants as members of the in-group, and explore the implications of that assumption, first theoretically and then empirically. All proofs of propositions are in Appendix Section E.1.

Setup

The population consists of two groups, an in-group and an out-group. For our purposes we define the in-group to be the group containing native whites of Anglo-Saxon origin, and the out-group to be everyone else. Prior to the Great Migration, the out-group consists of

immigrants of European descent, but later on is expanded to include African Americans. Building on a large literature in social psychology documenting the existence of consensual ethnic and racial hierarchies in social distance in multiethnic and multiracial societies (Berry and Kalin, 1979; Duckitt, 1992; Hagendoorn, 1995), we define h to be the social distance of out-group members from the group of native whites. h can be thought of as a unidimensional summary measure of distances along multiple dimensions, such as skin color, language, culture or religion. Native whites are share n of the population and have $h = 0$. Out-group members are a share $1 - n$ of the population and are distributed on the line $[0, H^{max}]$.

We assume that native whites engage in taste-based discrimination. In particular, we assume that interactions with individuals who are distant to them ($h > 0$) induce a psychological cost for native whites. This cost depends on *perceived distance* $\lambda(h, \bar{h})$, where \bar{h} is the average distance of the out-group from the in-group. \bar{h} can be thought of as the degree of difference between a native white and the members of the out-group he is faced with *on average* in his daily life or in his neighborhood. As such, it captures the degree of white natives' familiarity with more distant "outsiders". Perceived distance is increasing in actual distance and so $\lambda_h > 0$ and $\lambda_{\bar{h}} < 0$. Furthermore, $\lambda(0, \bar{h}) = 0$ for any \bar{h} .

Crucially, we assume that $\lambda_{h, \bar{h}} \leq 0$, since higher average familiarity with more distant out-group members makes an outsider of any given distance appear more similar to white natives. This assumption is one of the core tenets of self categorization theory, known as the *meta contrast principle* (Turner et al., 1987, 1994). Categorization of stimuli into groups is context-dependent. Humans are more likely to classify a collection of stimuli as a single grouping if differences between those stimuli are smaller than differences between the grouping that they form and other groupings. In our case, an immigrant is more likely to be classified as a member of the in-group if the difference in social distance between the immigrant and the natives is smaller than the difference between the average distance of the in-group and the out-group (\bar{h}).¹⁵

The role of relative distance is also consistent with the historical narrative on the progressive incorporation of European immigrants into the white Anglo-Saxon majority. In his study of this process, for example, Jacobson (1999) states that "In racial matters above all else, the eye that sees is 'a means of perception conditioned by the tradition in which its possessor has been reared.' The American eye sees a certain person as black, for instance, whom Haitian or Brazilian eyes might see as white. Similarly, an earlier generation of Americans saw Celtic, Hebrew, Anglo-Saxon, or Mediterranean physiognomies where today we see only subtly varying shades of a mostly undifferentiated whiteness."

¹⁵A related, though multidimensional, framework emphasizing context dependence with similar predictions on reclassification is the literature on stereotypes (Bordalo et al., 2016), which relies on the representativeness heuristic (Kahneman and Tversky, 1972). Stereotypes about groups depend on the reference group. Irish immigrants may be thought of primarily as Catholic when compared to native Anglo-Saxons. However, when compared to African Americans, they are more likely to be perceived as white, since skin color is the dimension in which Irish and African Americans differ the most.

Other than through higher average out-group distance, the disutility that white natives endure when interacting with outsiders can be further reduced through the efforts that out-group members themselves exert in order to assimilate. This assimilation effort, denoted by e , encompasses a range of behaviors that reduce the out-group's perceived distance to the in-group, such as learning their language, adopting their habits and mode of dress, or their characteristic naming patterns.

Discrimination. With each interaction, native whites are faced with a binary decision of whether to treat an individual as a member of their group or as an outsider, that is, whether to engage in discriminatory behavior. Discrimination relieves native whites of the psychological cost of interacting with distant others, but it comes at a cost F , which captures both the actual effort of engaging in discriminatory behavior and the cost of foregone monetary or social transactions with members of the out-group. We denote the decision to discriminate or not by $d \in [0, 1]$, so that utility for an in-group member is given by

$$U = \begin{cases} -W(\lambda, e), & \text{if } d = 0 \\ -F, & \text{if } d = 1 \end{cases}$$

where W is the psychological cost of interacting with an individual at perceived distance λ and who provides effort e , with $W_\lambda > 0$, $W_e < 0$ and $W_{\lambda,e} < 0$. $W(0, e) = 0$ for any e . Native whites engage in discrimination whenever its benefits exceed its costs, so whenever $W > F$.

Assimilation decisions of out-group members. Out-group members provide assimilation effort in the hope of becoming part of the native white in-group. Membership in the native group confers benefits, both material (e.g. access to better jobs or housing) and psychological (e.g. avoiding the cost of discrimination or harassment). The problem of out-group members is given by

$$\max_e P(e, h)B - c(e)$$

where B is the benefit from assimilation and c is a convex cost function with $c(0) = 0$. $P(e, h)$ is an indicator for successful assimilation and thus $P = \mathbb{1}(F > W)$. Consequently, $P_e \geq 0$, $P_h \leq 0$, and $P_{eh} \geq 0$.

The minimum amount of effort needed to ensure assimilation into the native white group is implicitly defined by

$$F = W(\lambda, \underline{e}(h)) \tag{4}$$

Out-group members can decide between exerting effort \underline{e} and enjoying the benefits of assimilation at cost $c(\underline{e})$, or avoiding the costs of effort and foregoing the assimilation benefit. An out-group member of type h will thus provide assimilation effort $e = \underline{e}(h)$ whenever $B > c(\underline{e}(h))$ and choose $e = 0$ otherwise. To ensure an interior solution we will always assume $c(\underline{e}(H^{max})) > B$.

Proposition 1. *There exists a threshold h^* , given by $B = (e(h^*))$, such that individuals with a distance below h^* optimally choose to provide sufficient assimilation effort to ensure integration in the native group. Those with a distance exceeding h^* provide zero effort and do not become assimilated. Formally,*

$$e^* = \begin{cases} e(h), & \text{for } h \leq h^* \\ 0, & \text{for } h > h^* \end{cases} \quad (5)$$

The appearance of a more distant out-group

In this context, we can think of the increasing presence of African Americans in the North as an increase in H^{max} , the perceived distance of the most distant out-group member. While no studies measure the precise social distance hierarchy in the early 20th century United States, studies in later years and in a number of countries suggest a clear ranking, with Northern Europeans at the top and Africans at the bottom. This ranking is furthermore agreed upon by members of all ethnic and racial groups (Berry and Kalin, 1979; Duckitt, 1992; Pettigrew, 1960). Our assumption builds on this empirical literature. This gives us the following proposition:

Proposition 2. *An increase in H^{max} leads to an increase in h^* .*

The increase in distance of the most distant out-group member has implications for the optimal level of effort exerted by out-group members of different types h . In particular, there is an intermediate range of types who would not have optimally exerted assimilation effort under the old distance cutoff, but who now find it profitable to do so.¹⁶

The empirical implication of Proposition 2 is that immigrant assimilation should increase on average in response to the Great Migration. This is our main empirical result presented in Section 4. This prediction directly follows from the assumption that the psychological cost of natives is decreasing in the average distance of the out-group. More interestingly, the model generates testable implications on heterogeneous patterns of assimilation. Specifically, we can state the effects of an increase in H^{max} on the optimal assimilation effort exerted by out-group members as a function of their type in the following proposition.

Proposition 3. *Consider an increase in H^{max} from H_1^{max} to H_2^{max} and the corresponding increase in the threshold distance level for assimilation from h_1^* to h_2^* . Denote with $e_1^*(h)$ and $e_2^*(h)$ the optimal level of effort provided by an immigrant of distance h , before and after the increase in H^{max} , respectively. We then have*

¹⁶The *accentuation* component of self categorization theory implies that people classify objects into groups in order to minimize within group differences and maximize across group differences (Turner et al., 1994; Haslam et al., 1995). This classification rule directly generates re-classification of white Europeans as members of the in-group when African Americans arrive and increase the variance of the out-group. Fryer and Jackson (2008) show that this rule of classification can derive from a utility maximization problem. Gennaioli and Tabellini (2018) rely on a similar assumption to study political identities.

$$\begin{cases} e_2^*(h) \leq e_1^*(h), & \text{if } h \leq h_1^* \\ e_2^*(h) > e_1^*(h), & \text{if } h_1^* < h \leq h_2^* \\ e_2^*(h) = e_1^*(h) = 0, & \text{if } h > h_2^* \end{cases}$$

Proposition 3 implies that the inflow of African Americans has a different effect on each of three distinct groups of immigrants, characterized by their baseline distance from the natives. This result is illustrated by Figure 5. The red line shows optimal effort levels for different values of h before an increase in H^{max} . Individuals who are sufficiently close to natives ($h \leq h_1^*$) exert the necessary effort to achieve assimilation, which is increasing in h . After a threshold where the costs of necessary effort equal the benefits of assimilation, immigrants “give up” and efforts (and consequently assimilation outcomes) drop to zero. Outcomes after an increase in H^{max} are shown by the blue line. A first group, with $h \leq h_1^*$, was already considered part of the in-group before the arrival of African Americans. They remain assimilated, but due to lower requirements from the side of the natives, they are allowed to somewhat decrease their assimilation efforts. A second group with intermediate levels of h used to be unable to achieve membership to the in-group, but now benefits from the reduction in necessary effort. This group substantially increases effort, and becomes assimilated. The final group, with largest distance $h > h_2^*$, will remain unassimilated. Notice that if the increase in H^{max} is sufficiently large, none of the existing immigrants will fall into the last group, i.e. if $H_1^{max} < h_2^*$, then all existing immigrants will provide sufficient effort and will become assimilated. Finally, note that changes in H^{max} will induce changes in effort along both the intensive (i.e. some groups who were not assimilating start to exert effort and assimilate) and the extensive (i.e. some groups who were already assimilating decrease their effort while remaining assimilated) margin.

In Appendix Section E.2, we present a generalization of the model where we relax the assumption that assimilation effort deterministically leads to assimilation success. Instead, we assume that, after deciding how much effort to provide, each out-group member receives an idiosyncratic assimilation shock. Individuals then optimize over the probability of a positive assimilation outcome, conditional on their distance, the distance of the average out-group member, and their personal effort provision. As a result, for each value of h , only a share of out-group members will be assimilated, despite the fact that all out-group members of the same h provide the same effort level.

The main insight gained from the deterministic model is retained in the case of stochastic assimilation: the response of effort to a change in \bar{h} , follows an inverted U-shape with respect to distance h . The resulting change in assimilation also has a U-shaped pattern, but interestingly, the effect on assimilation rates peaks at a significantly lower value of h compared to the effect on effort. Relaxed requirements by the natives allow for the immediate assimilation of the closest (still unassimilated) out-group members, who tend to be individuals with low values of h . On the other hand, incentives to exert more effort reach out further, including individuals with a large distance from natives. It is worth noting that alternative ways of thinking about the effect of black inflows – for example as an increase in direct assimilation benefits B that

would take place if black competition increases the incentives to Americanization effort for immigrants – do not produce a similar distinction in responses between effort provision and assimilation rates.

In sum, Propositions 3 and 4 translate into the following two testable predictions:

- (P1) There is an inverted U-shaped relationship between immigrants’ assimilation effort (or successful assimilation) and their distance from the native white in-group.
- (P2) The effect of black inflows on assimilation effort is largest for immigrant groups that are relatively distant to the natives. The effect of black inflows on successful assimilation is instead largest for groups of relatively low distance.

Our model implies that, while the Great Migration lowered barriers to assimilation for European immigrants, not all groups of immigrants profited equally. Groups relatively close to native whites in terms of skin color or cultural distance, but who were still considered outsiders before the arrival of blacks, should have experienced a large increase in assimilation rates, despite that fact that their increase in effort provision would have been relatively small when compared to more distant groups. On the other hand, groups sufficiently far away would not have benefited enough from reduced prejudice to become part of the in-group in significant numbers. In the following section we test and find support for these predictions in the data.

5.2 Empirical evidence on mechanisms

This section provides evidence for our proposed mechanism. In section 5.2.1 we provide evidence supportive of our main assumption: that black inflows reduced the perceived distance between immigrants and natives, and lowered prejudice towards immigrants. In section 5.2.2 we test the model’s main empirical implications.

5.2.1 Anti-immigrant sentiment and immigrant stereotypes in the press

The first piece of evidence for a mechanism emphasizing lower relative distance and reduced prejudice towards European immigrants comes from the historical press. Newspaper language and sentiment largely responds to readers’ demands (Gentzkow and Shapiro, 2010) and thus anti-immigrant sentiment in local newspapers should capture the public’s attitudes toward immigrants in each MSA. We compile a list of articles from Newspapers.com, covering cities in 69 of the 108 MSAs in our sample. In particular, we compute the MSA-level frequency of immigration-related terms appearing in local newspapers and investigate how that responds to the Great Migration. Results are presented in Table 4. The dependent variable is constructed as the number of articles containing the expression at the top of each column, scaled by the total number of articles containing the word “and” in newspapers of a given MSA in a given decade. To ease interpretation, the dependent variable is standardized by subtracting the mean and dividing by the standard deviation in 1910.

The inflow of African Americans had a negative and marginally significant effect on generic terms related to immigration (“immigration”; “immigrants”; “aliens”), suggesting that black arrivals lowered the salience of the immigration issue (column 1). Next, to more directly test our proposed mechanism, we focus on terms that either reflect concerns over immigration or capture cultural or ethnic prejudice. In column 2, we show that the Great Migration had negative and statistically significant effect on the relative frequency of the term “quotas”. Consistent with our previous findings, this result indicates that demand for immigration restrictions fell as European immigrants became increasingly perceived as less distant to native whites. In line with this idea, black inflows reduced the relative frequency of the word “Dago” (column 3), which was often used when describing Italian immigrants in disparaging terms. Even more direct evidence in support of our proposed mechanism is presented in columns 4 to 6, where we document that mentions to both “Catholic” or “Catholic threat” (columns 4 and 5) and the Ku Klux Klan (column 6) fell in response to black in-migration.

At the beginning of the 20th century, and especially during the 1920s, religion was a highly salient issue, and nativism was often associated with anti-Catholicism (see [Higham \(1998\)](#) and [D’Amico and Tabellini \(2017\)](#), among others). Indeed, the revival of the KKK that took place during the 1920s did not have an anti-black but, rather an anti-Catholic focus ([Dumenil, 1991](#)). Hence, we interpret findings in columns 4, 5, and 6 as consistent with the idea that the Great Migration softened the widespread anti-immigration sentiments (in particular towards Catholic immigrants) prevailing at the time.¹⁷

Alongside prejudice towards immigrants, the inflow of African Americans also reduced stereotyping. We provide evidence of this by searching for the co-occurrence of words commonly or stereotypically associated with two of the largest European immigrant groups of the time, the Irish and the Italians. Figure [6](#) and Table [A.6](#) present 2SLS estimates from regressions where the dependent variable is the relative frequency of such co-occurrences. In the left panel of Figure [6](#) we normalize each frequency with the frequency of the word “and”, which is proxying for the total number of articles published in an MSA in a decade. Black inflows reduced the probability that the Irish and Italians were mentioned jointly with the term “Catholic”. This lends support to a theoretical mechanism whereby immigrants are reclassified as in-group members, and are no longer associated with features that previously distinguished them from the natives, such as religion. This pattern is also consistent with context-dependent stereotyping, as in [Bordalo et al. \(2016\)](#). Similarly, the relative frequency of other stereotypical associations – such as that of the Irish with violence, or that of the Italians with the mafia – also declined following black in-migration.

The right panel of Figure [6](#) normalizes the frequency of each co-occurrence with the frequency of each stereotypical term. This is a complementary statistical test, since it requires

¹⁷The decline in KKK mentions reported in column 6 likely represents a lower bound for the effects of black inflows on the reduction in native whites’ prejudice towards European immigrants, since the arrival of African Americans likely increased KKK activity.

that the frequency of nationality-stereotypical term associations represents a smaller fraction of the total mentions of a given stereotypical term, and thus controls for the fact that terms like “crime” or “alcohol” may have become less widely discussed topics over time. The qualitative message is unchanged: also in this case, the Irish were significantly less likely to be associated with Catholicism, alcohol and crime, and Italians were significantly less likely to be associated with the mafia. Overall, these results indicate that the Great Migration reduced natives’ propensity to link immigrants to negative stereotypes. Moreover, they indicate that cultural traits like religion lost importance, in turn suggesting that features like skin color became more salient.

5.2.2 Heterogeneity across immigrant groups

The main predictions of our theoretical framework in Section 5.1 concern heterogeneity across immigrant groups. We start by testing prediction P1, which refers to the non-monotonic effect of the Great Migration on immigrant assimilation efforts. We proxy for the distance between immigrants and natives in two ways. First, we use a measure of genetic distance from the UK, compiled by Spolaore and Wacziarg (2009). Second, we rely on the measure of linguistic distance from English constructed by Chiswick and Miller (2005). The latter measure assigns lower values to languages in which English speakers had reached a higher degree of proficiency after several weeks of instruction. In Table 6, we examine if the effect of black inflows varies non-linearly with these two measures of distance at the nationality level. Figure 7 plots the implied effect of black inflows on intermarriage (left) and naturalization (right), by values of genetic (upper panel) and linguistic distance (lower panel) respectively.

According to the model, the response of assimilation effort should follow an inverted U-shape. Consistent with this prediction, we observe this pattern for naturalization rates. The effect is less pronounced, but still present for intermarriage in the case of genetic distance. The relationship with linguistic distance is flat in the case of this outcome. However, this is entirely driven by the Irish, who are assigned a zero linguistic distance from English, but actually experienced a drop in intermarriage rates – and, as Catholics, were religiously and culturally quite distant from the native Anglo-Saxons, and discriminated throughout the 19th and early 20th centuries. These results are encouraging for our proposed mechanism, but should be interpreted with caution since the presence of multiple interactions affects the first-stage F-statistic and increases concerns related to the presence of weak instruments.

The second empirical implication of the model (P2) suggests that, when assimilation effort does not deterministically translate into actual assimilation, the increase in assimilation effort peaks for distant immigrant groups, but the increase in successful assimilation peaks for groups that are closest to natives. We plot heterogeneous responses across immigrant groups in Figure 8. We report 2SLS coefficients for the effects of the Great Migration on intermarriage (left panel) and the probability of naturalization (right panel). Regression results underlying Figure 8 are reported in Table A.7. A clear pattern emerges. With the exception of the UK, New source regions exhibit the highest increase in naturalization rates. On the contrary, Germany, Northern and Western Europe experienced smaller or negative

changes. This pattern is consistent with Americanization efforts increasing the most for groups that were initially most discriminated against, and tried to exploit the “window of opportunity” offered by the Great Migration in order to fit in.

At the same time, results for intermarriage rates follow the opposite pattern. While effects are positive for most groups, with the notable exception of the Irish, acceptance by the native group increased the most for Old source regions, whose members had the smallest distance from native Anglo-Saxon whites. These patterns empirically verify prediction P2 and suggest that, despite increased efforts, native attitudes still constituted an important barrier to actual integration of distant immigrant groups.

Finally, in Table A.8 in the Appendix, we present results for our main outcomes for non-white immigrants (Mexicans and Chinese) who were viewed as even farther away in distance from natives. We confirm that for these very distant groups there is indeed no effect of black inflows on either assimilation effort or actual assimilation.

5.3 Alternative mechanisms

Competition and signaling. Blacks who moved to the North during the period were disproportionately low-skilled and were absorbed in sectors that were until then immigrant-dominated. The ensuing competition between low-skilled immigrants and blacks has been highly emphasized in the historical literature. Violent conflict between ethnic minorities and African Americans was common (Rieder, 1987; McDevitt, Levin and Bennett, 2002; Cho, 1993). Indeed, such struggles predated the Great Migration. Already before the Civil War, Irish immigrants reacted to their deplorable living conditions in northern cities with resentment against blacks, which was demonstrated in practice through their participation in anti-abolitionist riots and mobbing of African Americans.

Competition with blacks might have induced some immigrant groups to either invest in skill acquisition or actively try to differentiate themselves from their competitors, perhaps by signaling their Americanization as an asset in order to become more attractive to employers. Skill acquisition leading to occupational upgrading seems unlikely, since it would have implied, counterintuitively, that immigrants could have invested in their human capital and advanced occupationally even prior to the Great Migration, but they chose not to do so. Signaling American identity as a means to deal with competition is instead a channel that has been highlighted by the historical literature (Olzak and Shanahan, 2014). Ignatiev (1995), in his book *How the Irish became white*, documents how the Irish before the civil war facilitated their assimilation by emphasizing their differences from African Americans. Roediger (1991) shows how “immigrants in dirty and disease-ridden cities countered nativist assertions of racial difference with a determined focus on their own whiteness, on “the Negro”, and on slavery” (Guterl, 2001). Gunnar Myrdal, in his 1944 classical study of race relations in the US, suggests that prejudice against blacks was itself another means of signaling Americanization to the natives. “[T]he development of prejudice against Negroes [was] usually one of [the] first lessons in Americanization for [new immigrants residing in the North]. Because they are of low status, they like to have a group like the Negroes to which they can be superior” (Myrdal,

1944).

Heterogeneous effects on naturalization rates, displayed in the left panel of Figure 8, are consistent with increased efforts for groups most likely to experience labor market or other forms of competition from incoming blacks, and who might have used naturalization as a means to signal an American identity. However, if competition were the main channel at work, then we should also observe the largest increases in assimilation rates among those groups that increased their efforts the most. But this is not the case. Social assimilation, as proxied by intermarriage rates (right panel of Figure 8), displays the highest increase for Old source country nationals. These patterns can instead be reconciled using our proposed mechanism. The fact that groups that only marginally increased their assimilation effort experienced the largest gains in assimilation suggests that native attitudes were a crucial mediator of the effects of the Great Migration.

A more direct test of the role of labor market competition is consistent with this interpretation. Table 5 presents interactions of black inflows with the share of an immigrant group at the MSA-level employed in manufacture (columns 1 and 2) or in unskilled occupations (columns 3 and 4) in 1900, prior to the Great Migration. The positive effect of black inflows on naturalization rates seems to stem entirely from groups of immigrants employed in these sectors and that were disproportionately exposed to black competition. This confirms the origin region-level results of Table A.7, and suggests that labor market exposure to black migrants induced immigrants to increase their assimilation efforts. The effect of employment sector on intermarriage rates on the other hand is muted, suggesting that competition may have driven assimilation efforts, but not actual assimilation. This again strongly suggests that, while competition may have driven immigrant responses, it could not have led to their assimilation were it not for a change in the attitudes of natives.

Labor market complementarities. A second alternative mechanism is also economic in nature. If immigrants exhibited some degree of complementarity with African Americans, black arrivals may have not constituted direct competition, but instead may have led to immigrants' occupational upgrading (Peri and Sparber, 2009; Fogel and Peri, 2016). Such economic advancement could then have fostered Europeans' social incorporation. Figure A.3 shows that the skills of African Americans were very similar to those of Eastern and Southern Europeans, but quite different from those of more skilled natives and immigrants from Old source countries. It is possible that the heterogeneity patterns observed for social assimilation, with Northern and Western Europeans exhibiting the highest increases in intermarriage, are a result of blacks (positively) affecting immigrants' economic status.

If economic complementarities were the main drivers of our results, we would also expect similar patterns of heterogeneity for economic outcomes. This is not what we find. Figure A.4 in the Appendix presents the effects of black inflows, separately for each immigrant group, on two economic outcomes that we use as proxies for the economic distance between immigrants and natives. First, we consider the employment share of immigrants in manufacturing – a sector where the majority of immigrants used to be employed, but where much fewer natives were working. This was also a sector that absorbed many African Americans during

the Great Migration (e.g. [Boustan, 2016](#)). Second, we construct the native-immigrant gap in the log occupational scores.¹⁸ Unlike social assimilation, economic assimilation displays little heterogeneity, and does not indicate that Old source immigrants were favored by the Great Migration relative to New ones. First, the reduction in the share of immigrants employed in manufacturing is rather uniform across groups. Second, there is no clear trend in the effects of black inflows on the native-immigrant gap in occupational scores across ethnic groups. In fact, if anything, the gap becomes larger for Germans, contrary to what the economic complementarity mechanism would predict. Taken together, the results in [Figure A.4](#) suggest that the patterns observed for intermarriage are unlikely to be mediated by differential economic advancement for the English, Western and Northern Europeans.

Political competition and mobilization. The last alternative mechanism behind the effects of the Great Migration on immigrant assimilation is political in nature. Specifically, it is possible that black arrivals altered politicians’ incentives, inducing them to shift their rhetoric from its previous anti-immigrant nativist focus to concerns related to the changing racial composition and its concomitant dangers, such as crime. For instance, [Erie \(1988\)](#) discusses how Irish immigrants achieved occupational and economic advancement through patronage jobs in law enforcement, a sector that grew as a result of the Great Migration and the disproportional rise in the rates of African American incarceration ([Muller, 2012](#)). We provide evidence against this mechanism by showing that black inflows did not increase immigrants’ employment in the public sector ([Table A.9](#)) – either overall (column 1) or in specific occupations, such as firemen or policemen (columns 2 and 3). These patterns suggest that our main results are not driven by immigrants being able to advance economically and socially by directly benefitting from patronage jobs.

Black arrivals might have also increased immigrants’ incentives to mobilize and push for political inclusion. The labor movement and unionization represented one of the most common forms of social mobilization at the time. Since blacks likely increased labor market competition mostly for Eastern and Southern Europeans, these immigrant groups might have been induced to join local labor unions and eventually assimilate at a faster pace. However, as with the signaling scenario, without assuming a change in the attitudes of natives, this mechanism can explain heterogeneous patterns for assimilation effort, but not for actual assimilation. In [Table A.10](#), we directly address this concern, and show that black inflows did not have any significant effect on the presence of local chapters of the Industrial Workers of the World (IWW), one of the largest industrial labor unions active in the US in the first decades of the 20th century. The IWW was founded in 1905 and data on IWW locals, collected as part of the IWW history project at the University of Washington ([Gregory, n.d.](#)), are available between 1906 and 1917. For this reason results in [Table A.10](#) include only 1910 and 1920. To further

¹⁸Occupational scores assign to an individual the median income of his job category in 1950, and can be used as a proxy for lifetime earnings ([Abramitzky, Boustan and Eriksson, 2014](#)).

corroborate the validity of our empirical strategy, we also check that the local presence of IWW unions in 1910 is uncorrelated with the 1910 to 1920 change in predicted black population predicted (Figure [A.5](#)). This result rules out the possibility that early settlements of blacks (and thus our instrument) were correlated with pre-determined labor and social mobilization, which might have in turn favored the assimilation of European immigrants.

6 Conclusion

The Great Migration of African Americans and the mass migration of Europeans are two processes that critically contributed to the formation of the modern American racial and ethnic landscape. Their interaction allows us to study the effects of race on ethnic identity formation in the US context, as well as draw more general conclusions on how the appearance of new out-groups drives the process of assimilation and broadens the boundaries of the in-group.

In this paper, we leverage exogenous variation in black inflows by exploiting 1900 black settlements in the North, and the persistent nature of migrant location decisions to show that the massive arrivals of African Americans to the US North during the first Great Migration facilitated the assimilation of European immigrants living in non-southern urban centers. We observe assimilation along dimensions that reflect increased immigrant efforts, such as naturalization decisions and naming patterns, but also equilibrium assimilation, such as intermarriage rates and occupational upgrading.

To explain these effects, we build on insights from social psychology and formalize the hypothesis – also proposed by the historical literature – that black inflows changed perceptions of native whites toward Europeans, making this group seem “closer” to them than before. Using measures of anti-immigrant sentiment from local historical newspapers, we provide direct evidence that black in-migration reduced national stereotyping and lowered concerns about immigration among natives. Assimilation effort increased the most among groups who faced the highest competition with African Americans, but assimilation itself rather followed a skin color gradient that benefited the “whitest” immigrant groups – primarily old source country nationals. Groups sufficiently distant from natives, that could not signal whiteness or did not expect that their efforts would result in acceptance, did not exhibit a pronounced response to the Great Migration.

While one needs to be cautious when extrapolating these findings to other historical periods, we believe this study can provide answers to questions relevant today. To what extent has the distinct racial profile of the US contributed to its multiculturalism and relative success in integrating immigrants? Does the racial or cultural distance of new immigrant arrivals matter for the assimilation of existing immigrant stock? The framework used here can be extended to examine more broadly the interactions of earlier and later immigrant arrivals and the role of new immigration on the assimilation of more established immigrant groups.

References

- Abramitzky, Ran, and Leah Boustan.** 2017. “Immigration in American Economic History.” *Journal of Economic Literature*, 55(4): 1311–45.
- Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson.** 2014. “A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration.” *Journal of Political Economy*, 122(3): 467–717.
- Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson.** 2018. “Cultural Assimilation during the Two Ages of Mass Migration.” Working Paper.
- Ager, Philipp, and Casper Worm Hansen.** 2017. “Closing Heaven’s Door: Evidence from the 1920s US Immigration Quota Acts.” Working Paper.
- Akerlof, George A., and Rachel E. Kranton.** 2000. “Economics and Identity.” *The Quarterly Journal of Economics*, 115(3): 715–753.
- Algan, Yann, Thierry Mayer, and Mathias Thoenig.** 2013. “The Economic Incentives of Cultural Transmission: Spatial Evidence from Naming Patterns across France.” CEPR Discussion Paper 9416.
- Angrist, Josh.** 2002. “How do sex ratios affect marriage and labor markets? Evidence from America’s second generation.” *The Quarterly Journal of Economics*, 117(3): 997–1038.
- Bandiera, Oriana, Imran Rasul, and Martina Viarengo.** 2013. “The Making of Modern America: Migratory Flows in the Age of Mass Migration.” *Journal of Development Economics*, 102: 23–47.
- Bandiera, Oriana, Myra Mohnen, Imran Rasul, and Martina Viarengo.** forthcoming. “Nation-Building Through Compulsory Schooling During the Age of Mass Migration.” *The Economic Journal*.
- Bauer, Michal, Christopher Blattman, Julie Chytilová, Joseph Henrich, Edward Miguel, and Tamar Mitts.** 2016. “Can war foster cooperation?” *Journal of Economic Perspectives*, 30(3): 249–74.
- Bernhard, Helen, Urs Fischbacher, and Ernst Fehr.** 2006. “Parochial Altruism in Humans.” *Nature*, 442(7105): 912.
- Berry, John W, and Rudolf Kalin.** 1979. “Reciprocity of Inter-Ethnic Attitudes in a Multicultural Society.” *International Journal of Intercultural Relations*, 3(1): 99–111.
- Biavaschi, Costanza, Corrado Giulietti, and Zahra Siddique.** 2017. “The Economic Payoff of Name Americanization.” *Journal of Labor Economics*, 35(4): 1089–1116.
- Bisin, Alberto, and Thierry Verdier.** 2001. “The Economics of Cultural Transmission and the Dynamics of Preferences.” *Journal of Economic Theory*, 97(2): 298–319.
- Black, Dan A, Seth G Sanders, Evan J Taylor, and Lowell J Taylor.** 2015. “The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South.” *The American Economic Review*, 105(2): 477–503.
- Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer.** 2016. “Stereotypes.” *The Quarterly Journal of Economics*, 131(4): 1753–1794.
- Borjas, George J.** 1987. “Self-Selection and the Earnings of Immigrants.” *The American Economic Review*, 77(4): 531–553.
- Borusyak, Kirill, Peter Hull, and Xavier Jaravel.** 2018. “Quasi-Experimental Shift-Share Research Designs.” NBER Working Paper w24997.
- Boustan, Leah Platt.** 2010. “Was Postwar Suburbanization “White Flight”? Evidence from the Black Migration.” *Quarterly Journal of Economics*, 125(1): 417–443.
- Boustan, Leah Platt.** 2016. *Competition in the Promised Land: Black Migrants in Northern Cities and Labor Markets*. Princeton, New Jersey: Princeton University Press.
- Boustan, Leah Platt, M. E. Kahn, and P. W. Rhode.** 2012. “Moving to Higher Ground: Migration Response to Natural Disasters in the Early Twentieth Century.” *American Eco-*

- conomic Review*, 102(3): 238–244.
- Card, David.** 2001. “Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration.” *Journal of Labor Economics*, 19(1): 22–64.
- Card, David, Alexandre Mas, and Jesse Rothstein.** 2008. “Tipping and the Dynamics of Segregation.” *Quarterly Journal of Economics*, 123(1): 177–218.
- Chen, Yan, and Sherry Xin Li.** 2009. “Group Identity and Social Preferences.” *American Economic Review*, 99(1): 431–57.
- Chiswick, Barry R.** 1978. “The Effect of Americanization on the Earnings of Foreign-Born Men.” *Journal of Political Economy*, 86(5): 897–921.
- Chiswick, Barry R, and Paul W Miller.** 2005. “Linguistic Distance: A Quantitative Measure of the Distance between English and other Languages.” *Journal of Multilingual and Multicultural Development*, 26(1): 1–11.
- Choi, Jung-Kyoo, and Samuel Bowles.** 2007. “The Coevolution of Parochial Altruism and War.” *Science*, 318(5850): 636–640.
- Cho, Sumi K.** 1993. “Korean Americans vs. African Americans: Conflict and Construction.” In *Reading Rodney King: Reading Urban Uprising.*, ed. Robert Gooding-Williams. New York:Routledge.
- Collins, William J.** 1997. “When the Tide Turned: Immigration and the Delay of the Great Black Migration.” *The Journal of Economic History*, 57(3): 607–632.
- Collins, William J., and Marianne H. Wanamaker.** 2014. “Selection and Income Gains in the Great Migration of African Americans.” *American Economic Journal: Applied Economics*, 6(1): 220–252.
- Collins, William J., and Marianne H. Wanamaker.** 2015. “The Great Migration in Black and White: New Evidence on the Selection and Sorting of Southern Migrants.” *The Journal of Economic History*, 75(4): 947–992.
- D’Amico, Leonardo, and Marco Tabellini.** 2017. “Measuring Attitudes Towards Immigration Using Local Newspapers’ Data and Congressional Speeches.” *Working Paper*.
- Duckitt, John H.** 1992. *The Social Psychology of Prejudice*. Praeger Publishers/Greenwood Publishing Group.
- Dumenil, Lynn.** 1991. “The Tribal Twenties: “Assimilated” Catholics’ Response to Anti-Catholicism in the 1920s.” *Journal of American Ethnic History*, 21–49.
- Echenique, Federico, and Roland G. Fryer, Jr.** 2007. “A Measure of Segregation Based on Social Interactions.” *Quarterly Journal of Economics*, 122(2): 441–485.
- Edin, Per-Anders, Peter Fredriksson, and Olof Åslund.** 2003. “Ethnic Enclaves and the Economic Success of Immigrants – Evidence from a Natural Experiment.” *The Quarterly Journal of Economics*, 118(1): 329–357.
- Enos, Ryan D.** 2017. *The Space between Us: Social Geography and Politics*. Cambridge University Press.
- Erie, Steven P.** 1988. *Rainbow’s End: Irish-Americans and the Dilemmas of Urban Machine Politics, 1840–1985*. Berkeley and Los Angeles:University of California Press.
- Eriksson, Katherine, and Zachary A. Ward.** 2018. “The Ethnic Segregation of Immigrants in the United States from 1850 to 1940.” NBER Working Paper w24764.
- Ferenczi, Imre, and Walter F. Willcox.** 1929. *International Migrations Volume I: Statistics*. New York:National Bureau of Economic Research.
- Ferrie, Joseph.** 1996. “A New Sample of Males Linked from the Public Use Micro Sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census Manuscript Schedules.” *Historical Methods*, (29): 141–156.
- Foged, Mette, and Giovanni Peri.** 2016. “Immigrants’ effect on native workers: New analysis on longitudinal data.” *American Economic Journal: Applied Economics*, 8(2): 1–34.

- Fouka, Vasiliki.** 2018. “Backlash: The Unintended Effects of Language Prohibition in US Schools after World War I.” Stanford University Working Paper.
- Fouka, Vasiliki.** forthcoming. “How do Immigrants Respond to Discrimination: The Case of Germans in the US during World War I.” *American Political Science Review*.
- Fowler, James H, and Nicholas A Christakis.** 2010. “Cooperative Behavior Cascades in Human Social Networks.” *Proceedings of the National Academy of Sciences*, 200913149.
- Fryer, Roland, and Matthew O Jackson.** 2008. “A Categorical Model of Cognition and Biased Decision Making.” *The BE Journal of Theoretical Economics*, 8(1).
- Fryer, Roland G., and Steven D. Levitt.** 2004. “The Causes and Consequences of Distinctively Black Names.” *The Quarterly Journal of Economics*, 119(3): 767–805.
- Gaertner, Samuel, and John Dovidio.** 2000. *Reducing Intergroup Bias: The Common Ingroup Identity Model*. Psychology Press.
- Gennaioli, Nicola, and Guido Tabellini.** 2018. “Identity, Beliefs, and Political Conflict.” CEPR Discussion Paper DP13390.
- Gentzkow, Matthew, and Jesse M. Shapiro.** 2010. “What Drives Media Slant? Evidence from U.S. Daily Newspapers.” *Econometrica*, 78(1): 35–71.
- Gentzkow, Matthew, and Jesse M. Shapiro.** 2011. “Ideological Segregation Online and Offline.” *Quarterly Journal of Economics*, 126(4): 1799–1839.
- Goette, Lorenz, David Huffman, and Stephan Meier.** 2006. “The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups.” *American Economic Review*, 96(2): 212–216.
- Goldin, Claudia.** 1994. “The Political Economy of Immigration Restriction in the United States, 1890 to 1921.” In *The Regulated Economy: A Historical Approach to Political Economy*. 223–258. University of Chicago Press.
- Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift.** 2018. “Bartik Instruments: What, When, Why, and How.” NBER Working Paper w24408.
- Gordon, Milton A.** 1964. *Assimilation in American Life: The Role of Race, Religion, and National Origins*. New York:Oxford University Press.
- Gregory, James N.** n.d.. “IWW History Project.”
- Gregory, J. N.** 2005. *The Southern Diaspora: How the Great Migrations of Black and White Southerners Transformed America*. Chapel Hill:University of North Carolina Press.
- Guglielmo, Thomas A.** 2003. *White on Arrival: Italians, Race, Color, and Power in Chicago, 1890-1945*. Oxford University Press.
- Guterl, Matthew Pratt.** 2001. *The Color of Race in America, 1900-1940*. Cambridge, MA:Harvard University Press.
- Hagendoorn, Louk.** 1995. “Intergroup Biases in Multiple Group Systems: The Perception of Ethnic Hierarchies.” *European review of social psychology*, 6(1): 199–228.
- Hargreaves Heap, Shaun P, and Daniel John Zizzo.** 2009. “The Value of Groups.” *American Economic Review*, 99(1): 295–323.
- Haslam, S Alexander, Penelope J Oakes, John C Turner, and Craig McGarty.** 1995. “Social Categorization and Group Homogeneity: Changes in the Perceived Applicability of Stereotype Content as a Function of Comparative Context and Trait Favourableness.” *British Journal of Social Psychology*, 34(2): 139–160.
- Hatton, Timothy J., and Jeffrey G. Williamson.** 1998. *The Age of Mass Migration: Causes and Economic Impact*. New York:Oxford University Press.
- Higham, John.** 1998. *Strangers in the Land: Patterns of American Nativism, 1860-1925*. New Brunswick:Rutgers University Press.
- Hjort, Jonas.** 2014. “Ethnic Divisions and Production in Firms.” *Quarterly Journal of Economics*, 129(4): 1899–1946.
- Hochschild, Jennifer L., and Brenna Marea Powell.** 2008. “Racial Reorganization and

- the United States Census 1850-1930: Mulattoes, Half-Breeds, Mixed Parentage, Hindoos, and the Mexican Race.” *Studies in American Political Development*, 22(1): 59–96.
- Hornbeck, Richard, and Suresh Naidu.** 2014. “When the Levee Breaks: Black Migration and Economic Development in the American South.” *American Economic Review*, 104(3): 963–990.
- Ignatiev, Noel.** 1995. *How the Irish Became White*. New York:Routledge.
- Jacobson, Matthew Frye.** 1999. *Whiteness of a Different Color*. Cambridge, MA:Harvard University Press.
- Jaeger, David A, Joakim Ruist, and Jan Stuhler.** 2018. “Shift-Share Instruments and the Impact of Immigration.” NBER Working Paper w24285.
- Jordan, Jillian J, David G Rand, Samuel Arbesman, James H Fowler, and Nicholas A Christakis.** 2013. “Contagion of Cooperation in Static and Fluid Social Networks.” *PloS One*, 8(6): e66199.
- Kahneman, Daniel, and Amos Tversky.** 1972. “Subjective Probability: A Judgment of Representativeness.” In *The Concept of Probability in Psychological Experiments*. 25–48. Springer.
- Katz, Larry, and Robert Margo.** 2013. “Technical Change and the Relative Demand for Skilled Labor: The United States in Historical Perspective.” NBER Working Paper 18752.
- Lange, F. A., L. Olmstead, and P. W. Rhode.** 2009. “The Impact of the Boll Weevil, 1892-1932.” *Journal of Economic History*, 69(3): 685–718.
- Lee, Everett S., Ann Ratner Miller, Carol P. Brainerd, and Richard A. Easterlin.** 1957. “Methodological Considerations and Reference Tables.” In *Population redistribution and economic growth: United States, 1870-1950.* , ed. Simon Kuznets and Dorothy Swaine. Philadelphia:American Philosophical Society.
- Lleras-Muney, Adriana, and Allison Shertzer.** 2015. “Did the Americanization Movement Succeed? An Evaluation of the Effect of English-Only and Compulsory Schooling Laws on Immigrants.” *American Economic Journal: Economic Policy*, 7(3): 258–290.
- Logan, Trevon D., and John M. Parman.** 2017. “The National Rise in Residential Segregation.” *Journal of Economic History*, 77(1): 127–170.
- Massey, Douglas S, and Nancy A Denton.** 1993. *American Apartheid: Segregation and the Making of the Underclass*. Harvard University Press.
- Mazumder, Soumyajit.** 2018. “Becoming White: How Mass Warfare Transformed Immigrants into Americans.” Harvard University Working Paper.
- McAdams, Richard H.** 1995. “Cooperation and Conflict: The Economics of Group Status Production and Race Discrimination.” *Harvard Law Review*, 108(5): 1003–1084.
- McConnell, Brendon, and Imran Rasul.** 2018. “Contagious Animosity in the Field: Evidence from the Federal Criminal Justice System.” University College London Working Paper.
- McDevitt, Jack, Jack Levin, and Susan Bennett.** 2002. “Hate Crime Offenders: An Expanded Typology.” *Journal of Social Issues*, 58(2): 303–317.
- McGarty, Craig.** 1999. *Categorization in Social Psychology*. London, Thousand Oaks, New Delhi:Sage Publications.
- Muller, Christopher.** 2012. “Northward Migration and the Rise of Racial Disparity in American Incarceration, 1880–1950.” *American Journal of Sociology*, 118(2): 281–326.
- Myrdal, Gunnar.** 1944. *An American Dilemma: The Negro Problem and Modern Democracy*. New York:Harper & Bros.
- Olzak, Susan.** 2013. “Competition Theory of Ethnic/Racial Conflict and Protest.” In *The Wiley-Blackwell Encyclopedia of Social and Political Movements..* , ed. David A. Snow, Bert Klendermans, Donatella della Porta and Doug McAdam. Oxford, UK:Blackwell Publishers.
- Olzak, Susan, and Suzanne Shanahan.** 2003. “Racial Policy and Racial Conflict in the

- Urban United States, 1869–1924.” *Social Forces*, 82(2): 481–517.
- Olzak, Susan, and Suzanne Shanahan.** 2014. “Prisoners and Paupers: The Impact of Group Threat on Incarceration in Nineteenth Century American Cities.” *American Sociological Review*, 79(3): 392–411.
- Peri, Giovanni, and Chad Sparber.** 2009. “Task Specialization, Immigration, and Wages.” *American Economic Journal: Applied Economics*, 1(3): 135–69.
- Pettigrew, Thomas F.** 1960. “Social Distance Attitudes of South African Students.” *Social Forces*, 246–253.
- Rieder, Jonathan.** 1987. *Canarsie. The Jews and Italians of Brooklyn against Liberalism.* Cambridge, MA:Harvard University Press.
- Roediger, David R.** 1991. *The Wages of Whiteness.* London, New York:Verso.
- Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek.** 2010. *Integrated Public Use Microdata Series: Version 5.0 [Machine-Readable Database].* University of Minnesota.
- Sambanis, Nicholas, and Moses Shayo.** 2013. “Social Identification and Ethnic Conflict.” *American Political Science Review*, 107(2): 294–325.
- Shayo, Moses.** 2009. “A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution.” *American Political Science Review*, 103(2): 147–174.
- Shertzer, Allison.** 2016. “Immigrant Group Size and Political Mobilization: Evidence from European Migration to the United States.” *Journal of Public Economics*, 139: 1–12.
- Shertzer, Allison, and Randall Walsh.** 2016. “Racial Sorting and the Emergence of Segregation in American Cities.” NBER Working Paper 22077.
- Spiro, Jonathan P.** 2008. *Defending the Master Race: Conservation, Eugenics, and the Legacy of Madison Grant.* Burlington:University of Vermont Press.
- Spolaore, Enrico, and Romain Wacziarg.** 2009. “The Diffusion of Development.” *The Quarterly Journal of Economics*, 124(2): 469–529.
- Suri, Siddharth, and Duncan J Watts.** 2011. “Cooperation and Contagion in Web-Based, Networked Public Goods Experiments.” *PloS One*, 6(3): e16836.
- Swift, Elaine K., Robert G. Brookshire, David T. Canon, Evelyn C. Fink, John R. Hibbing, Brian D. Humes, Michael J. Malbin, and Kenneth C. Martis.** 1989. *Database of [United States] Congressional Historical Statistics, 1789-1989.* Ann Arbor, MI:Inter-university Consortium for Political and Social Research [distributor].
- Tabellini, Marco.** 2018a. “Gifts of the Immigrants, Woes of the Natives: Lessons from the Age of Mass Migration.” Harvard Business School Working Paper 19-005.
- Tabellini, Marco.** 2018b. “Racial Heterogeneity and Local Government Finances: Evidence from the Great Migration.” Harvard Business School Working Paper 19-006.
- Tajfel, Henri.** 1974. “Social Identity and Intergroup Behavior.” *Social Science Information*, 13(2): 65–93.
- Tajfel, Henri.** 1979. “Individuals and Groups in Social Psychology.” *British Journal of Clinical Psychology*, 18(2): 183–190.
- Tajfel, Henri, Michael G Billig, Robert P Bundy, and Claude Flament.** 1971. “Social Categorization and Intergroup Behaviour.” *European Journal of Social Psychology*, 1(2): 149–178.
- Tolnay, S., and E. M. Beck.** 1990. “Black Flight: Lethal Violence and the Great Migration, 1900 to 1930.” *Social Science History*, 14(3): 347–370.
- Turner, John C, Michael A Hogg, Penelope J Oakes, Stephen D Reicher, and Margaret S Wetherell.** 1987. *Rediscovering the Social Group: A Self-Categorization Theory.* Basil Blackwell.
- Turner, John C, Penelope J Oakes, S Alexander Haslam, and Craig McGarty.** 1994. “Self and Collective: Cognition and Social Context.” *Personality and Social Psychology*

Bulletin, 20(5): 454–463.

Vigdor, Jacob L. 2010. *From Immigrants to Americans: The Rise and Fall of Fitting In*. Rowman & Littlefield Publishers.

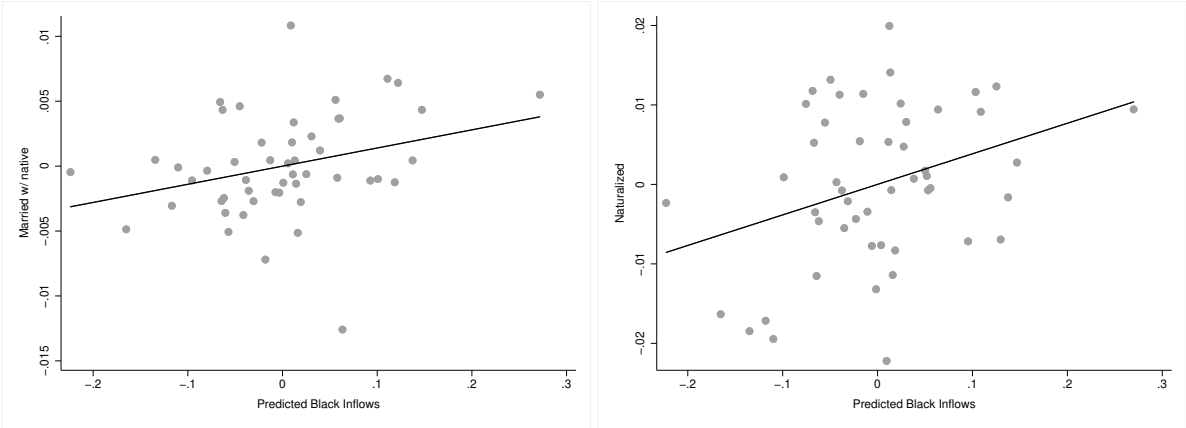
Watkins, Susan Cotts (ed.). 1994. *After Ellis Island: Newcomers and Natives in the 1910 census*. Russell Sage Foundation.

Wilkerson, I. 2010. *The Warmth of Other Suns: The Epic Story of Americas Great Migration*. Random House LLC.

Wright, Gavin. 1986. *Old South, New South*. NY:Basic Books.

Figures and Tables

Figure 1. Effects of black inflows on intermarriage and naturalization rates



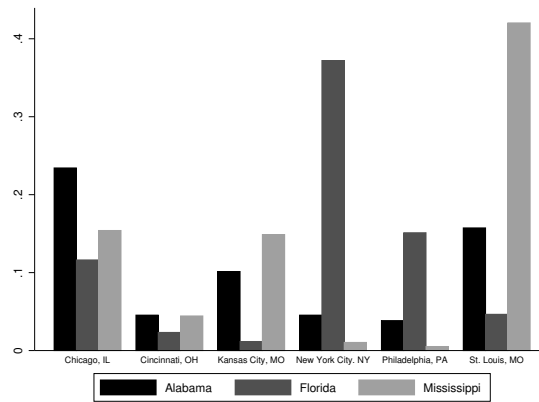
Binned scatterplot of the relationship between the probability of marrying a native-born person (left panel) or of being a naturalized citizen (right panel) and predicted black population for the years 1910-1930. Variables on the x- and y-axis represent residual changes, after partialling out total MSA population, MSA and region by year fixed effects and indicators for age, nationality group and years in the US. See Section 3.3 for details on the construction of the instrument for black population.

Figure 2. Immigrants and African Americans in sample MSAs



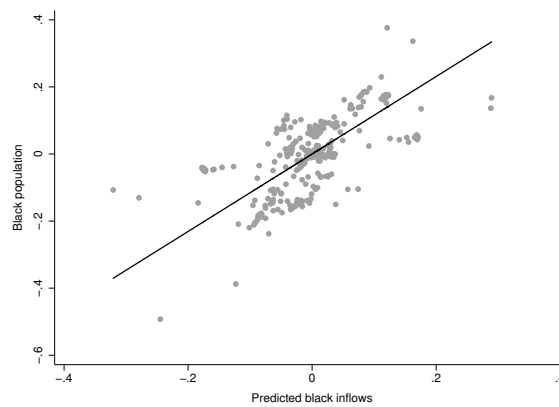
The map depicts the total number of foreign-born in 1910 (left panel) and the change in the number of African Americans between 1910 and 1930 (right panel) in the 108 MSAs in our sample.

Figure 3. Share of blacks from selected southern states in northern MSAs, 1900



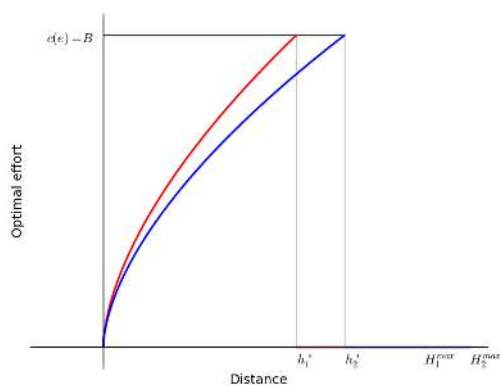
The figure shows the fraction of southern born blacks from a given state residing in the North living in one of the selected northern MSAs in 1900. Data are from the 5% 1900 IPUMS sample.

Figure 4. First stage



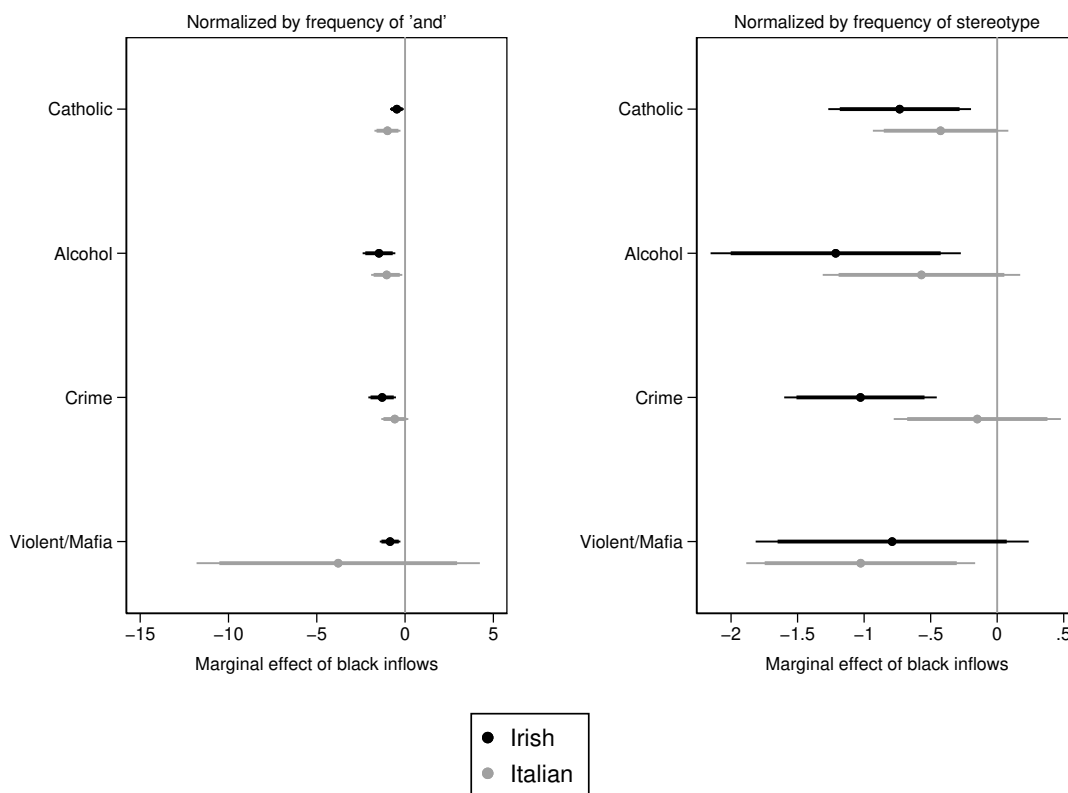
The figure shows the relationship between actual and predicted black population for the years 1910 to 1930. Each point represents the residual change in an MSA's actual and predicted number of blacks after partialling out total MSA population and MSA and region by year fixed effects.

Figure 5. The effects of an increase in H^{max}



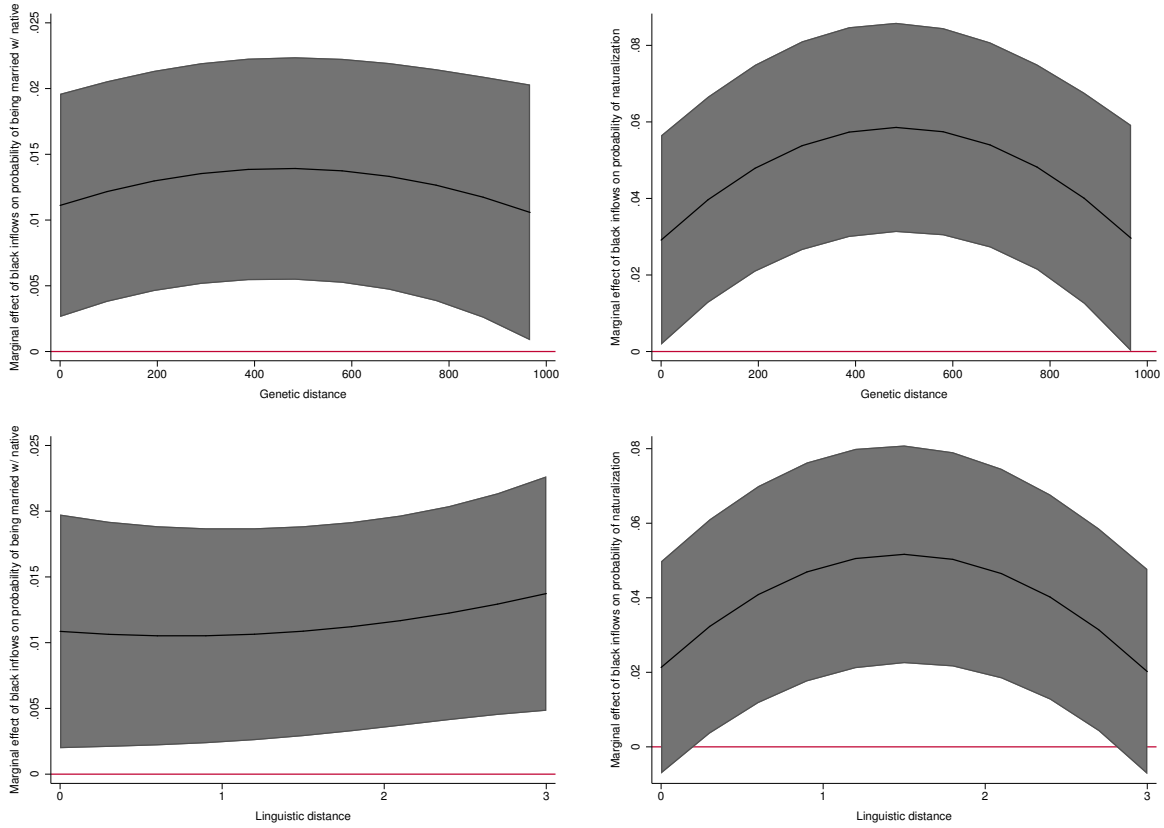
The graph assumes disutility of natives of the form $w(\cdot) = (1/\nu)h^{2/3}\bar{h}^{-1}e^{-1}$ and a quadratic cost function.

Figure 6. Effects on the frequency of national stereotypes



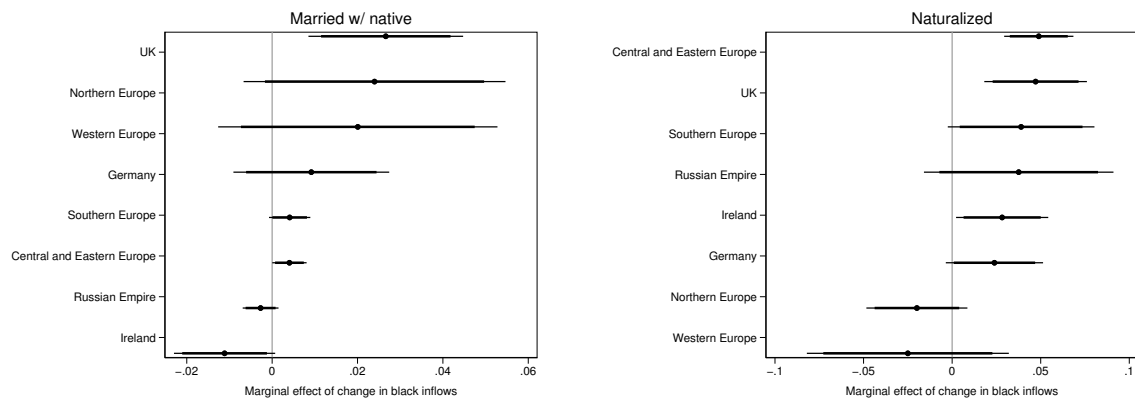
The figure plots estimates of the marginal effect of black inflows on the relative frequency of national stereotypes. Each line corresponds to a separate regression. The dependent variable is the frequency of the co-occurrence of each term on the y-axis with the word “Irish” (black lines) or “Italian” (gray lines), normalized by the frequency of the word “and” (left panel) or by the frequency of the respective term (right panel). Dependent variables are standardized by subtracting the mean and dividing by the standard deviation in 1910. Thin and thick lines indicate 95% and 90% confidence intervals, respectively.

Figure 7. Heterogeneity by distance from natives



The figure plots the marginal effect of black inflows on the probability of being married with a native (upper panel) and the probability of being a US citizen (lower panel) against values of genetic (left) and linguistic distance (right). The underlying regressions are reported in Table 6.

Figure 8. Heterogeneity by national origin



The figure plots 2SLS estimates of the marginal effect of black inflows on the probability of being married to a native (left panel) or being a naturalized citizen (right panel) from regressions separately run by national origin group that control for age, years in the US, MSA and year by region fixed effects. Thick and thin lines represent 90% and 95% confidence intervals, respectively.

Table 1. Summary statistics

| Variables | Mean | Median | S.D. | Min | Max | N |
|-------------------------------|---------|---------|---------|---------|------------|-----|
| <u>MSA characteristics</u> | | | | | | |
| MSA population | 420,640 | 159,581 | 992,887 | 23,606 | 10,900,000 | 324 |
| Number of immigrants | 94,648 | 21,478 | 308,618 | 781 | 3,338,862 | 324 |
| Number of blacks | 15,294 | 2,349 | 44,830 | 29 | 485,750 | 324 |
| Predicted number of blacks | 3,350 | 328 | 13,308 | -14,096 | 153,907 | 324 |
| Share foreign-born | 0.163 | 0.161 | 0.091 | 0.009 | 0.457 | 324 |
| Share blacks | 0.029 | 0.013 | 0.039 | 0.001 | 0.293 | 324 |
| <u>Main outcome variables</u> | | | | | | |
| Married with native | 0.150 | 0.132 | 0.085 | 0.036 | 0.498 | 324 |
| Share naturalized | 0.568 | 0.569 | 0.141 | 0.125 | 0.916 | 324 |

Notes: The sample consists of the 108 non-southern MSAs for which the instrument could be constructed, and is restricted to census years 1910, 1920, and 1930. Married with native refers to the share of immigrant men who are married with a native born spouse of native parentage. Share naturalized refers to the share of immigrant men who are US citizens.

Table 2. First stage

| Dep. variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Predicted black inflows | 1.064*** (0.313) | 1.155*** (0.268) | 0.899*** (0.284) | 1.125*** (0.249) | 1.101*** (0.279) | 1.090*** (0.260) | 1.152*** (0.268) |
| Observations | 324 | 324 | 324 | 324 | 324 | 324 | 324 |
| R-squared | 0.972 | 0.983 | 0.918 | 0.983 | 0.983 | 0.983 | 0.983 |
| F-stat | 11.38 | 17.46 | 9.444 | 19.08 | 14.54 | 15.31 | 17.27 |
| MSAs | 108 | 108 | 108 | 108 | 108 | 108 | 108 |
| Region \times Year FE | | X | X | X | X | X | X |
| Weighted | X | X | | X | X | X | X |
| Fr. black 1900 \times Year FE | | | | X | | | |
| Fr. foreign-born 1900 \times Year FE | | | | | X | | |
| Log Value manuf. output 1900 \times Year FE | | | | | | | |
| Fr. working age males in manuf. 1900 \times Year FE | | | | | | X | X |

Notes: The analysis is restricted to 108 non-southern MSAs for which the IV could be constructed, and to years 1910 to 1930. All regressions control for total population and MSA fixed effects. Column (1) includes year fixed effects. Standard errors clustered at the MSA level in parentheses. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table 3. Assimilation and assimilation effort

| | (1) | (2) | (3) | (4) | (5) | (6) |
|--|--------------------|-------------------|--------------------|---------------------|--------------------|---------------------|
| | OLS | OLS | 2SLS | 2SLS | 2SLS | 2SLS |
| Panel A: Married w/ native (<i>1910 mean: 0.071</i>) | | | | | | |
| Num. Blacks | 0.007 (0.006) | 0.005 (0.004) | 0.012** (0.005) | 0.013*** (0.005) | 0.012** (0.005) | 0.019 (0.017) |
| Observations | 9,323,126 | 9,323,126 | 9,323,126 | 9,323,109 | 9,323,109 | 88,892 |
| R-squared | 0.124 | 0.124 | 0.124 | 0.139 | 0.140 | 0.131 |
| F-stat | | | 23.33 | 23.71 | 23.83 | 32.75 |
| Panel B: Naturalized (<i>1910 mean: 0.491</i>) | | | | | | |
| Num. Blacks | 0.056** (0.025) | 0.030* (0.016) | 0.033** (0.015) | 0.038*** (0.013) | 0.033** (0.014) | 0.144*** (0.031) |
| Observations | 15,267,846 | 15,267,846 | 15,267,846 | 15,267,844 | 15,267,844 | 80,866 |
| R-squared | 0.345 | 0.346 | 0.346 | 0.352 | 0.355 | 0.569 |
| F-stat | | | 24.23 | 24.38 | 24.51 | 32.74 |
| Individual controls | X | X | X | X | X | |
| Region \times Year | | X | X | X | X | |
| MSA \times Origin region | | | | X | X | |
| Origin region \times Year | | | | | X | |
| Linked sample | | | | | | X |

Notes: The table presents results for immigrant men living in the 108 non-southern MSAs for which the instrument could be constructed in Census years 1910, 1920, and 1930. In Panel A the sample is restricted to married men. *Married w/ native* is a dummy equal to 1 if the individual is married to a native woman of native parentage. Cols 1 to 5 present results obtained from the repeated cross-sections (Section 3.2.1), while Col 6 shows results from the linked panel of men (Section 3.2.2) who always remained in the same MSA in the three Census years. Cols 1-2 (resp. 3-5) present OLS (resp. 2SLS results). Individual controls include fixed effects for age, years in the US and origin region. All regressions control for MSA and year fixed effects and for total MSA population. In Col 6 of Panel A the sample is restricted to men who were not married in the previous decade (Panel A) or who were not naturalized in the previous decade (Panel B). Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4. Frequency of anti-immigrant terms in press

| Dep. Variable | Relative frequency of word over frequency of word “and” | | | | | |
|---------------|---|---------------------|---------------------|----------------------|------------------------|---------------------|
| | (1) Immigration | (2) Quotas | (3) Dago | (4) Catholic | (5) Catholic Threat | (6) KKK |
| Num. Blacks | -1.005* (0.557) | -2.000** (0.819) | -1.382** (0.689) | -0.923*** (0.305) | -1.283*** (0.364) | -3.837** (1.878) |
| Observations | 195 | 195 | 195 | 195 | 195 | 195 |
| R-squared | 0.607 | 0.603 | 0.678 | 0.734 | 0.717 | 0.140 |
| F-stat | 17.67 | 17.67 | 17.67 | 17.67 | 17.67 | 17.67 |
| MSAs | 69 | 69 | 69 | 69 | 69 | 69 |

Notes: 2SLS estimates reported. The analysis is restricted to 65 non-southern MSAs for which the IV and the word frequency measure could be constructed, and to years 1910 to 1930. Dependent variables are standardized by subtracting the mean and dividing by the standard deviation in 1910. All regressions control for MSA and year by region fixed effects and MSA total population and are weighted by population in 1900. Standard errors clustered at the MSA level in parentheses. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

Table 5. Heterogeneity by sector of employment and skill in 1900

| Dep. variable | (1) Married w/ native | (2) Naturalized | (3) Married w/ native | (4) Naturalized |
|---|--------------------------|---------------------|--------------------------|---------------------|
| Num. blacks | 0.011* (0.006) | -0.005 (0.018) | 0.006 (0.007) | -0.005 (0.017) |
| Num. blacks \times Share in manufacture in 1900 | 0.004 (0.009) | 0.135*** (0.018) | | |
| Num. blacks \times Share unskilled in 1900 | | | 0.012 (0.008) | 0.080*** (0.019) |
| Observations | 9,322,784 | 15,266,394 | 9,322,784 | 15,266,394 |
| R-squared | 0.140 | 0.355 | 0.140 | 0.355 |
| F-stat | 12.32 | 12.57 | 11.95 | 12.29 |

Notes: 2SLS estimates reported. The analysis is restricted to 108 non-southern MSAs for which the IV could be constructed, and to years 1910 to 1930. The sample is restricted to foreign-born men. In Columns 1 and 3 it is further restricted to married individuals. All regressions control for MSA by origin region, origin region by MSA and year by region fixed effects and MSA total population. Standard errors clustered at the MSA level in parentheses. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table 6. Heterogeneity by distance from natives

| Dep. variable | (1) Married w/ native | (2) Naturalized | (3) Married w/ native | (4) Naturalized |
|---|--------------------------|----------------------|--------------------------|----------------------|
| Num. blacks | 0.011** (0.004) | 0.029** (0.014) | 0.011** (0.005) | 0.021 (0.015) |
| Num. blacks \times Genetic distance | 0.000** (0.000) | 0.000*** (0.000) | | |
| Num. blacks \times Genetic distance ² | -0.000** (0.000) | -0.000*** (0.000) | | |
| Num. blacks \times Linguistic distance | | | -0.001 (0.002) | 0.041*** (0.006) |
| Num. blacks \times Linguistic distance ² | | | 0.001 (0.001) | -0.014*** (0.002) |
| Observations | 9,220,664 | 14,978,414 | 9,219,280 | 15,012,424 |
| R-squared | 0.144 | 0.358 | 0.145 | 0.358 |
| F-stat | 8.145 | 8.371 | 7.969 | 8.211 |

Notes: 2SLS estimates reported. The analysis is restricted to 108 non-southern MSAs for which the IV could be constructed, and to years 1910 to 1930. The sample is restricted to foreign-born men. In Columns 1 and 3 it is further restricted to married individuals. All regressions include fixed effects for age, years in the US, MSA by nationality, nationality by year and year by region, and control for MSA total population. Standard errors clustered at the MSA level in parentheses. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

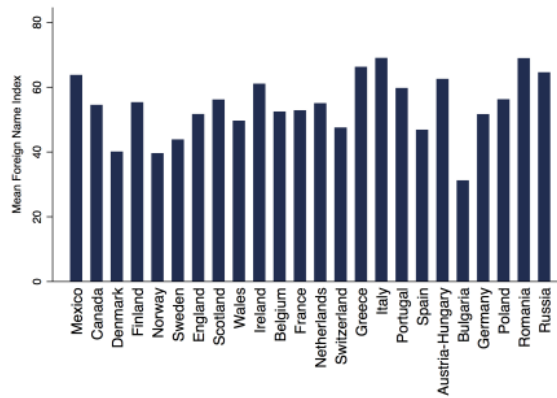
Appendix (Not for publication)

Table of Contents

| | |
|---|-----------|
| A Additional Figures and Tables | 51 |
| B Push instrument | 64 |
| C Robustness Checks | 66 |
| D Additional outcomes | 78 |
| D.1 Additional Economic and Social Outcomes | 78 |
| D.2 Residential Segregation | 80 |
| E Additional model results | 82 |
| E.1 Proofs | 82 |
| E.2 Stochastic extension | 82 |

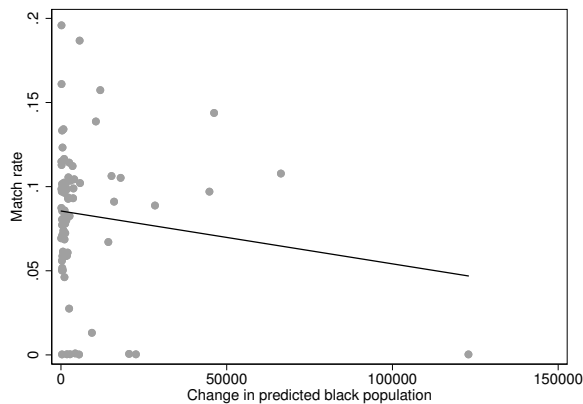
A Additional Figures and Tables

Figure A.1. Foreign name index by nationality in 1910



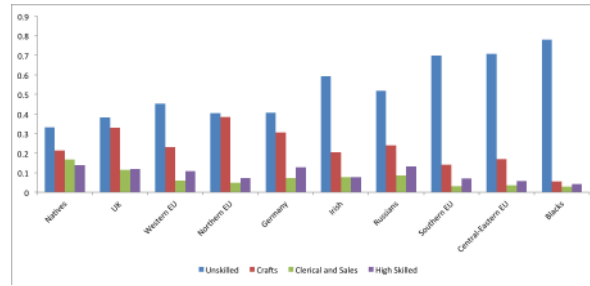
The figure depicts the mean Foreign name index among US-born men of foreign-born fathers from the origin counties listed on the x-axis.

Figure A.2. Correlation between match rate and change in predicted black inflows



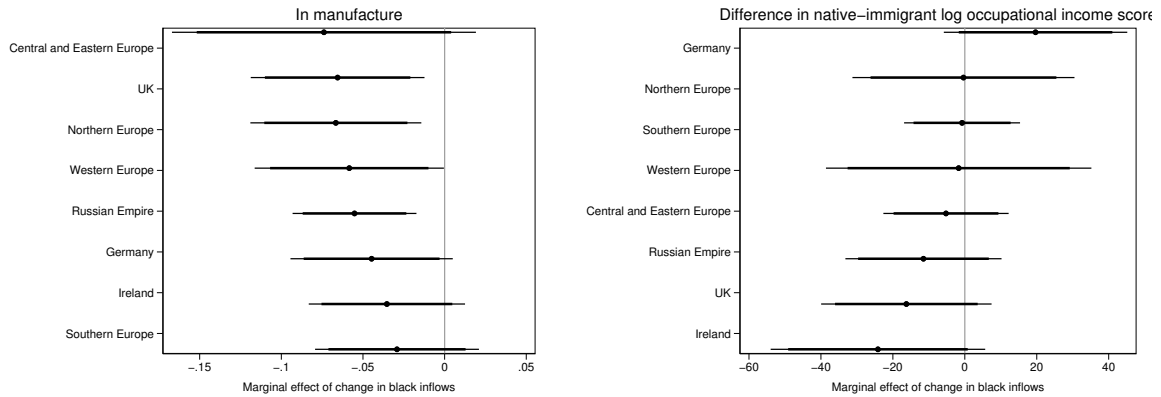
The figure plots the match rate in the dataset of non-movers linked across three census decades against the change in the predicted number of black arrivals in each period.

Figure A.3. Skill distribution for men aged 15-65 in 1910



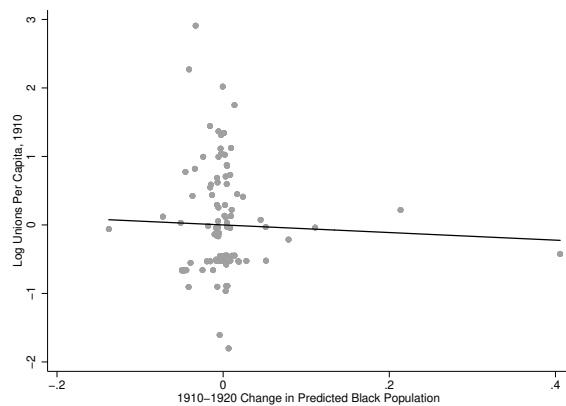
The figure plots the share of men aged 15-65 in each group who were employed in each sector. Skill and occupational categories were defined following the classification in [Katz and Margo \(2013\)](#). Authors' calculations from the 1910 full count Census of Population made available by IPUMS.

Figure A.4. Heterogeneity by national origin - Economic outcomes



The figure plots 2SLS estimates of the marginal effect of black inflows on the probability of being employed in manufacture (left panel) or on the difference in log occupational income score between natives and immigrants (right panel) from regressions separately run by national origin group that control for age, years in the US, MSA and year by region fixed effects. The sample is restricted to men aged 15 to 65. Thick and thin lines represent 90% and 95% confidence intervals, respectively.

Figure A.5. Correlation between local labor unions in 1910 and change in predicted black inflows



The figure represents the residual scatterplot for a regression of the logarithm of local IWW chapters in 1910 (y-axis) against the 1910 to 1930 change in predicted black in-migration (x-axis), after partialling out region dummies and changes in total MSA population.

Table A.1. List of MSAs

| | | | |
|-------------------------------------|-------------------------------|------------------------------|---------------------------|
| Akron, OH | Flint, MI | New York, NY | Scranton, PA |
| Albany-Schenectady-Troy, NY | Fort Wayne, IN | Omaha, NE/IA | Wilkes-Barre-Hazleton, PA |
| Albuquerque, NM | Fresno, CA | Peoria, IL | Seattle-Everett, WA |
| Allentown-Bethlehem- Easton, PA | Grand Rapids, MI | Philadelphia, PA/NJ | Sioux City, IA/NE |
| Altoona, PA | Green Bay, WI | Phoenix-Mesa, AZ | Sioux Falls, SD |
| Atlantic City, NJ | Hamilton-Middletown, OH | Pittsburgh, PA | South Bend, IN |
| Baltimore, MD | Harrisburg-Leban-Carlisle, PA | Pittsfield, MA | Spokane, WA |
| Binghamton, NY | Hartford, CT | Portland, ME | Springfield, IL |
| Boston, MA | Huntington, WV/KY/OH | Portland-Vancouver, OR/WA | Springfield, MO |
| Bridgeport, CT | Indianapolis, IN | Providence, RI | Springfield-Holyoke, MA |
| Brockton, MA | Jackson, MI | Fall River, MA/RI | Springfield, OH |
| Buffalo, NY | Johnstown, PA | Pueblo, CO | Stockton, CA |
| Canton, OH | Kalamazoo, MI | Racine, WI | Syracuse, NY |
| Cedar Rapids, IA | Kansas City, MO/KS | Reading, PA | Tacoma, WA |
| Chicago, IL | Kenosha, WI | San Bernardino, CA | Terre Haute, IN |
| Cincinnati-Hamilton, OH/KY/IN | Lancaster, PA | Rochester, NY | Toledo, OH/MI |
| Cleveland, OH | Lansing-East Lansing, MI | Rockford, IL | Topeka, KS |
| Columbus, OH | Lima, OH | Sacramento, CA | Trenton, NJ |
| Davenport-Moline-Rock Island, IA/IL | Lincoln, NE | Saginaw-Bay City-Midland, MI | Utica-Rome, NY |
| Dayton, OH | Lorain-Elyria, OH | Bay City, MI | Washington DC, MD/VA/WV |
| Decatur, IL | Los Angeles-Long Beach, CA | St. Joseph, MO | Waterloo-Cedar Falls, IA |
| Denver, CO | Louisville, KY/IN | St. Louis, MO/IL | Wheeling, WV/OH |
| Des Moines, IA | Manchester, NH | Salt Lake City, UT | Wichita, KS |
| Detroit, MI | Milwaukee-Waukesha, WI | Ogden, UT | Wilmington, DE/NJ/MD |
| Duluth-Superior, MN/WI | Minneapolis-St. Paul, MN | San Diego, CA | Worcester, MA |
| Erie, PA | Muncie, IN | San Francisco, CA | York, PA |
| Evansville, IN/KY | New Haven-Meriden, CT | San Jose, CA | Youngstown-Warren, OH |

Table A.2. Summary statistics (individual level)

| Sample | Repeated Cross-Section | | | Linked Sample | | |
|---------------------|------------------------|--------|------------|---------------|-------|---------|
| | Mean | SD | N | Mean | SD | N |
| Married with native | 0.076 | 0.265 | 9,323,128 | 0.105 | 0.306 | 228,677 |
| Naturalized | 0.517 | 0.500 | 15,267,853 | 0.735 | 0.441 | 293,350 |
| In manufacture | 0.252 | 0.434 | 14,055,931 | 0.213 | 0.409 | 269,657 |
| Foreign name index | 59.065 | 28.342 | 4,499,505 | - | - | - |
| Endogamous marriage | 0.780 | 0.414 | 9,323,128 | 0.698 | 0.459 | 228,677 |
| Unskilled | 0.411 | 0.492 | 14,055,931 | 0.278 | 0.448 | 269,657 |
| Employed | 0.888 | 0.315 | 14,055,931 | 0.912 | 0.283 | 269,657 |
| Speaks English | 0.871 | 0.335 | 15,365,327 | 0.951 | 0.216 | 302,619 |
| Literate | 0.875 | 0.331 | 15,889,418 | 0.937 | 0.243 | 312,635 |

Notes: Note: The repeated cross-section sample consists of foreign born men living in the 108 non-southern MSAs for which the instrument could be constructed, in Census years 1910, 1920, and 1930. The linked sample of non-movers consists of foreign born men who could be linked across Census years (as described in the main text) always living in one of the 108 non-southern MSAs in the three Census years, 1910, 1920, and 1930. *Married with native* and *Endogamous marriage* refer to the share of immigrant men who are married, respectively, with a native born spouse of native parentage and with a spouse born in the same country of origin. When computing the share in manufacture and unskilled, the sample is restricted to immigrant men aged 15-65. See Section [3.1](#) for the construction of the Foreign name index.

Table A.3. Comparison of linked and non-linked records

| | Linked | Not-Linked | Difference |
|--------------------------------------|--------------------|--------------------|----------------------|
| <u>Demographic characteristics</u> | | | |
| Age | 32.716 (11.440) | 35.113 (12.470) | -2.397*** (0.014) |
| Years in the US | 17.686 (11.295) | 16.177 (12.951) | 1.509*** (0.025) |
| Northern Europe | 0.039 (0.195) | 0.058 (0.235) | -0.019*** (0.000) |
| UK | 0.067 (0.249) | 0.044 (0.206) | 0.022*** (0.000) |
| Ireland | 0.021 (0.142) | 0.040 (0.196) | -0.020*** (0.000) |
| Western Europe | 0.032 (0.176) | 0.015 (0.123) | 0.017*** (0.000) |
| Southern Europe | 0.044 (0.205) | 0.076 (0.265) | -0.032*** (0.000) |
| Central and Eastern Europe | 0.032 (0.176) | 0.080 (0.271) | -0.048*** (0.000) |
| Germany | 0.065 (0.250) | 0.088 (0.284) | -0.022*** (0.000) |
| Russian Empire | 0.036 (0.185) | 0.064 (0.245) | -0.028*** (0.000) |
| <u>Socioeconomic characteristics</u> | | | |
| Married | 0.586 (0.492) | 0.548 (0.500) | 0.038*** (0.001) |
| Married with native | 0.220 (0.414) | 0.224 (0.417) | -0.004*** (0.001) |
| Endogamous marriage | 0.672 (0.469) | 0.700 (0.458) | -0.027*** (0.001) |
| In manufacture | 0.227 (0.419) | 0.191 (0.393) | 0.036*** (0.000) |
| Unskilled | 0.311 (0.463) | 0.379 (0.485) | -0.068*** (0.001) |
| Naturalized | 0.612 (0.487) | 0.491 (0.500) | 0.121*** (0.001) |
| N | 799,193 | 7,727,528 | |
| Match rate | | 9.4% | |

Notes: The table reports means and standard deviations (in parentheses) for various characteristics of matched and unmatched records. The numbers in parentheses under the mean differences are standard errors on point estimates of differences. *Linked* refers to records that could be matched forward from 1910 to both 1920 and 1930.

Table A.4. First stage with push instrument

| Dep. variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|--------------------|---------------------|--------------------|---------------------|---------------------|---------------------|---------------------|
| | | | | Number of blacks | | | |
| Predicted black inflows | 0.553** (0.266) | 0.781*** (0.252) | 0.717** (0.284) | 0.710*** (0.233) | 0.808*** (0.250) | 0.659*** (0.202) | 0.778*** (0.256) |
| Observations | 324 | 324 | 324 | 324 | 324 | 324 | 324 |
| R-squared | 0.952 | 0.969 | 0.887 | 0.970 | 0.973 | 0.972 | 0.969 |
| F-stat | 12.77 | 21.34 | 8.134 | 22.22 | 21.05 | 22.55 | 20.70 |
| MSAs | 108 | 108 | 108 | 108 | 108 | 108 | 108 |
| Region \times Year FE | | X | X | X | X | X | X |
| Weighted | | X | | X | X | X | X |
| Fr. black 1900 \times Year FE | X | | | X | | | |
| Fr. foreign-born 1900 \times Year FE | | | | X | | | |
| Log Value manuf. output 1900 \times Year FE | | | | | X | | |
| Fr. working age males in manuf. 1900 \times Year FE | | | | | | X | |

Notes: The analysis is restricted to 108 non-southern MSAs for which the IV could be constructed, and to years 1910 to 1930. All regressions control for total population and MSA fixed effects. Column (1) includes year fixed effects. Standard errors clustered at the MSA level in parentheses. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A.5. Assimilation and assimilation effort – Alternative measures

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|----------------------|----------------------|----------------------|----------------------|---------------------|---------------------|
| | OLS | OLS | 2SLS | 2SLS | 2SLS | 2SLS |
| Panel A: In manufacture (<i>1910 mean: 0.272</i>) | | | | | | |
| Num. Blacks | -0.027*** (0.010) | -0.034*** (0.011) | -0.058** (0.027) | -0.056** (0.025) | -0.054** (0.025) | -0.020** (0.010) |
| Observations | 14,055,931 | 14,055,931 | 14,055,931 | 14,055,929 | 14,055,929 | 261,867 |
| R-squared | 0.070 | 0.070 | 0.070 | 0.078 | 0.080 | 0.016 |
| F-stat | | | 24.54 | 24.72 | 24.82 | 28.56 |
| Panel B: Log Foreign name index (<i>1910 mean: 3.683</i>) | | | | | | |
| Num. Blacks | -0.036 (0.025) | -0.053** (0.023) | -0.103*** (0.033) | -0.086*** (0.029) | -0.050** (0.022) | - |
| Observations | 4,499,505 | 4,499,505 | 4,499,505 | 4,499,394 | 4,499,394 | - |
| R-squared | 0.043 | 0.043 | 0.043 | 0.049 | 0.053 | - |
| F-stat | | | 24.87 | 26.59 | 27.04 | - |
| Individual controls | X | X | X | X | X | |
| Region × Year | | X | X | X | X | |
| MSA × Origin region | | | | X | X | |
| Origin region × Year | | | | | X | |
| Linked sample | | | | | | X |

Notes: The table presents results for men living in the 108 non-southern MSAs for which the instrument could be constructed in Census years 1910, 1920, and 1930. In panel A the sample consists of foreign-born men aged 15-65. In panel B it consists of US-born men of foreign-born fathers who were born in the 10 years before each census year. The construction of the Foreign name index is explained in Section 3.1. Col 6 of Panel A shows results from the linked panel of individuals who always remained in the same MSA in the three Census years. Individual controls include fixed effects for age and years in the US (Panel A) or birth year (Panel B) and origin region. All regressions control for MSA and year fixed effects and for total MSA population. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.6. Frequency of national stereotypes

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|--|----------------------|----------------------|----------------------|----------------------|----------------------|---------------------|-------------------|---------------------|
| | Irish&Catholic | Irish&alcohol | Irish&crime | Irish&violent | Italian&Catholic | Italian&alcohol | Italian&crime | Italian&mafia |
| Panel A: Normalized by frequency of word "and" | | | | | | | | |
| Num. Blacks | -0.464** (0.199) | -1.477*** (0.472) | -1.298*** (0.399) | -0.848*** (0.298) | -0.996*** (0.376) | -1.039** (0.450) | -0.581 (0.391) | -3.786 (4.092) |
| Observations | 197 | 197 | 197 | 197 | 197 | 197 | 197 | 197 |
| R-squared | 0.723 | 0.713 | 0.723 | 0.799 | 0.818 | 0.630 | 0.778 | 0.084 |
| F-stat | 17.51 | 17.51 | 17.51 | 17.51 | 17.51 | 17.51 | 17.51 | 17.51 |
| MSAs | 69 | 69 | 69 | 69 | 69 | 69 | 69 | 69 |
| Panel B: Normalized by frequency of the stereotype | | | | | | | | |
| Num. Blacks | -0.733*** (0.274) | -1.214** (0.480) | -1.027*** (0.293) | -0.789 (0.523) | -0.425 (0.260) | -0.569 (0.379) | -0.149 (0.320) | -1.026** (0.439) |
| Observations | 192 | 188 | 194 | 192 | 192 | 188 | 194 | 187 |
| R-squared | 0.389 | 0.675 | 0.389 | 0.255 | 0.718 | 0.397 | 0.640 | 0.617 |
| F-stat | 17.66 | 17.75 | 17.63 | 17.82 | 17.66 | 17.75 | 17.63 | 18.24 |
| MSAs | 68 | 66 | 68 | 68 | 68 | 66 | 68 | 67 |

Notes: 2SLS estimates reported. The analysis is restricted to 65 non-southern MSAs for which the IV and the word frequency measure could be constructed, and to years 1910 to 1930. Dependent variables are standardized by subtracting the mean and dividing by the standard deviation in 1910. All regressions control for MSA and year by region fixed effects and MSA total population and are weighted by population in 1900. Standard errors clustered at the MSA level in parentheses. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A.7. Heterogeneity by national origin - European immigrants

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|----------------------------|-------------------|---------------------|--------------------|-------------------|-------------------|----------------------------|-------------------|-------------------|
| | Northern Europe | UK | Ireland | Western Europe | Southern Europe | Central and Eastern Europe | Germany | Russian Empire |
| Panel A: Married w/ native | | | | | | | | |
| Num. blacks | 0.024 (0.015) | 0.027*** (0.009) | -0.011* (0.006) | 0.020 (0.017) | 0.004* (0.002) | 0.004** (0.002) | 0.009 (0.009) | -0.003 (0.002) |
| Observations | 640,387 | 835,193 | 674,360 | 278,255 | 1,528,411 | 2,004,635 | 1,295,258 | 1,292,126 |
| R-squared | 0.076 | 0.094 | 0.057 | 0.082 | 0.041 | 0.030 | 0.072 | 0.019 |
| F-stat | 27.72 | 22.85 | 25.67 | 18.32 | 28.44 | 23.61 | 23.25 | 26.68 |
| Panel B: Naturalized | | | | | | | | |
| Num. blacks | -0.020 (0.014) | 0.047*** (0.015) | 0.028** (0.013) | -0.025 (0.029) | 0.039* (0.021) | 0.049*** (0.010) | 0.024* (0.014) | 0.038 (0.027) |
| Observations | 1,095,177 | 1,318,735 | 1,205,241 | 450,354 | 2,630,147 | 3,191,466 | 1,923,930 | 2,004,345 |
| R-squared | 0.322 | 0.321 | 0.301 | 0.312 | 0.228 | 0.288 | 0.364 | 0.289 |
| F-stat | 27.09 | 23.13 | 26.59 | 18.41 | 28.60 | 24.19 | 23.06 | 27.04 |

Notes: 2SLS estimates reported. The analysis is restricted to 108 non-southern MSAs for which the IV could be constructed, and to years 1910 to 1930. All regressions include fixed effects for age, years in the US, MSA and year by region, and control for total population. The sample is restricted to married men in Panel A. Standard errors clustered at the MSA level in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A.8. Effects for non-Europeans

| | (1) | (2) | (3) | (4) | (5) | (6) |
|--------------|-------------------|--------------------|------------------|-------------------|-------------------|-------------------|
| | Mexico | Canada | China | Mexico | Canada | China |
| | Married w/ native | | | Naturalized | | |
| Num. blacks | -0.032 (0.026) | 0.028** (0.013) | 0.012 (0.062) | -0.012 (0.016) | 0.052* (0.028) | -0.003 (0.016) |
| Observations | 81,415 | 686,760 | 6,252 | 211,618 | 1,137,272 | 99,490 |
| R-squared | 0.106 | 0.102 | 0.191 | 0.102 | 0.274 | 0.054 |
| F-stat | 66.81 | 11.05 | 79.43 | 73.53 | 12.20 | 84.24 |

Notes: 2SLS estimates reported. The analysis is restricted to 108 non-southern MSAs for which the IV could be constructed, and to years 1910 to 1930. All regressions include fixed effects for age, years in the US, MSA and year by region and control for total population. The sample is restricted to married men in Columns 1-3. Standard errors clustered at the MSA level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.9. Public sector employment

| Dep. Variable | Employed in | | |
|---|-----------------------|------------------|------------------|
| | Public administration | Police | Fire protection |
| | (1) | (2) | (3) |
| Panel A: All immigrant men of working age | | | |
| Num. Blacks | -0.165 (0.134) | 0.005 (0.024) | 0.005 (0.012) |
| Observations | 14,599,200 | 14,599,200 | 14,599,200 |
| R-squared | 0.010 | 0.008 | 0.002 |
| F-stat | 25.07 | 25.07 | 25.07 |
| Mean dep. variable | 0.919 | 0.287 | 0.087 |
| Panel B: Immigrant men in the labor force | | | |
| Num. Blacks | -0.130 (0.117) | 0.017 (0.027) | 0.010 (0.014) |
| Observations | 13,005,817 | 13,005,817 | 13,005,817 |
| R-squared | 0.011 | 0.009 | 0.002 |
| F-stat | 24.89 | 24.89 | 24.89 |
| Mean dep. variable | 1.008 | 0.259 | 0.071 |
| Panel C: Immigrant men employed | | | |
| Num. Blacks | -0.128 (0.126) | 0.025 (0.027) | 0.008 (0.013) |
| Observations | 12,958,064 | 12,958,064 | 12,958,064 |
| R-squared | 0.011 | 0.009 | 0.002 |
| F-stat | 25.85 | 25.85 | 25.85 |
| Mean dep. variable | 0.994 | 0.315 | 0.095 |

Notes: The table presents results for immigrant men living in the 108 non-southern MSAs for which the instrument could be constructed in Census years 1910, 1920, and 1930. The dependent variable is an indicator for individuals employed in the sector indicated in each column heading. All regressions include fixed effects for age, nationality, years in the US, MSA and year by region and control for total population. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.10. Local labor unions

| Dep. Variable | Log(IWW Locals) | | | |
|---------------|------------------|------------------|------------------|------------------|
| | (1) | (2) | (3) | (4) |
| Num. Blacks | 0.075 (1.177) | 0.436 (0.529) | 0.175 (0.746) | 0.209 (0.672) |
| Local unions | | Per capita | Over 1900 Pop. | Over 1910 Pop. |
| Observations | 216 | 216 | 216 | 216 |
| R-squared | 0.381 | 0.259 | 0.305 | 0.317 |
| F-stat | 19.55 | 19.55 | 19.55 | 19.55 |

Notes: The table presents results for the 108 non-southern MSAs for which the instrument could be constructed in Census years 1910, 1920, and 1930. All regressions include fixed effects for age, nationality, years in the US, MSA and year by region and control for total population. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B Push instrument

Even if early settlements were as good as randomly assigned, one remaining concern is that black outflows from each southern state, O_{jt} , might have been differentially affected by specific, time-varying shocks in northern destinations. To deal with this potential threat, as in [Boustan \(2010\)](#), we construct a modified version of the instrument in [\(3\)](#) by replacing O_{jt} with predicted (rather than actual) outmigration. In a “zeroth stage”, we start by estimating:

$$mig_{sjt} = \alpha_j + \gamma Push_{sjt-10} + e_{sjt} \quad (6)$$

where mig_{sjt} is the net black migration rate from county s in southern state j between t and $t-10$. We use beginning of decade county “push factors”, $Push_{sjt-10}$, since contemporaneous variables are likely to be themselves affected by outmigration. In our most preferred specification, we include state fixed effects, α_j , and collect in the vector $Push_{sjt-10}$ the following variables: the black share of the population; the share of the population living in rural areas; the share of land cultivated in cotton; and an indicator for the arrival of the boll weevil in the previous decade. [Table B.1](#) presents results for equation [\(6\)](#). In columns 1 to 3 (resp. 4 to 6), we report results obtained without (resp. with) state fixed effects. Reassuringly, and consistent with estimates in [Boustan \(2010, 2016\)](#), the inclusion of state fixed effects does not significantly alter the main message emerging from [Table B.1](#).

In line with the historical evidence (e.g. [Boustan \(2016\)](#)), a higher black share and a higher fraction of the population living in rural areas are associated with larger black departures during the subsequent decade. Also, counties with a larger share of land cultivated in cotton were more likely to attract blacks between 1900 and 1920, but this pattern was reversed during the 1920-1930 decade. Indeed, after 1920, cotton mechanization began to spread around the South, reducing demand for black labor in agriculture and increasing incentives to migrate northward ([Wright, 1986](#)). Finally, in line with findings in [Collins and Wanamaker \(2015\)](#), the arrival of the boll weevil is significantly associated with black outflows only for the 1920-1930 decade.¹⁹

After estimating [\(6\)](#), we compute predicted migration flows from each county by multiplying the fitted values from [\(6\)](#) with the county initial black population. Finally, we aggregate these flows to the state level to obtain the predicted number of blacks leaving each southern state j in each decade, \hat{O}_{jt} . We then replace O_{jt} with \hat{O}_{jt} in [\(3\)](#) to derive the (push-factors induced) predicted number of blacks moving to city c in year t . By construction, this (predicted) measure of black outmigration from the South is orthogonal to any specific shock occurring in the North. Moreover, by exploiting southern shocks to agricultural conditions, this instrument is less likely to suffer from the problem of high serial correlation in migration

¹⁹Results are very similar when including only a subset of the push factors used in [\(6\)](#), or when adding additional controls such as the share of a county cultivated with tobacco, the presence of railroads, or average farm values.

Table B.1. Zeroth stage

| Dep. variable | (1) | (2) | (3) | (4) | (5) | (6) |
|----------------|--------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| | Net black migration rate | | | | | |
| Share Blacks | -0.170*** (0.048) | -0.041 (0.045) | -0.215*** (0.052) | -0.165*** (0.058) | -0.075* (0.045) | -0.184*** (0.052) |
| Rural Share | -0.257*** (0.064) | -0.163*** (0.045) | -0.333*** (0.052) | -0.230*** (0.059) | -0.143*** (0.044) | -0.302*** (0.048) |
| Share Cotton | 0.292*** (0.105) | 0.295*** (0.100) | -0.105 (0.085) | -0.167 (0.171) | 0.291** (0.128) | -0.233** (0.099) |
| 1[Boll Weevil] | -0.034 (0.051) | 0.030 (0.019) | -0.052** (0.020) | -0.205*** (0.073) | 0.053 (0.036) | -0.000 (0.043) |
| Observations | 1,002 | 989 | 937 | 1,002 | 989 | 937 |
| R-squared | 0.030 | 0.036 | 0.104 | 0.166 | 0.110 | 0.245 |
| State FEs | | | | X | X | X |
| Decade | 1900-1910 | 1910-1920 | 1920-1930 | 1900-1910 | 1910-1920 | 1920-1930 |

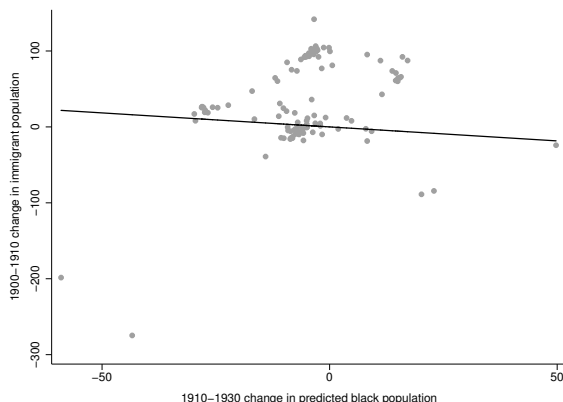
Notes: Each observation is a southern county. Net black migration rates are constructed using the forward survival method (Gregory, 2005). Data is from the full count IPUMS and ICPSR. Standard errors clustered at the state level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

patterns between sending and receiving areas – a possible concern for standard shift-share instruments (Jaeger, Ruist and Stuhler, 2018). We show that both first stage and 2SLS results are robust to using this, instead of our baseline version of the instrument.

C Robustness Checks

The validity of our identification strategy rests on the assumption that 1900 black settlements are uncorrelated to characteristics of northern places which vary within regions and which affected immigrant assimilation.²⁰ For instance, if early blacks were more likely to settle in booming northern places that also attracted foreign born individuals, and if these immigrants were in turn more likely to assimilate, we would be erroneously attributing this assimilation to black in-migration. To address this and related concerns we perform a number of robustness and falsification tests. We start by showing that the 1900 to 1910 change in European immigration is not correlated with subsequent black inflows between 1910 and 1930, as predicted by the instrument (Figure C.1). This result is very important, and suggests that predicted black inflows and European migration patterns did not overlap, and thus immigrant outcomes were unlikely to be influenced by unobservables correlated with black inflows.

Figure C.1. Pre-trends



The figure represents the residual scatterplot for a regression of 1900-1910 change in the number of immigrants (y-axis) against the 1910 to 1930 change in predicted black in-migration (x-axis), after partialling out region dummies and changes in total MSA population.

Next, in Table C.1 (columns 3 and 4), we more formally test for pre-trends by regressing the 1900 to 1910 change in the key outcomes of interest (see Table 3) against the 1910 to 1930 change in black population as predicted by the instrument. Reassuringly, in all cases, coefficients are statistically indistinguishable from zero and quantitatively different from 2SLS estimates from our baseline specification which, for convenience, are reported in columns 1 and 2 of Table C.1.²¹ These results indicate that, before 1900, southern born blacks did not systematically settle in MSAs that were already undergoing immigrant social or economic

²⁰As discussed and formally shown in Goldsmith-Pinkham, Sorkin and Swift (2018), the key identifying assumption of Bartik-style instruments (a class of instruments that includes the shift-share) is best stated in terms of the initial shares.

²¹Specifically, columns 1 and 2 report results presented in column 3 of Table 3.

Table C.1. Pre-trends

| | (1) | (2) | (3) | (4) |
|--------------|------------------------|--------------------|--|-------------------|
| | Married w/ native | Naturalized | Married w/ native | Naturalized |
| | Baseline specification | | Dependent variable is the 1900-1910 change | |
| Num. Blacks | 0.012** (0.005) | 0.033** (0.015) | -0.028 (0.034) | -0.030 (0.030) |
| Observations | 9,323,126 | 15,267,846 | 2,919,024 | 4,992,949 |
| R-squared | 0.124 | 0.346 | 0.422 | 0.423 |
| F-stat | 23.33 | 24.33 | 13.56 | 13.84 |

Notes: 2SLS estimates reported. The analysis is restricted to 108 non-southern MSAs for which the IV could be constructed, and to years 1910 to 1930. The sample consists of foreign-born men, and is further restricted to married individuals in columns 1 and 3. Columns 1 and 2 replicate the specification in column 3 of Table 3 for marriage and naturalization. Columns 3 and 4 estimate the effect of the 1910-1930 change in black population (instrumented) on the 1900-1910 change in marriage and naturalization. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

As a further check, we show that our results are robust to interacting year dummies with several 1900 MSA characteristics, including the share of blacks, the share of foreign born, the value of manufacturing output, and the share of workers in manufacturing (Table C.2). In the first row of Table C.2, we report the F-stat as well as the 2SLS coefficient on black population from our baseline specification, while in subsequent rows we present results from each different specification. As it appears, the F-stat always remains above conventional levels, and the point estimate is stable across specifications, suggesting that the characteristics which attracted more blacks (from different southern states) before 1900 did not correlate with stronger (or weaker) economic and social assimilation of immigrants. In particular, controlling for the interaction between the 1900 share of blacks and year dummies (second row of Table C.2) implies that the effects of black in-migration are identified exploiting variation only in the (southern state) composition of African Americans' enclaves across MSAs, holding constant the size of their black populations.

A related concern is that our instrument might be spuriously correlated with the differential impact of the immigration quotas across MSAs. The Immigration Acts restricted access to the US disproportionately more for immigrants from Eastern and Southern Europe, and so the bite of the quotas was stronger in MSAs that had a larger concentration of immigrants from these sending regions. If settlements of southern born blacks were correlated with en-

²²In unreported results, we also checked that the 1910 to 1930 change in predicted black inflows is not correlated with the 1900 to 1910 change in employment and manufacturing activity.

Table C.2. Robustness

| Dep. variable | (1) Married w/ native | (2) Naturalized | (3) In manufacture | (4) Log Foreign name index |
|--|--------------------------|---------------------|-----------------------|-------------------------------|
| 2SLS coefficient on number of blacks | | | | |
| Baseline | 0.012** (0.005) | 0.033** (0.014) | -0.054** (0.025) | -0.050** (0.022) |
| F-stat | 23.83 | 24.51 | 24.82 | 27.04 |
| Share blacks 1900 \times Year FE | 0.012*** (0.004) | 0.033** (0.014) | -0.049 (0.031) | -0.064*** (0.023) |
| F-stat | 25.13 | 25.04 | 25.81 | 30.59 |
| Share foreign-born 1900 \times Year FE | 0.012** (0.006) | 0.044*** (0.017) | -0.061** (0.030) | -0.055** (0.025) |
| F-stat | 19.66 | 20.20 | 20.36 | 20.91 |
| Log Value manufacturing output 1900 \times Year FE | 0.013** (0.005) | 0.027* (0.015) | -0.051* (0.026) | -0.047* (0.027) |
| F-stat | 23.24 | 23.81 | 24.03 | 25.50 |
| Share in manufacture 1900 \times Year FE | 0.012** (0.005) | 0.032** (0.014) | -0.052** (0.022) | -0.050** (0.022) |
| F-stat | 23.64 | 24.54 | 24.84 | 26.83 |
| All 1900 controls \times Year FE | 0.012** (0.005) | 0.033* (0.019) | -0.050 (0.035) | -0.066** (0.033) |
| F-stat | 19.35 | 19.88 | 20.28 | 21.75 |
| Observations | 9,323,109 | 15,267,844 | 14,055,929 | 4,499,394 |

Notes: The analysis is restricted to 108 non-southern MSAs for which the IV could be constructed, and to years 1910 to 1930. The sample consists of foreign-born men in columns 1 to 3 and US-born men of foreign-born fathers who were born in the 10 years before the census year in column 4. It is further restricted to married individuals in column 1 and to individuals aged 15-65 in column 3. All regressions include fixed effects for age, years in the US, MSA by nationality, nationality by year and year by region, and control for MSA total population. Standard errors clustered at the MSA level in parentheses. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

claves of specific groups of European immigrants, we might be incorrectly attributing to black in-migration the separate effects that the reduction in European immigration could have had on immigrants’ assimilation.

To address this concern, we construct a measure of “quota exposure” that captures the extent to which each MSA was affected by the Immigration Acts. Specifically, following [Ager and Hansen \(2017\)](#), we define

$$Quota_Shock_n = \frac{1}{Pop_{n,1920}} \sum_{k \in Europe} \lambda_{kn} \left(\hat{M}_{k,22-30} - Q_{k,22-30} \right) \quad (7)$$

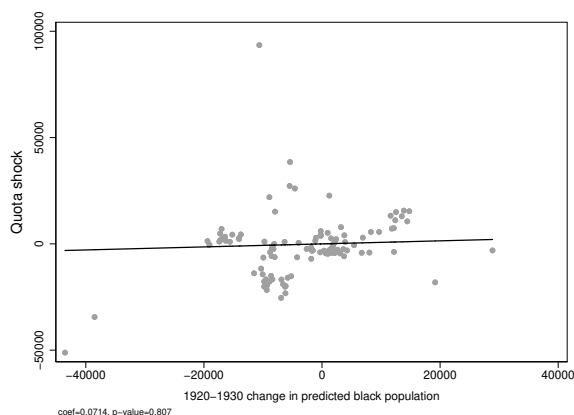
the share of immigrants (relative to the 1920 MSA population) “lost” by each MSA between 1921 and 1930 due to the quotas. In particular, $\hat{M}_{k,22-30}$ is the predicted number of immigrants from k that would have entered the US, had the quota system not been introduced. As in [Ager and Hansen \(2017\)](#), $\hat{M}_{k,22-30}$ is predicted by first estimating a regression of the form: $M_{kt} = \beta_1 \ln(t) + \beta_2 \ln(t^2) + \varepsilon_{kt}$, where M_{kt} is the actual number of immigrants from country k in each year t between 1900 and 1914.²³ $Q_{k,22-30}$ is the total number of immigrants from k that were allowed to enter the US according to the yearly quotas.²⁴ Whenever the difference between $\hat{M}_{k,22-30}$ and $Q_{k,22-30}$ in (7) is negative, i.e. whenever the quotas were not binding, we set it to zero (but results are unchanged when we allow $(\hat{M}_{k,22-30} - Q_{k,22-30})$ to be negative). The “missing” immigrants from each sending country are apportioned across MSAs according to the share of individuals from k who were living in MSA n in 1900, relative to all immigrants from k in the US in that year, $\lambda_{kn} \equiv \frac{Imm_{kn}^{900}}{Imm_k^{900}}$. Finally, for each MSA, we sum over all immigrant groups k to obtain the total number of missing immigrants in MSA n between 1922 and 1930, and we then divide this number by the 1920 MSA population ($Pop_{n,1920}$).

With this variable at hand, we can check that the 1920 to 1930 change in predicted black in-migration, i.e. the instrument used in our paper, is uncorrelated with the 1920-1930 quota exposure across MSAs. Results for this exercise are reported in [Figure C.2](#): reassuringly, there is no correlation between the predicted change in black population (x-axis) and the quota shock defined in equation (7) above (y-axis). This result strongly suggests that the effects of black inflows on immigrant assimilation are not driven by the differential effect that the quota system might have had across MSAs.

²³Data were taken from [Ferenczi and Willcox \(1929\)](#); see footnote 29 in [Ager and Hansen \(2017\)](#) for a detailed description of this data.

²⁴The original quota system introduced in 1921 was revised in 1924 to make the immigration restrictions even more stringent (see [Goldin \(1994\)](#)).

Figure C.2. Correlation of quota shock with instrument



The figure represents the residual scatterplot for a regression of the “quota shock”, as specified in [Ager and Hansen \(2017\)](#) and described in the text (y-axis) against the 1910 to 1930 change in predicted black in-migration (x-axis), after partialling out region dummies and changes in total MSA population.

We perform an additional check to ensure that our instrument is orthogonal to the effect of the quotas. If predicted black population captures disproportionately larger movements of African Americans to MSAs more affected by the quotas, we would expect changes in the instrument to be correlated with changes in immigrant characteristics likely caused by the quotas. One such characteristic is years spent in the US: in MSAs where the quotas were more binding, the average length of stay of immigrants in the country should increase. In [Table C.3](#) we estimate our baseline specification with years spent in the US as a dependent variable. There is no correlation between our instrument and the length of stay of the average immigrant in MSAs in our sample. Because the average number of years spent in the country masks substantial heterogeneity – depending on whether a nationality was favored or not by the quotas – columns 2 to 9 of [Table C.3](#) repeat this exercise separately by region of immigrant origin. With the exception of the UK – for which average years in the US increase, despite the fact that new arrivals from this country were favored by the quotas – there is no indication that our instrument correlates with quota-induced changes in immigrant profiles, that could also be correlated to better assimilation outcomes.

Yet another set of concerns is related to the possibility that the Great Migration triggered a (selective) European flight. This would be problematic because our estimates could then be affected by compositional changes. A related possibility is that, if the arrival of blacks was associated with smaller enclaves of international immigrants (due either to immigrants flight or to lower in-migration rates), the declining size of immigrant population, rather than the inflow of a new out-group, might have favored the Americanization of Europeans. We tackle this issue in several ways. First, as already shown above, all our results are robust, and if anything stronger, when using a linked sample of individuals who were always observed in the same MSA in each census year between 1910 and 1930.

Table C.3. Effect on average years in the US

| Dep. variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---------------|-------------------------|-------------------|-------------------|------------------|------------------|-------------------|------------------------|-------------------|-------------------|
| | Average years in the US | | | | | | | | |
| | All nationalities | Northern EU | UK | Ireland | Western EU | Southern EU | Central and Eastern EU | Germany | Russia |
| Num. blacks | 0.097 (0.406) | -0.411 (0.447) | 1.672* (0.903) | 0.330 (0.671) | 0.783 (0.552) | -0.251 (0.179) | -0.048 (0.522) | -0.648 (0.683) | -0.062 (0.332) |
| Observations | 15,889,409 | 1,119,820 | 1,359,759 | 1,217,539 | 463,489 | 2,777,117 | 3,321,056 | 1,964,497 | 2,162,993 |
| R-squared | 0.606 | 0.662 | 0.559 | 0.686 | 0.586 | 0.488 | 0.539 | 0.618 | 0.501 |
| F-stat | 24.64 | 27.35 | 22.98 | 26.57 | 18.28 | 28.58 | 25.04 | 23.02 | 26.78 |

Notes: The analysis is restricted to 108 non-southern MSAs for which the IV could be constructed, and to years 1910 to 1930. Column 1 replicates the specification of column 5, Table 3 without indicators for years in the US. Columns 2-9 replicate the specification in column 3 of Table 3 separately for each region of origin. Standard errors clustered at the MSA level in parentheses. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table C.4. Placebo check: actual immigrants

| | (1) | (2) | (3) | (4) |
|-------------------------|-------------------------------------|-------------------|-------------------|-------------------|
| | Dep. Variable: Number of immigrants | | | |
| Num. blacks | -0.820 (0.610) | -0.663 (0.429) | -0.261 (0.471) | -0.074 (0.327) |
| Immigrants | All | Europeans | All | Europeans |
| Year by 1900 Population | | | X | X |
| F-stat | 17.46 | 17.46 | 16.66 | 16.66 |
| Observations | 324 | 324 | 324 | 324 |

Notes: 2SLS estimates reported. The analysis is restricted to 108 non-southern MSAs for which the IV could be constructed, and to years 1910 to 1930. All regressions control for MSA and region by year fixed effects, and total MSA population and are weighted by 1900 MSA population. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Second, we show that the inflow of African Americans did not have any statistically significant effect on the number of international immigrants (Table C.4). While the point estimate is negative, standard errors are very large, and coefficients are never significant at conventional levels. Moreover, once we control for the interaction between 1900 population and year dummies, the coefficient moves closer to zero and, again, remains statistically indistinguishable from zero (columns 3 and 4). Also, and importantly, in Table C.5 we document that the inflow of blacks did not significantly affect the share of immigrants (over all foreign born individuals) in an MSA. This finding further reduces concerns that the Great Migration triggered compositional changes in the population of international immigrants.

In Table C.6, we also check that black in-migration did not alter the sex ratio within the immigrant group, something that could be driving our results on intermarriage. Reassuringly, there is no statistically significant relationship between black inflows and sex ratios, neither for younger cohorts nor for older ones.²⁵ In unreported results, we also verified that our findings in Table 3 (Panel A) and C5 (column 1) remain unchanged when controlling directly for sex ratios.

²⁵Following Angrist (2002), we compute sex ratios for younger (resp. older) cohorts by taking the ratio of foreign born men 20-35 (resp. above 35) over foreign born women 18-33 (resp. above 33). Results are very similar also when considering the second generation, and when looking at sex ratios separately for each ethnic group.

Table C.5. Change in the ethnic composition of the foreign-born

| Dep. variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---------------|--------------------------------------|-------------------|--------------------|-------------------|------------------|------------------------|------------------|------------------|
| | Share of immigrants from each region | | | | | | | |
| | Northern EU | UK | Ireland | Western EU | Southern EU | Central and Eastern EU | Germany | Russia |
| Num. blacks | -0.007 (0.005) | -0.014 (0.009) | -0.014* (0.007) | -0.005 (0.004) | 0.019 (0.012) | 0.016 (0.017) | 0.001 (0.022) | 0.006 (0.013) |
| Observations | 324 | 324 | 324 | 324 | 324 | 324 | 324 | 324 |
| R-squared | 0.157 | 0.331 | 0.862 | 0.428 | 0.799 | 0.750 | 0.808 | 0.527 |
| F-stat | 17.46 | 17.46 | 17.46 | 17.46 | 17.46 | 17.46 | 17.46 | 17.46 |

Notes: The analysis is restricted to 108 non-southern MSAs for which the IV could be constructed, and to years 1910 to 1930. All regressions control for MSA and year by region fixed effect, total population and are weighted by population in 1900. Standard errors clustered at the MSA level in parentheses. Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1.

Table C.6. Sex ratios

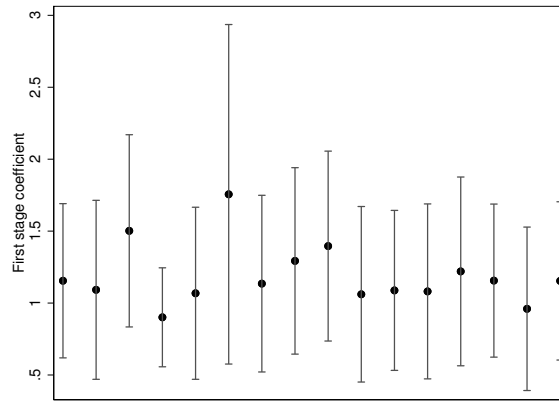
| Dep. variable | (1) | (2) | (3) |
|---------------|-------------------|-------------------|-------------------|
| | | Sex ratios | |
| | All | Young | Old |
| Num. blacks | -0.084 (0.102) | -0.042 (0.113) | -0.102 (0.089) |
| F-stat | 17.46 | 17.46 | 17.46 |
| Observations | 324 | 324 | 324 |

Notes: The analysis is restricted to 108 non-southern MSAs for which the IV could be constructed, and to years 1910 to 1930. All regressions control for MSA and year by region fixed effect, total population and are weighted by population in 1900. Standard errors clustered at the MSA level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Additionally, and to more directly tackle the potential concerns on the validity of Bartik instruments discussed in Goldsmith-Pinkham, Sorkin and Swift (2018), we replicate our analysis by interacting year dummies with the share of southern born blacks from each state, i.e. α_{jn}^{1900} in equation (2).²⁶ We plot 2SLS coefficients from this exercise in Figure C.4 for both intermarriage (Panel A) and naturalization rates (Panel B). The first point estimate on the left plots the coefficient from our baseline specification (i.e. column 3 of Table 3), while subsequent point estimates result from regressions including each interaction separately. As it appears, with two exceptions, results for both intermarriage and naturalization rates are very stable and remain close to the baseline effect. Only when interacting year dummies with the share of African Americans from Georgia (resp. Delaware) the point estimate for intermarriage (resp. naturalization) becomes smaller in magnitude and imprecisely estimated. Overall, this exercise is reassuring for the validity of the instrument, since it suggests that results are unlikely to be driven by any specific black enclave that happened to locate in e.g. booming MSAs before 1900.

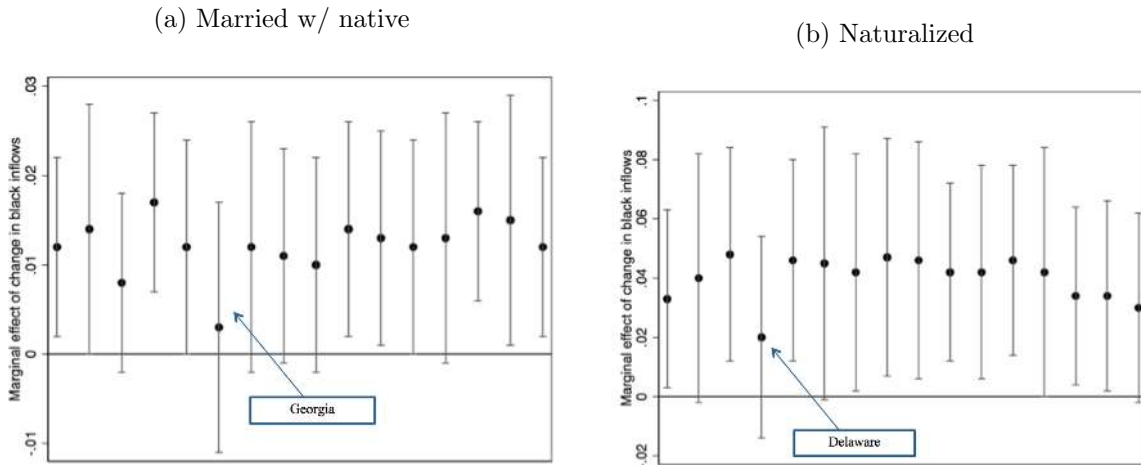
²⁶Specifically, Goldsmith-Pinkham, Sorkin and Swift (2018) argue that the key identification assumption behind Bartik instruments is the exogeneity of the initial shares (α_{jn}^{1900}). Because of perfect multicollinearity, we cannot include interactions between year dummies and all the shares simultaneously. Hence, we perform this exercise by interacting year dummies with α_{jn}^{1900} for each southern state j , one at the time.

Figure C.3. First stage robustness to controlling for black shares from each southern state



The figure plots the first stage estimate and 95% confidence intervals. The first point estimate on the left results from a regression identical to the one reported in Column 2 of Table 2. Each subsequent regression includes an interaction with the 1900 share of blacks from each southern state.

Figure C.4. Robustness to controlling for black shares from each southern state



Each subfigure plots 2SLS estimates of the marginal effect of a change in black inflows on the outcome indicated in each title. The first point estimate on the left results from a regression identical to the one reported in column 5 of Table 3. Each subsequent regression includes an interaction with the 1900 share of blacks from each southern state. Vertical lines are 95% confidence intervals.

The validity of the IV strategy relies on black outflows from each southern state not being driven by time-specific shocks in northern MSAs. To address concerns related to the violation of this assumption, in section 3.3 we construct an alternative version of the instrument that uses county-level push factors to predict net black migration rates from the South, following Boustan (2010). In Table C.7 we replicate the results of Table 3 using this version of the instrument. This exercise leaves both the magnitude and significance of coefficients practically unchanged.

Table C.7. Main results with push instrument

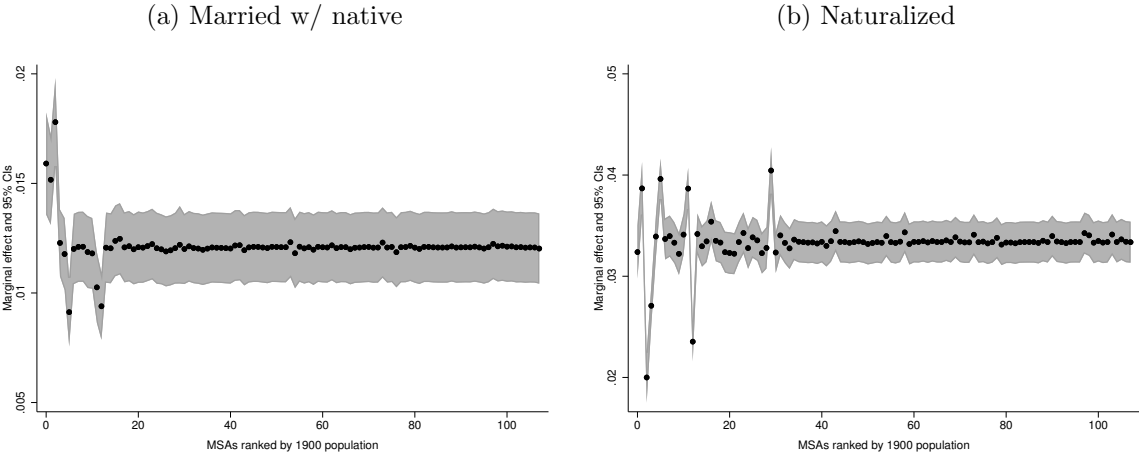
| | (1) | (2) | (3) | (4) | (5) | (6) |
|--|--------------------|-------------------|--------------------|---------------------|---------------------|---------------------|
| | OLS | OLS | 2SLS | 2SLS | 2SLS | 2SLS |
| Panel A: Married w/ native (<i>1910 mean: 0.071</i>) | | | | | | |
| Num. Blacks | 0.007 (0.006) | 0.005 (0.004) | 0.013** (0.005) | 0.016*** (0.004) | 0.014*** (0.004) | 0.025 (0.024) |
| Observations | 9,323,126 | 9,323,126 | 9,323,126 | 9,323,109 | 9,323,109 | 88,892 |
| R-squared | 0.124 | 0.124 | 0.124 | 0.139 | 0.140 | 0.131 |
| F-stat | | | 32.22 | 32.42 | 32.80 | 37.25 |
| Panel B: Naturalized (<i>1910 mean: 0.491</i>) | | | | | | |
| Num. Blacks | 0.056** (0.025) | 0.030* (0.016) | 0.030* (0.016) | 0.038*** (0.014) | 0.034** (0.014) | 0.190*** (0.036) |
| Observations | 15,267,846 | 15,267,846 | 15,267,846 | 15,267,844 | 15,267,844 | 80,866 |
| R-squared | 0.345 | 0.346 | 0.346 | 0.352 | 0.355 | 0.569 |
| F-stat | | | 33.39 | 33.41 | 33.83 | 35.45 |
| Individual controls | X | X | X | X | X | |
| Region \times Year | | X | X | X | X | |
| MSA \times Origin region | | | | X | X | |
| Origin region \times Year | | | | | X | |
| Linked sample | | | | | | X |

Notes: The table presents results for immigrant men living in the 108 non-southern MSAs for which the instrument could be constructed in Census years 1910, 1920, and 1930. In Panel A the sample is restricted to married men. *Married w/ native* is a dummy equal to 1 if the individual is married to a native woman of native parentage. Cols 1 to 5 present results obtained from the repeated cross-sections (Section 3.2.1), while Col 6 shows results from the linked panel of men (Section 3.2.2) who always remained in the same MSA in the three Census years. Cols 1-2 (resp. 3-5) present OLS (resp. 2SLS results). Individual controls include fixed effects for age, years in the US and origin region. All regressions control for MSA and year fixed effects and for total MSA population. In Col 6 of Panel A the sample is restricted to men who were not married in the previous decade (Panel A) or who were not naturalized in the previous decade (Panel B). Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

One remaining concern with a specification in levels is that results might be at least in part driven by outliers. In Figure C.5, we show that, reassuringly, this is not the case. We plot IV coefficients for the effects of the Great Migration on our two key outcomes, i.e. intermarriage and naturalization, by dropping each MSA in our sample at a time, ranked by 1900 population. While the exclusion of large MSAs like New York or Chicago has some impact on the magnitude and on the precision of coefficients, our estimates are stable and not driven by any one MSA in particular. Figure C.6 conducts a similar exercise for the first stage F-statistic. While omitting Chicago, the second largest MSA in 1900, from the sample affects

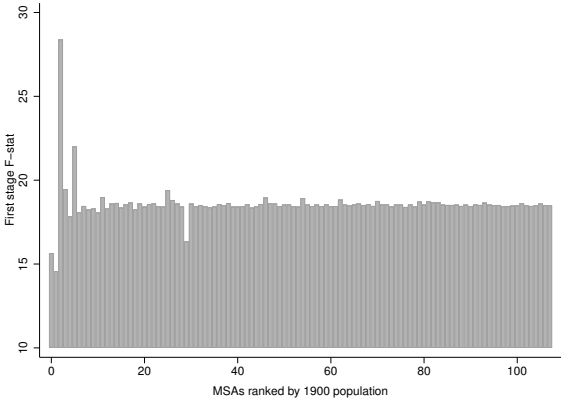
the power of the first stage, the F-statistic is always larger than 10. These results suggest that our findings are not driven by outliers.

Figure C.5. Robustness to dropping outliers



Each subfigure plots 2SLS estimates of the marginal effect of a change in black inflows on the outcome indicated in each title from regressions identical to the one reported in column 5 of Table 3 that drop one MSA at a time. MSAs are ranked according to 1900 population.

Figure C.6. Robustness to dropping outliers - First stage



The figure plots the F-statistic of the first stage as specified in Column 2 of Table 2 dropping one MSA at a time. MSAs are ranked by 1900 population.

D Additional outcomes

D.1 Additional Economic and Social Outcomes

In Section 4, we showed that the Great Migration increased the probability of intermarriage between immigrants and natives and lowered the share of foreign born working in the “immigrant-intensive” sector, namely manufacturing. We interpreted these findings as evidence that black inflows fostered assimilation of the previous out-group. We also showed that the arrival of African Americans fostered immigrants’ effort to assimilate, proxied as the share of naturalized immigrants and the names chosen by foreign born parents for their kids. In Table D.1, we explore the robustness of our main results to the use of alternative proxies for successful assimilation, and investigate the effects of the Great Migration on additional outcomes.

First, we document that black in-migration lowered the frequency of endogamous marriage among immigrants (column 1). While the point estimate is not statistically significant in the repeated cross-sections sample (Panel A), when turning to the linked dataset (Panel B), the 2SLS coefficient is precisely estimated, and the implied magnitude is close to that reported in the main text. Second, in column 2, we proxy for occupational or skill upgrading using the share of immigrants working in the unskilled sector. As it appears, consistent with our findings for manufacturing (see Table A.5, Panel A), there is a negative and significant effect of black inflows on the share of immigrants working in the unskilled sector.

In column 3, we show that this was not due to immigrants becoming unemployed: indeed, we do not find any statistically significant effect of black arrivals on the employment to population ratio of immigrant men in working age. Also, in unreported regressions we found that black inflows raised the share of immigrant men who were homeowners, suggesting that, if anything, immigrants’ socio-economic status increased because of the Great Migration.

Finally, in columns 4 and 5 we investigate the impact of black in-migration on immigrants’ ability to speak English and literacy. As for endogamy, results in the cross-sectional sample are not statistically significant. However, when turning to the linked dataset and restricting attention to individuals who were illiterate in the previous period, we find a positive and significant effect of black in-migration to immigrants’ ability to read and write. This suggests that effects were present on literacy, but concentrated within groups that could actually respond to black migration. Similarly, while the coefficient for English proficiency in column 4 is not statistically significant, it suggests that black inflows might have increased the share of immigrants who could speak English among those who did not speak English prior to black arrivals. While we do not want to over-emphasize these findings, we nonetheless view them as consistent with our previous results for both immigrants’ skill upgrading and assimilation effort.

The fact that results for the linked sample are stronger than those from the repeated cross-sections sample in columns 4 and 5 should not be surprising for an additional reason: after 1920, new immigrants were significantly more likely to be literate and to speak English. On the one hand, the literacy test of 1917 restricted access to the US to immigrants who

Table D.1. Additional outcomes

| | (1) | (2) | (3) | (4) | (5) |
|------------------------|----------------------|----------------------|-------------------|------------------|--------------------|
| | Endogamous | Unskilled | Employed | Speak English | Literacy |
| Repeated cross-section | | | | | |
| Num. Blacks | -0.007 (0.010) | -0.021*** (0.007) | 0.016 (0.013) | 0.013 (0.023) | -0.009 (0.011) |
| Observations | 9,323,109 | 14,055,929 | 14,055,929 | 15,365,318 | 15,889,409 |
| R-squared | 0.198 | 0.110 | 0.083 | 0.282 | 0.235 |
| F-stat | 23.83 | 24.82 | 24.82 | 24.68 | 24.66 |
| Linked sample | | | | | |
| Num. Blacks | -0.035*** (0.010) | -0.014** (0.006) | -0.016 (0.012) | 0.053 (0.034) | 0.063** (0.031) |
| Observations | 88,892 | 261,867 | 261,867 | 18,298 | 23,295 |
| R-squared | 0.346 | 0.040 | 0.040 | 0.827 | 0.845 |
| F-stat | 32.75 | 28.56 | 28.56 | 56.56 | 46.89 |

Notes: The table presents results for men living in the 108 non-southern MSAs for which the instrument could be constructed in Census years 1910, 1920, and 1930. The sample consists of foreign-born men. In panel A, it is further restricted to married individuals in column 1, and to individuals aged 15-65 in columns 2 and 3. In panel B, it is restricted to men not married in the previous decade (column 1), unable to speak English (column 4) or illiterate (column 5) in the previous decade. Individual controls include fixed effects for age and years in the US (Panel A) or birth year (Panel B) and origin region. All regressions control for MSA and year fixed effects and for total MSA population. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

could read and write. On the other, the Immigration Acts of the 1920s disproportionately favored immigrant groups that were more skilled and whose linguistic distance from English was lower (i.e. those from Western and Northern Europe). It follows that the “margin of adjustment” for these outcomes was certainly larger for immigrants arrived in the US before 1915, i.e. those that we observe in the linked sample, than for more recent ones.

D.2 Residential Segregation

Given the close-link between social identity and geographic segregation (Echenique and Fryer, Jr., 2007; Card, Mas and Rothstein, 2008; Gentzkow and Shapiro, 2011; Enos, 2017; Logan and Parman, 2017), we investigate whether there is a link between the arrival of African Americans as a result of the Great Migration and patterns of residential segregation. Building on a new method developed by Logan and Parman (2017) we use information on the relative locations of household heads to calculate ethnic segregation. This method utilizes the way in which census enumerators traversed around geographies to induct the identity of a given household’s neighbors. In particular, it is based on the relative ordering on a census enumeration form as a proxy for the geographic location of one’s neighbors. For example, if the household head enumerated right before a given household head is an immigrant, then this measure codes that given household head as being neighbors with an immigrant. From this, one can calculate a neighbor-based index of segregation along any binary dimension. Using the complete count U.S. Censuses from 1910-1930 we construct a measure of segregation on the basis of the country of birth of (first-generation) European immigrants. Specifically, we compute

$$Segregation = \frac{E(\bar{x}_c) - \bar{x}_c}{E(\bar{x}_c) - E(x_c)}$$

where x_c is the number of households nationality n (as defined by the household head) with a next-door neighbor of a different nationality. $E(\bar{x}_c)$ is the expected number of households with neighbors of different nationality under random assignment of neighbors, and $E(x_c)$ is the respective number under complete segregation.²⁷ We calculate four different versions that adjust for the presence of foreign-born non co-nationals on either side of a given household and for the presence of foreign-born non co-nationals on both sides, in addition to adjusting the measure to look at neighbors on the same street among the households for which we have data on street names. We denote these outcomes as *Segregation(P – One)* and *Segregation(P – Both)* when not adjusting for streets and using information on one or both sides respectively, in addition to *Segregation(S – One)* and *Segregation(S – Both)* when adjusting for streets.²⁸ We compute these measures at the MSA-level for the sixteen European countries that had more than 10,000 migrants living in the US in 1900 and stack the data, so that the level of

²⁷For more details on the construction of the measure and its comparison with traditional indices of segregation we refer readers to Logan and Parman (2017).

²⁸The correlation among all of these measures is generally quite high.

Table D.2. Segregation

| Dep. Variable | (1) | (2) | (3) | (4) |
|-----------------------|-------------------|-------------------|-------------------|-------------------|
| | Segregation index | | | |
| | <i>P – One</i> | <i>P – Both</i> | <i>S – One</i> | <i>S – Both</i> |
| Panel A: OLS | | | | |
| Num. blacks | -0.012 (0.014) | -0.010 (0.012) | -0.010 (0.011) | -0.015 (0.014) |
| Observations | 4,854 | 4,839 | 4,823 | 4,781 |
| R-squared | 0.541 | 0.488 | 0.522 | 0.494 |
| Panel A: 2SLS | | | | |
| Num. blacks | -0.026 (0.019) | -0.015 (0.016) | -0.022 (0.018) | -0.028 (0.019) |
| Observations | 4,854 | 4,839 | 4,823 | 4,781 |
| R-squared | 0.541 | 0.488 | 0.522 | 0.493 |
| F-stat | 18.71 | 18.74 | 18.73 | 18.83 |
| MSA-nationality cells | 1,637 | 1,635 | 1,632 | 1,624 |

Notes: 2SLS estimates reported. The analysis is restricted to 108 non-southern MSAs for which the IV could be constructed, and to years 1910 to 1930. All regressions control for MSA by nationality, MSA by region and year by region fixed effects and MSA total population and are weighted by population in 1900. Standard errors clustered at the MSA level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

observation in our final dataset is an MSA-nationality-year cell.²⁹ Table I presents summary statistics for all social and economic outcome variables, as well as for the segregation measures.

Table D.2 reports OLS and 2SLS results on immigrant segregation. The effect of black inflows is consistently negative across all four segregation measures, but not statistically significant. IV estimates are larger in magnitude than OLS ones, indicating that blacks may have sorted into initially more segregated areas. Substantively, these effects are modest relative to the variation across MSAs: a one standard deviation increase in black population across MSAs leads to about a one-tenth standard deviation decrease in segregation. Given the limited over time variation in segregation, however, it is important to also note substantive significance with regard to the variation *within* MSAs. Within MSAs, a one-standard deviation increase in black population leads to roughly a fifth of a standard deviation decrease in segregation. These effects are in line with previous results and provide suggestive evidence that the first Great Migration led to immigrants' residential assimilation with natives.

²⁹In choosing these nationalities, we follow Abramitzky, Boustan and Eriksson (2014). The countries in the dataset are Denmark, Norway, Sweden, England, Wales, Ireland, Belgium, France, the Netherlands, Switzerland, Italy, Portugal, Austria-Hungary, Germany, Poland and Russia.

E Additional model results

E.1 Proofs

Proof of Proposition 1. If $h = 0$ then $\underline{e}(h = 0) = 0$ and $c(0) = 0$ which is less than B , so immigrants of distance $h = 0$ will provide sufficient effort to reap the assimilation benefit.

If $h = H^{max}$ we assumed that $c(e(H^{max})) > B$, so immigrants of distance $h = H^{max}$ will provide no effort and will not become assimilated.

From equation 4, the implicit function theorem gives that $e_h = -\frac{W_h}{W_e} > 0$, so that the minimum effort necessary for assimilation, as well as its associated costs, are increasing in an out-group member's distance from the native white group.

Proof of Proposition 2. An increase in H^{max} raises \bar{h} , or the average distance of the out-group from the white natives. Writing 4 explicitly as

$$F = W(\lambda(h, \bar{h}), \underline{e}(h, \bar{h}))$$

we get immediately that

$$e_{\bar{h}} = -\frac{W_{\bar{h}}}{W_e} < 0$$

so that for all out-group members of a given distance h , necessary assimilation effort is reduced. This implies that after the inflow of African Americans we have $B > (\underline{e}(h_1^*))$ with h_1^* denoting the old level of the distance threshold that guaranteed assimilation. Since $e_h > 0$, the new threshold level h_2^* for which $B = (\underline{e}(h_2^*))$ must be greater than h_1^* . Thus, the arrival of a new out-group that is more distant to existing ones will necessarily increase the share of existing out-group members that are allowed membership into the native white in-group.

Proof of Proposition 3. If $h \leq h_1^*$ an immigrant provides the minimum necessary effort both before and after the increase. Since for any h \underline{e} goes down when H^{max} increases, immigrants who provide \underline{e} decrease their optimal level of provided effort. An immigrant with $h > h_1^*$ has provided zero effort under H_1^{max} . If $h \leq h_2^*$ immigrants now provide the (strictly positive) necessary effort level to ensure assimilation and if $h > h_2^*$ then efforts stay at zero.

E.2 Stochastic extension

So far, assimilation was assumed to be a deterministic process. We now generalize the model and include stochastic individual level characteristics that alter an out-group member's assimilation outlook. Let the disutility that an in-group member suffers from interaction with out-group member be given by

$$W(\lambda, e) + \mu$$

where $\mu \sim N(0, \sigma^2)$, representing unalterable characteristics of the individual. We assume that its value is unknown to the individual at the time when effort provision is decided. This may be due to the fact that out-group members cannot properly predict how the in-group will respond to their personal characteristics in specific interactions.

The out-group member then solves

$$\max_e Pr(F > W(\lambda, e) + \mu)B - c(e)$$

which can be written as

$$\max_e G(F - W(\lambda, e))B - c(e)$$

where $G()$ represents the *cdf* of μ . The first order condition for a relative extremum is then

$$g(F - W(\lambda, e))(-W_e)B = c'(e)$$

so that the marginal costs of effort are equalized to the *pdf* of μ , scaled by the benefits of assimilation and the marginal effect of effort on the psychological costs of natives. To make progress with the generalized model, we specify that this disutility W is linear in λ and e and that costs of effort are quadratic, in particular

$$\begin{aligned} W &= \omega_1 \lambda - \omega_2 e \\ c &= ce^2 \end{aligned}$$

with ω_1, ω_2 and $c > 0$. Finally we will assume that $\lambda_{h, \bar{h}} = 0$, implying that a change in \bar{h} has the same immediate effect on all out-group members³⁰. Optimal effort is then implicitly defined by

$$g(F - W(\lambda, e^*))\omega_2 B = 2ce^*$$

We then derive the effect of a marginal increase in \bar{h} on e^* , given by

$$\frac{\partial e^*}{\partial \bar{h}} = \frac{g'(\cdot)B(-\lambda_{\bar{h}})\omega_1\omega_2}{2c - g'(\cdot)B\omega_2^2} \quad (8)$$

which is positive as long as $g'(\cdot)$ is positive. Thus individuals respond to a larger \bar{h} with an increase in their assimilation efforts, as long as they had an initial assimilation probability of less than 50%, and reduce efforts otherwise.

³⁰We also assume that σ is sufficiently large so that $g'(-\sigma)\omega_2^2 B < 2c$ and the marginal cost curves thus crosses the scaled *pdf* only once

Equation (8) also shows that the response in effort has an inverted U-shape. The effect of \bar{h} on e^* is increasing in $g'()$ and thus maximal for individuals with \tilde{h} such that $F - W(\lambda(\tilde{h}, \bar{h}), e^*(\tilde{h}, \bar{h})) = -\sigma$. Effects are monotonically decreasing as h moves away from \tilde{h} in either direction. A similar result holds for resulting rates of assimilation into the in-group. Assimilation rates of individuals with distance h are given by

$$A(h, \bar{h}) = Pr(F > W(\lambda, e^*) + \mu) = G(F - W(\lambda(h, \bar{h}), e^*(h, \bar{h})))$$

where individuals with a lower h always have higher equilibrium assimilation rates. The response of assimilation rates to an increase in \bar{h} is

$$\frac{dA(h, \bar{h})}{d\bar{h}} = g(F - W(\lambda, e^*)) \left[(-\omega_1 \lambda_{\bar{h}}) + \omega_2 \frac{\partial e^*}{\partial \bar{h}} \right] > 0$$

Notice that, from the previous discussion, the expression in brackets on the right hand side is unimodal with peak at \tilde{h} , while the density $g()$ is naturally unimodal and maximized at a value $\tilde{\tilde{h}}$, $\tilde{\tilde{h}} < \tilde{h}$ such that $F = W(\lambda(\tilde{\tilde{h}}, \bar{h}), e^*(\tilde{\tilde{h}}, \bar{h}))$. The resulting response in A is thus also unimodal with a peak at h_A , with $\tilde{h} > h_A > \tilde{\tilde{h}}$. In other words, the response in assimilation rates to an increase in \bar{h} is also of inverted U-shape, but unlike assimilation efforts, the largest responses will be among individuals with relatively low distance h .

We summarize this finding in the following proposition.

Proposition 4. *The responses of optimal effort and of assimilation rates to an increase in H_{max} are both unimodal with respect to distance h . The peak of assimilation rates occurs at a lower level of h than the peak of optimal effort.*

We illustrate this result with a specific example, considering the functional form $W = \omega_1 h - \omega_3 \bar{h} - \omega_2 e$ and parameter values $\omega_1 = 3$, $\omega_2 = \omega_3 = c = \sigma = 1$, $B = F = 0.5$ and an increase of H_{max} from 1 to 1.5. Figure (E.1) plots optimal effort (left panel) and assimilation rates (right panel) as a function of h . In the upper row, the blue line indicates initial values ($H_{max} = 1$) while the orange line depicts values after the increase in H_{max} . The lower row plots the difference between the two. Both effort and assimilation rates have an inverted U-shape, but responses are larger for individuals with high h for effort and vice versa for assimilation.

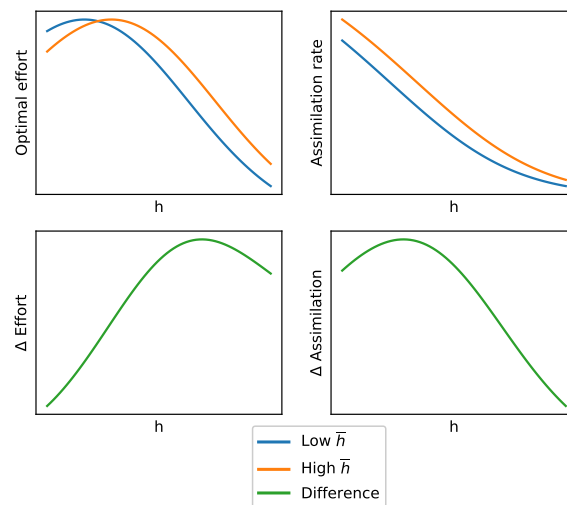


Figure E.1. Response of optimal effort (left panel) and assimilation rates (right panel) to an inflow of a distant out-group (increase in \bar{h}), as a function of individual distance h . The response of both measures has an inverted U-shape, yet effects on assimilation rates peak at much lower values of distance h .