DISCUSSION PAPER SERIES

DP14396
(v. 3)

FROM IMMIGRANTS TO AMERICANS: RACE AND ASSIMILATION DURING THE GREAT MIGRATION

Vasiliki Fouka, Soumyajit Mazumder and Marco Tabellini

ECONOMIC HISTORY

LABOUR ECONOMICS

PUBLIC ECONOMICS

ISSN 0265-8003

FROM IMMIGRANTS TO AMERICANS: RACE AND ASSIMILATION DURING THE GREAT MIGRATION

Vasiliki Fouka, Soumyajit Mazumder and Marco Tabellini

Discussion Paper DP14396
First Published 09 February 2020
This Revision 27 October 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

FROM IMMIGRANTS TO AMERICANS: RACE AND ASSIMILATION DURING THE GREAT MIGRATION

Abstract

How does the appearance of a new minority group affect the social acceptance and outcomes of existing minorities? We study this question in the context of the First Great Migration. Between 1915 and 1930, 1.5 million African Americans moved from the US South to Northern urban centers, which were home to millions of European immigrants arrived in previous decades. We formalize and empirically test the hypothesis that Black inflows changed perceptions of outgroup distance among native-born whites, reducing the barriers to the social integration of European immigrants. Predicting Black in-migration with a version of the shift-share instrument, we find that immigrants living in areas that received more Black migrants experienced higher assimilation along a range of outcomes, such as naturalization rates and intermarriages with native-born spouses. Evidence from the historical press and patterns of heterogeneity across immigrant nationalities provide additional support to the role of shifting perceptions of the white majority.

JEL Classification: J11, J15, N32

Keywords: Immigration, Assimilation, Great Migration, race, group identity

Vasiliki Fouka - vfouka@stanford.edu
Stanford University, Department of Political Science

Soumyajit Mazumder - smazumder@g.harvard.edu
Harvard University

Marco Tabellini - mtabellini@hbs.edu
Harvard Business School and CEPR

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

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

October 2020

Abstract

How does the appearance of a new minority group affect the social acceptance and outcomes of existing minorities? We study this question in the context of the First Great Migration. Between 1915 and 1930, 1.5 million African Americans moved from the US South to Northern urban centers, which were home to millions of European immigrants arrived in previous decades. We formalize and empirically test the hypothesis that Black inflows changed perceptions of outgroup distance among native-born whites, reducing the barriers to the social integration of European immigrants. Predicting Black in-migration with a version of the shift-share instrument, we find that immigrants living in areas that received more Black migrants experienced higher assimilation along a range of outcomes, such as naturalization rates and intermarriages with native-born spouses. Evidence from the historical press and patterns of heterogeneity across immigrant nationalities provide additional support to the role of shifting perceptions of the white majority.

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 Costas Arkolakis, Thomas Chaney, Nicola Gennaioli, Melanie Krause, David Laitin, Petra Moser, Salma Mousa, Agustina Paglayan, Giovanni Peri, Imran Rasul, Ken Scheve, Alain Schlaepfer, Bryan Stuart, Guido Tabellini, Hans-Joachim Voth, Katia Zhuravskaya, 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, the ENS de Lyon Workshop on Political Economy of Migration, the Washington PECO, the San Diego ASSA Meetings as well as seminar participants at LSE, UCL, Durham, Yale, Harvard, MIT, PSE, Stanford, Johns Hopkins SAIS and NYU for helpful comments and suggestions. Silvia Farina, Valentin Figueroa, Ludovica Mosillo, Arjun Shah, and Monia Tomasella 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 and CEPR. Email: mtabellini@hbs.edu

1 Introduction

In recent decades, immigration and rising diversity have fundamentally reshaped societies in the Western world. Economists and political scientists have shown increasing interest in the effects of immigration on social outcomes, including trust (Putnam, 2007), prejudice towards immigrants (Hainmueller and Hopkins, 2014), and political backlash (Halla et al., 2017; Dustmann et al., 2019).[1] Yet, scholars have paid less attention to how the arrival of new social groups affects earlier generations of migrants in already diverse societies. Do new groups facilitate the incorporation of existing minorities by re-directing prejudice away from the latter, or do they hinder it by fueling native-born backlash against all minorities? More broadly, how does the appearance of a new minority group affect the majority's attitudes towards other minorities, and shape social boundaries across groups?

In this paper, we build on insights from social psychology and advance a hypothesis on how the appearance of a new social group changes the social categorization of extant groups and, as a consequence, their socioeconomic outcomes. Specifically, we propose that social categorization is based on the *perceived distance* – cultural or social – across groups, and that such distance is context-dependent. With the appearance of a new minority group that is relatively distant from the majority, existing outgroups appear closer, and are more likely to be categorized as members of the ingroup.

We test this hypothesis in the context of US history. From 1915 to 1930, 1.5 million African Americans migrated from the US South to urban centers in the North and West, in a movement that was termed the First Great Migration. The same parts of the country had been major destinations for more than 30 million Europeans during the previous 50 years. White Americans of Anglo-Saxon origin made fine-grained ethnoracial distinctions across immigrant groups, which then formed the basis of prejudiced attitudes and nativist sentiment, particularly targeted against the most culturally and religiously different nationalities (Southern and Eastern Europeans). Historians have suggested that the arrival of Southern Blacks in the US North and West redrew racial boundaries around skin color rather than ethnicity, and facilitated the incorporation of formerly discriminated Europeans into "white" American society (Guterl, 2001).

We present a simple formalization of this process, drawing from the cognitive psychology literature on group categorization (Turner et al., 1987) and from related work in political science and economics (Shayo, 2009). We assume that ethnic and racial groups are ranked in terms of their (social, cultural) distance from the native-born white ingroup. Ingroup

[1] For the social costs and benefits of diversity more generally see Alesina and La Ferrara (2005) and Ray and Esteban (2017). Bazzi et al. (2019) and Charnysh (2019) show how diversity can promote the formation of superordinate social identities.

members engage in taste-based discrimination in order to avoid the psychological cost of interaction with the outgroup. Outgroup members can partly counter discrimination by exerting effort to assimilate. The cost of intergroup interactions for ingroup members is decreasing in assimilation effort and increasing in an outgroup member's perceived distance, which is a function of both actual distance and the social context – in this case captured by the distances of other groups. Context-dependence follows the *meta contrast* principle (Turner et al., 1987, 1994), which posits that categorization minimizes within group differences and maximizes across group differences. When an outgroup of high actual distance to the ingroup, relative to existing outgroups, appears, the perceived distance of existing outgroup members drops, leading to effort adjustment and recategorization of some outgroup members as members of the ingroup.

We test the predictions of this theory by using information on the universe of foreign-born US residents 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 of the United States during this period. We construct repeated cross-sections and a linked sample of immigrants who remained in the same MSA during the entire 1910-1930 period. We examine multiple measures of assimilation, focusing on naturalization and marriages with native-born spouses of native parentage. We combine this information with data from historical newspapers to capture changing attitudes of Northern white Anglo-Saxons towards immigrants and Black Americans.

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 an increase in naturalization rates and in the likelihood of intermarriage. These effects are quantitatively large. An inflow of Black migrants such as that experienced by Detroit (around 130,000 Blacks between 1910 and 1930) increased the share of naturalized immigrants by 4.3 percentage points, or 8.8% relative to the 1910 mean, and raised the probability of intermarriage between immigrants and natives by 1.7 percentage points, or 24% relative to the 1910 mean. Black migration also lowered residential segregation between immigrants and the native-born and induced foreign-born parents to choose less ethnically distinctive names for their children. 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 in turn favored (or hindered) assimilation. To address this concern, we construct a "shift-share" instrument (Card, 2001; Boustan, 2010) 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. The shift-share instrument does not simply assign more Blacks to areas with larger 1900 Black population. Rather, it combines two sources of variation: variation in the "mix" of Blacks born in different Southern states and living in different

Northern MSAs in 1900; and time-series variation in Black emigration rates from different Southern states for each decade between 1910 and 1930. Thus, conditional on its 1900 share of Black residents, a Northern MSA is predicted to receive more Black migrants in a given decade only if its 1900 Black settlers originated from Southern states that experienced higher Black outflow in that decade. The identifying assumption behind the instrument is that, conditional on MSA and region by decade fixed effects, immigrant assimilation after 1910 should not be simultaneously correlated with the 1900 composition of African Americans' enclaves in Northern MSAs *and* with migration patterns across Southern states after 1900.

One threat to the validity of the instrument is that the characteristics of MSAs that attracted a specific mix of early Black settlers (in terms of the composition of their Southern states of origin) might be correlated with time-variant confounders affecting both Black migration patterns and immigrant assimilation (Goldsmith-Pinkham et al., 2020). We address this concern by testing for pre-trends, interacting decade dummies with 1900 MSA characteristics, and controlling for both changes in economic activity predicted by baseline industry composition and the instrumented number of European immigrants from different source countries. We also show that enclaves of 1900 Black settlers from the Southern sending states that account for the largest share of Black migrants are not correlated with key MSA characteristics, such as the share of immigrants or manufacturing employment (Goldsmith-Pinkham et al., 2020). Finally, we verify that the instrument is not correlated with either local exposure to the immigration quotas (Ager and Hansen, 2017; Collins, 1997) or direct effects of WWI (Fouka, 2019; Ferrara and Fishback, 2020).

The identifying assumption may also be violated if shocks to Northern MSAs both influenced the local economic and political landscape and simultaneously caused out-migration from specific Southern states that already had large enclaves in those MSAs before 1900 (Borusyak et al., 2018). To address this potential concern we replicate our results using a version of the instrument that exploits only variation in local push factors across Southern counties to predict Black outflows from the US South (Boustan, 2010; Derenoncourt, 2019).

The second part of the paper explores the channels through which the Great Migration affected immigrant assimilation. Consistent with changes in perceptions of native whites, MSAs that received more Blacks exhibited larger reductions in concerns about immigration, as measured by the frequency of terms expressing anti-immigrant sentiment in historical newspapers. Black inflows also lowered 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 also they became less likely to be perceived as Catholics. Since religious cleavages between Protestants and non-Protestants were highly salient at the time (Higham, 1998), these results suggest that the Great Migration reduced the importance of features such as religion, which differentiated immigrants from native-born whites, and in turn lowered anti-immigrant sentiment. Simultaneously, negative stereotyping of African Americans increased. Black migrants were more frequently referred to in the press as "violent" or "rapist", in line with high perceived distance of – and associated prejudice against – this group from the white majority.

We test two additional predictions of the theoretical framework. The first one is that assimilation exhibits an inverted U shape. According to our model, groups most affected by Black arrivals were those of intermediate distance from native whites – sufficiently distant to be excluded from the ingroup before the inflow of Blacks, but still close enough to benefit from the arrival of the new outgroup. Consistent with this prediction, the interaction of continuous measures of distance (linguistic and genetic) with Black population produces an inverted U-shaped pattern. Additionally, we find no effect of Black migration on non-European immigrant groups, such as the Mexicans and the Chinese, that remained too distant to be recategorized as white, even after the Great Migration.

The second prediction of the model is that assimilation effort and successful assimilation should display the largest increase at different points of the distribution of immigrant distance. With native whites becoming more inclusive, groups of relatively low distance could experience immediate increases in assimilation with no need to provide additional effort. On the other hand, more distant groups would have incentives to increase their assimilation efforts, even though the outcome of those efforts – though more likely than before to be successful – would still remain uncertain. While we lack pure measures of immigrant effort, we show that naturalization rates – at the time an immigrant action that was little impeded by host society's barriers – were highest for relatively distant, "New Source" immigrants, such as Southern and Eastern Europeans. Instead, intermarriage rates – a measure more heavily affected by the preferences of the host society – peaked for "Old Source" immigrant groups, such as Northern and Western Europeans.

Black migrants potentially affected immigrant assimilation also through economic channels. In particular, we find evidence for a role of competition between immigrants and Black migrants. The effect of Black inflows on naturalization rates (but not intermarriage rates) was larger for Eastern and Southern Europeans, who were more similar to Blacks in terms of skills. It was also larger for immigrant groups that were more likely to be employed in manufacture and in unskilled occupations in 1900. These groups were arguably more exposed to labor market pressure induced by Black migration, and thus had higher incentives to signal Americanization and, possibly, to invest in skills that could promote social assimilation. On the contrary, we find little evidence for labor market complementarities between immigrants and Blacks that could have led to occupational upgrading of the former (Peri and Sparber, 2009; Foged and Peri, 2016). Such a mechanism would imply more economic assimilation among relatively more skilled immigrants. Yet, the effect of Black migration on economic outcomes, such as manufacturing employment and the gap in occupational income scores between immigrants and natives, displays little heterogeneity by skill.

Our paper builds on insights from social psychology, and especially self-categorization theory (Turner et al., 1987; McGarty, 1999), which studies how individuals classify themselves and others in groups. A set of theoretical papers in economics microfound similar categorization principles (Fryer and Jackson, 2008) and use them to explain patterns in political economy (Bonomi et al., 2020). Relatedly, 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 outgroup. Our framework relies on the concept of perceived distance, which draws from Shayo (2009).

More broadly, our study contributes to a large literature on ingroup and outgroup 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. Most closely related, McConnell and Rasul (2020) examine how increased animosity towards Muslims spurred by 9/11 affected minorities in the US, as evidenced by decisions in the Federal Criminal Justice system. They find evidence of negative spillovers on Hispanics, but not Blacks. The authors exploit an exogenous shock to preferences, while we exploit an exogenous shock to the size of a distant (and thus more discriminated) group. Findings in the two papers are compatible. In our context, Black migration increases prejudice against African Americans in Northern MSAs. In theirs, consistent with our framework, groups of a similar distance from the ingroup as Muslims – as they show was the case for Hispanics since 1990 – are viewed with similar distaste, while groups sufficiently distant from the ingroup (Black Americans) are unaffected by the changed average distance of the outgroup.

The existing literature on immigrant assimilation is vast and has identified a number of determinants of integration and its speed.[2] To our knowledge, there has been no comprehensive quantitative study of the causal effect of race on immigrant outcomes, in particular through channels working on the white majority's perceptions. Another study that emphasizes the role of natives' preferences for immigrant assimilation is Bisin and Tura (2019), which shows theoretically that higher tolerance can slow down integration. Our framework and empirical results produce similar findings; immigrants of a low distance to the ingroup exhibit lower assimilation along dimensions like naturalization rates, as they can achieve acceptance with lower levels of assimilation effort. Yet, our theory is simpler and targeted towards understanding categorization, thus abstracting from immigrants' preferences for cultural maintenance and cultural transmission dynamics.

Finally, our paper is related to the large literature on the Great Migration, which has focused on "white flight", Black and white economic outcomes, city finances, crime, and intergenerational mobility (Boustan, 2010; Collins and Wanamaker, 2014; Boustan, 2016; Shertzer and Walsh, 2019; Tabellini, 2018; Stuart and Taylor, 2017; Derenoncourt, 2019).[3] Our study complements these works by extending the analysis 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

[2]These include immigrant group size (Shertzer, 2016; Eriksson, 2020), ethnic networks (Edin et al., 2003), as well as education and other government policies (Lleras-Muney and Shertzer, 2015; Mazumder, 2018; Bandiera et al., 2019; Fouka, 2020).

[3]See also Collins (2020) for a review of this literature.

period of reference. We show that, by inducing immigrants to assimilate, the Great Migration had effects beyond those on native-born whites, and that the assimilation of Europeans 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 discusses the historical background and presents the conceptual framework linking the Great Migration to immigrant assimilation. Section 3 describes our data and empirical strategy. Section 4 presents our main result, namely that Black inflows led to immigrant assimilation, and summarizes the robustness checks performed to address threats to identification. Section 5 provides evidence on changed perceptions of native whites, and tests additional implications of the theory. Section 6 concludes.

2 Historical background

2.1 The first Great Migration

Mass out-migration 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 skill – 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 (Collins, 1997).

Alongside pull factors in the North, a number of push factors in Southern states drove Black out-migration during this period. Natural disasters such as the 1927 Mississippi flood (Boustan et al., 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 et al., 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; Feigenbaum et al., 2020).

The combination of these forces led around 1.5 million African Americans to move from the South to the North of the US between 1915 and 1930 (Boustan, 2016), increasing the fraction of Blacks living in the North from 10% to more than 25% in the same period. The unprecedented inflow of African Americans and the induced change in the racial landscape of Northern cities triggered hostile reactions by white residents, who often engaged in coordinated activity to segregate Blacks (Massey and Denton, 1993). Boustan (2010) and Shertzer and Walsh (2019) show, respectively for the second and for the first wave of the Great Migration, that uncoordinated actions were equally important for the rise of the American ghetto. Often, whites reacted to Black inflows by leaving cities and neighborhoods (a phenomenon known as white flight).

2.2 The Great Migration and immigrant assimilation

Between 1850 and 1915, 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 European immigrants changed dramatically 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 native-born whites, and were 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 for and faced higher barriers to assimilation. Indeed, return migration prior to 1920 is estimated to have been 30% or higher (Bandiera et al., 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 regions.[4]

Borjas (1987) and, more recently, Abramitzky et al. (2014) have shown that immigrants did not experience substantial labor market assimilation, and their gap from native-born whites persisted well into the second generation. Abramitzky et al. (2020a) also show gradual, 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. Barriers to integration, such as prejudice and discrimination, can potentially explain why it took so long for immigrants to close the gap with the native-born, despite substantial efforts to assimilate. Discrimination and prejudice were most often directed toward new immigrant groups. Though the Irish, Italians, and Eastern Europeans were phenotypically white, their social status was in many respects that of an inferior race (Guglielmo, 2003). Discrimination against immigrants was also often reinforced or even encouraged by the US government (Hochschild and Powell, 2008).

An extensive historical literature suggests that the Great Migration catalyzed the assimilation of immigrants and contributed to their Americanization. One factor emphasized throughout this literature is the role of changing perceptions of native-born whites toward racial boundaries (Ignatiev, 1995; Guterl, 2001). In the first two decades of the 20th century, academic theories about race and eugenics emphasized fine grained racial distinctions among the various European groups. Northern and Western Europeans were placed higher in the racial hierarchy than "Alpines" and "Mediterraneans" (Spiro, 2008). Prominent eugenicists like Madison Grant – the author of the opus magnum of scientific racism, *The Passing of the*

[4]With 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).

Great Race – contributed to the design of the 1920s system of immigration restrictions that reflected this racial hierarchy.

The Great Migration shifted the focus, both of academics and of society at large, from ethnic differences to color as a racial group identifier. Lothrop Stoddard, another influential eugenicist and Klansman, 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). The salience of race reduced ethnic prejudice, and made it easier for immigrants to assimilate into US society.

2.3 A conceptual framework

We formalize these historical insights in a simple model. We present the formal framework in Appendix B, and provide a summary here. The framework draws from the cognitive psychology literature on group categorization (Turner et al., 1987) and from related work in economics and political science (Shayo, 2009, 2020). We start from the assumption that ethnic and racial groups are ranked in terms of their distance – social, cultural, or other – from native-born whites (the ingroup). Ingroup members in turn engage in taste-based discrimination in order to avoid the psychological cost of interaction with outgroup members. Outgroup members can partly counter discrimination by exerting (costly) effort to assimilate. The cost of intergroup interactions for ingroup members is decreasing in assimilation effort and increasing in an outgroup member's *perceived distance*, which is context dependent. The context here is the average distance of all other members of the outgroup.

The appearance of an outgroup of higher actual distance to natives from that of existing outgroups reduces the perceived distance of the latter. This, in turn, induces effort adjustment and recategorization of some outgroup members as members of the ingroup.[5] The process of context-dependent categorization, known as the *meta contrast principle*, is a central tenet of the self categorization theory in social psychology (Turner et al., 1987, 1994), and is documented in experimental studies (Tajfel and Wilkes, 1963).[6]

The role of context-dependent perceived distance is also consistent with the historical narrative on the gradual incorporation of European immigrants into the white Anglo-Saxon majority. Studying this process, Jacobson (1999) writes 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

[5]Similar predictions on reclassification are delivered by the literature on stereotypes (Bordalo et al., 2016), which relies on the representativeness heuristic (Kahneman and Tversky, 1974). Stereotypes about groups depend on the reference group. Irish immigrants may be thought of primarily as Catholics when compared to native-born 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.

[6]Bonomi et al. (2020) use the meta contrast principle in the context of political identities.

saw Celtic, Hebrew, Anglo-Saxon, or Mediterranean physiognomies where today we see only subtly varying shades of a mostly undifferentiated whiteness."

Mapping the model to this empirical context, we derive the following testable predictions:

1. Black inflows increased the outgroup's average distance from native whites, in turn reducing the perceived distance of European immigrants (Proposition 1).[7]

2. Anticipating the reduction in natives' discrimination, immigrants exerted more effort to assimilate. Thus, on average, Black inflows increased both assimilation effort and assimilation success among European immigrants (Proposition 2). This is the central empirical test of our paper.

3. The relationship between distance and the response of assimilation effort (and, consequently, successful assimilation) to Black inflows has an inverted-U shape (Proposition 3). Intuitively, immigrant groups of intermediate distance to native whites, such as Eastern and Southern Europeans, were discriminated prior to the Great Migration, but were seen more favorably after Black arrivals. Those groups had the strongest incentives to exert more assimilation effort, since that effort became more likely to pay off. Instead, nationalities that were culturally closer to native whites, such as Northern and Western Europeans, could be integrated with less effort than before. At the same time, the recategorization effect of Black inflows should not have affected non-white immigrant groups, who were sufficiently distant from native whites, and who would not have been accepted as part of native society despite any increase in their assimilation efforts.

4. The effect of Black inflows on assimilation effort was highest for relatively distant immigrants, but the effect on successful assimilation was highest for relatively less distant ones (Proposition 4). Intuitively, in the presence of idiosyncratic shocks, assimilation effort did not always translate into successful assimilation. Hence, some immigrants of relatively high distance would have increased their assimilation efforts, without necessarily being accepted by the native ingroup. For those immigrants, efforts increased more than assimilation. Instead, immigrant groups of relatively low distance could be accepted without much effort adjustment. Those groups experienced larger increases in assimilation than in assimilation effort. The peak of the effort distribution should thus be on the right of the peak of the distribution of successful assimilation.

Section 4 tests the central prediction of this framework, namely the positive effect of Black migration on immigrant assimilation. Section 5 provides evidence consistent with the model's additional empirical implications.

[7] There is little doubt that African Americans stood at the bottom of the social hierarchy in early 20th century US and were thus more distant to native whites than European immigrants (Myrdal, 1944).

3 Data and empirical strategy

3.1 Data

To examine the effects of the Great Migration on immigrant assimilation, we use data from the full count of the US Census for the period 1900 to 1930 (Ruggles et al., 2015). We restrict our analysis to the 108 MSAs outside the US South with a positive number of southern-born Black residents in 1900 – a requirement imposed by the construction of the instrument, as explained in detail in Section 3.3. The complete list of MSAs in our sample is reported in Table A.1 and displayed in Figure 2. 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 influenced the residential decision of native-born whites, often triggering white flight. During the First Great Migration such white flight was mostly confined within cities, and took place across neighborhoods (Shertzer and Walsh, 2019).[8] Since cities expanded substantially over this period, focusing on MSAs allows us to consider the entire city area, without worrying about changes in boundaries over time possibly triggered by the Great Migration.

We are interested in testing the effects of the Great Migration on immigrant assimilation through a social mechanism working via changed perceptions of the native white majority. Our framework further introduces a nuanced distinction between assimilation effort exerted by immigrants and actual assimilation. The latter is an equilibrium outcome, which depends on the actions of both immigrants and native-born whites. Empirically, we lack measures that unambiguously isolate immigrant effort, unaffected by majority actions. To map the model to the empirics, we consider two outcomes that reflect native acceptance to different degrees, and that can thus help us, albeit imperfectly, distinguish between effort and assimilation.

The first one is intermarriage between a foreign born individual and a native-born spouse of native-born parentage. Sociologists have termed intermarriage "the final stage of assimilation"(Gordon, 1964), which arguably captures both immigrants' desire to fit in and natives' willingness to accept them. We interpret this as a proxy for successful assimilation. Our second main outcome of interest 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 file a Declaration of Intent (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. The petitioning process, which entailed fulfilling the residency requirement and having access to two witnesses, was relatively straightforward. This is reflected in the high rates of naturalization between 1910 and 1930

[8] This differs from the Second Great Migration, when whites left central cities for the suburbs (Boustan, 2010).

(over 50%).[9] Rejection rates of petitions by the courts were very low in practice.[10] Even though naturalization rates might partly reflect native whites' actions, they do so to a lesser extent than intermarriage, especially in the early 20th century. As such, they can be viewed as a proxy for the effort of the foreign-born to "fit in" the American society.

Besides intermarriage and naturalization rates, we also examine a host of other outcomes that reflect social assimilation to various extents. First, we consider employment in manufacturing and in unskilled occupations – two sectors where immigrants were over-represented – to measure immigrants' economic advancement. We also examine the ability to speak English. In early censuses, this was assessed by the enumerator with a binary yes-no answer, and thus masks important heterogeneity in English proficiency. As a result, rates of English proficiency were as high as 90% among the foreign-born during the 1910-1930 period. Nonetheless, English knowledge is an important indicator of assimilation. To examine whether assimilation affected immigrants' decisions regarding their offspring, we measure the ethnic distinctiveness of names given to immigrant children, using an empirical index developed by Fryer and Levitt (2004) and used by Abramitzky et al. (2014) and Fouka (2019, 2020). Finally, we use the procedure developed by Logan and Parman (2017) to measure changes in residential segregation along ethnic lines. A detailed description of the construction of these outcomes can be found in Appendix C.1.[11]

Table 1 presents summary statistics for MSA-level characteristics (e.g. total, Black, and immigrant population), and for the outcome variables aggregated at the MSA level. Table A.2 presents individual-level summary statistics.[12] Only 7.1% of immigrant men were married with a native-born white of native-born parentage. At the same time, almost 50% of immigrants in our sample were naturalized. The difference between intermarriage and naturalization rates is consistent with the conjecture that immigrants displayed a high desire to

[9] Bureaucratic barriers were also significantly less important at the time than today. Court clerks would often fill out forms and otherwise assist foreign-born petitioners. In a 1921 congressional hearing, MR J.C.F. Gordon, chief naturalization examiner in the third district of Philadelphia, states: "...in my office...We give the alien all the assistance possible, filling out his papers for him and properly advising him... ."

[10] In a sample of approximately 3,300 naturalization petitions filed in New York City in 1930, Biavaschi et al. (2017) find that only 2.6% were rejected. In a sample of 1,464 petitions from Pennsylvania and Illinois, Fouka (2019) does not find differential rejection rates across nationalities. Even during WWI, when German immigrants were considered enemy aliens, their petitions were processed with delay by the court, but not entirely rejected.

[11] These outcomes, while important, are less easily mapped to the distinction between effort and successful assimilation discussed above. Some, such as first names given to children, may be a better reflection of the downstream consequences of successful social assimilation.

[12] For individual level outcomes, we focus on men only. In the case of intermarriage we avoid double-counting. Until 1922, only men could apply for citizenship, and immigrant women were naturalized through their husband, or through marriage with a citizen. For economic outcomes, we follow the convention in the literature to focus on men (Abramitzky et al., 2019, 2020a; Ager and Hansen, 2017; Tabellini, 2020), since women constituted less than 20% of the labor force until WWII (Goldin, 1991). Focusing on men for all outcomes also allows for comparability of results between repeated cross-sections and linked panel. Consistent with the literature, we link men, since women changed their maiden names upon marriage and were harder to track in successive census decades.

assimilate, which may have been frustrated by barriers to assimilation erected by native-born whites.

3.2 Empirical strategy

3.2.1 Difference-in-differences

Repeated cross-section. Our 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 time-varying shocks common to all MSAs within a census region. Stacking the data for the three decades between 1910 and 1930, we estimate the following equation:

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

where Y_{int} is the outcome of foreign-born individual i living in non-southern MSA n in census decade t, and B_{nt} is the number of African Americans living in MSA n in decade t. We always include MSA and decade fixed effects (α_n and t_t), and in our preferred specification, we also control for interactions between decade 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 $\mathbf{X}_{int}$.[13] Controlling for MSA and region by decade fixed effects implies that β_1 is estimated from changes in the number of African Americans within the same MSA over time, as compared to other MSAs in the same region in a given decade. Finally, following Boustan (2010), since growing areas might have attracted both African Americans and European immigrants, we control for total MSA population, Pop_{nt}. Standard errors are clustered at the MSA level.

Linked panel dataset. 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 white flight (Boustan, 2010; Shertzer and Walsh, 2019). Black inflows could have similarly led to 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.

[13] When defining regions, we follow the census division classification. We classify immigrants into eleven countries or country groupings, corresponding to broad ethnoracial classifications of the early 20th century. 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, the Russian Empire (Russia, Estonia, Latvia and Lithuania), Mexico, China and Canada. We exclude immigrants from countries that do not fall in one of these origin groups.

In addition to dealing with possible compositional effects, comparing the same immigrant over time is also desirable when considering outcomes such as marriage or naturalization, which are "absorbing states". Once an individual obtains citizenship, he does not go back to non-citizen status. A similar argument holds for intermarriage, with divorce rates before 1930 being lower than 1%. A panel dataset allows us to restrict attention to those immigrants "at risk" of responding to the Great Migration, i.e. those who were not already naturalized or married at the time of Black arrivals.

Details on the construction of the linked dataset are provided in Appendix C.2. The last three columns of Table A.2 present summary statistics for this dataset.

3.3 Instrument for Black population

The northward movement of African Americans was largely dictated by economic conditions in Northern cities. Those same conditions might have also affected – positively or negatively – the assimilation of European immigrants. For instance, the outbreak of WWI pulled many Black Americans to the US North (Boustan, 2016), and at the same time triggered widespread anti-German sentiment (Fouka, 2019, 2020; Ferrara and Fishback, 2020). Moreover, the introduction of the literacy test in 1917 and, more importantly, of immigration quotas in 1921 and 1924 drastically reduced immigration flows to the US (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 low-skilled workforce created by 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 predict the location decision of Black migrants using a version of the "shift-share" instrument commonly adopted in the immigration literature (Card, 2001; Boustan, 2010). This instrument exploits two sources of variation. The first one is cross-sectional variation across Northern MSAs in 1900 settlements of African Americans born in Southern states. The second one is time-series variation in the number of Black migrants who left the South from each state after 1900. The predicted number of Black migrants received by each Northern MSA thus depends both on the 1900 "mix", in terms of their state of origin, of Southern born African Americans and on the heterogeneity in out-migration from each Southern state between 1910 and 1930.

Because data on internal migration do not exist before 1940, we estimate migration rates from each Southern state in each decade using the forward survival method (Gregory, 2005). We detail this procedure in Appendix C.3. We then predict the number of Blacks received by each Northern MSA in any given decade 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 decade t is given by

$$Z_{nt} = \sum_{s=1910}^{t} \sum_{j \in South} \alpha_{jn}^{1900} O_{js} \tag{2}$$

where α_{jn}^{1900} is the share of Blacks born in Southern state j and living in MSA n in 1900 (as a fraction of all Blacks born in state j and residing outside the South), and O_{js} is the number of African Americans born in state j who left the South between $s-1$ and s. Since we are interested in instrumenting a stock, i.e. the total number of Blacks in the MSA, B_{nt}, we recursively sum flows across decades.

This instrument is based on the empirical regularity that settlement patterns of Blacks were 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). 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. This is visually confirmed in Figure A.1, which plots the share of southern-born Blacks from selected states living in a number of Northern MSAs in 1900.

Consistent with the persistence of Black settlements, the instrument has high predictive power. Figure 3 plots the correlation between actual and predicted Black population, controlling for total MSA population, MSA fixed effects, and region by decade fixed effects. The coefficient implies that every predicted new Black arrival in the MSA is associated with 1.1 more actual Black residents.[14] Appendix Table C.2 presents the regression analog of Figure 3, and verifies that the instrument remains strong when estimating unweighted regressions and when including additional controls.

Several recent papers have discussed conditions for the validity of shift-share instruments (Borusyak et al., 2018; Goldsmith-Pinkham et al., 2020; Jaeger et al., 2018; Adao et al., 2019). One way to express the key identifying assumption in our context is the following. Conditional on MSA and region by decade fixed effects, third factors affecting the trajectories of immigrant assimilation after 1910 must not be simultaneously correlated with both the 1900 mix, in terms of Southern state of origin, of Black enclaves across Northern MSAs *and* emigration rates from different Southern states after 1910. In Section 4.2, we summarize a number of robustness checks providing evidence that this assumption is not violated in our context. All checks are presented in more detail in Appendix D.2.

[14] Our estimates are similar to those reported in Shertzer and Walsh (2019) and Tabellini (2018) for the same historical period for neighborhoods and cities, respectively. The fact that the estimated coefficient is close to 1, as theoretically expected, is reassuring, both for interpretation and for the predictive power of the instrument. The latter is also more formally confirmed by the values of the F-stat, which are well above conventional levels.

4 Results

4.1 Social assimilation

We begin the empirical analysis by presenting results for intermarriage between immigrants and native-born whites (of native parentage) in Panel A of Table 2. Restricting attention to married immigrant men, we first estimate equation (1) with OLS. Column 1 only includes MSA and decade fixed effects, and controls for individual level characteristics (age, origin region, and years in the US fixed effects). Column 2 additionally includes region by decade fixed effects. In both cases, the point estimate is positive, but close to zero and not statistically significant.

Columns 3 to 6 present 2SLS results. Column 3 replicates column 2 instrumenting the number of Blacks with the shift-share instrument introduced in Section 3.3. The coefficient becomes larger in magnitude and statistically significant at the 5% level. The downward bias in the OLS estimates indicates that Black migrants may have selected MSAs where the prospects for immigrant assimilation were not that bright. The point estimate implies that one standard deviation increase in the number of Blacks (approximately 45,000 individuals) increases intermarriage rates by 0.54 percentage points, or 7.5% relative to the 1910 mean. For a city like Chicago, which received almost 230,000 Black migrants 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 decade by region of origin fixed effects. 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. The magnitude of the coefficient is larger than in the cross-sectional dataset, but substantially smaller sample size leads to a loss in precision.[15]

Panel B of Table 2 turns to naturalization. Both OLS (columns 1-2) and 2SLS results from the repeated cross-sections (columns 3-5) and linked sample (column 6) indicate that Black inflows significantly increase immigrant naturalization rates. According to our most preferred specification, reported in column 3, one standard deviation increase in the Black population raises naturalization rates by approximately 1.5 percentage points, or by 3% relative to the 1910 mean. In column 6, as for intermarriage, we restrict attention to individuals who were not naturalized or not married in the previous decade, and thus were able to register a change in those outcomes in response to Black arrivals. As a result, the magnitude of the coefficient is once again larger in the linked sample.

[15] We confirm that this loss of statistical significance is due to small sample size in an unreported bootstrap exercise. We constrain the true effect size to be equal to the one displayed in column 6. In a set of 10,000 draws with replacement, with sample size equal to the one of column 6, only 13.9% of estimates are significant at the 10% level.

4.2 Summary of robustness checks

In this section, we summarize the robustness checks performed to increase confidence in the validity of our empirical strategy and rule out alternative explanations. They are described in more detail in Appendix D.

There are two main threats to our identification strategy, which relies on a Bartik-type instrument. The first concern, formalized in recent work by Goldsmith-Pinkham et al. (2020), is that the 1900 composition of Black enclaves in Northern MSAs (i.e. the "share" component of the shift-share instrument) were correlated with both post-1900 migration patterns from the South and time-varying factors affecting immigrant assimilation in those MSAs. To address this concern, we show that: *i*) the 1900 to 1910 change in European immigration is uncorrelated with predicted Black inflows between 1910 and 1930 (Figure D.1); *ii*) there are no pre-trends for our outcome variables (Table D.1); and, *iii*) results are robust to interacting decade dummies with a number of 1900 MSA characteristics, including the 1900 share of Blacks in the MSA, to separately controlling for a (predicted) measure of industrialization, and to controlling for baseline employment (Tables D.2 and D.3). Controlling for the interaction between the 1900 fraction of Blacks and decade 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 initial Black populations.

We also address the possibility that the composition of 1900 Black enclaves and emigration rates from the South may be spuriously correlated with immigration patterns of specific European groups. We show that our estimates do not change when separately controlling for the instrumented number of European immigrants from "new" and "old" sending countries (Table D.4). In addition, following Goldsmith-Pinkham et al. (2020), we compute the Rotemberg weights, and verify that those driving the bulk of the variation in the instrument are not correlated with key 1900 MSA characteristics (Tables D.5 and D.6). Finally, in Appendix D.1.2, we extensively address the possibility that the immigration quotas of the 1920s or the outbreak of WWI may be correlated with the instrument, due to a spurious relationship between the baseline composition of Blacks across MSAs and the presence of immigrant groups either more exposed to the quotas (e.g. Southern and Eastern Europeans; Abramitzky et al., 2019) or more directly affected by WWI (e.g. German immigrants; Fouka, 2019).

The second threat to the validity of the instrument, formally described in Borusyak et al. (2018), is that out-migration from each Southern state might not be independent of cross-MSA pull factors systematically related to 1900 settlers' state of origin. To deal with this possibility, in Table D.15, we replicate our results using a version of the instrument that only exploits variation in "push" factors across Southern counties to predict outmigration from each Southern state in each decade, as in Boustan (2010) and Derenoncourt (2019).[16] This

[16]Using this and our baseline instrument, we also conduct an over-identification test to increase the confidence in the validity of our estimates (Table D.17), as suggested in Goldsmith-Pinkham et al. (2020).

alternative instrument also deals with the possibility that our 2SLS estimates may conflate the long and short run effects of Black migration, due to serial correlation (Jaeger et al., 2018).[17] Lastly, we address the concern that correlated shocks might generate a spurious correlation between origin states and receiving MSAs, causing standard errors to be too small (Adao et al., 2019). In Appendix D.2, we perform a placebo check that weighs against this possibility (Figures D.4 and D.5).

Even if the shift-share instrument were valid, our results may still be influenced 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 part of this concern, but we perform additional checks. We document that Black inflows did not affect either the share of immigrants – constructed as fraction of different (e.g., total, white, or native) MSA populations – or their composition (Table D.18). We also show that Black in-migration was not associated with changes in immigrants' and natives' profiles in the repeated cross-section (Tables D.19, D.20 and D.22), and use movers in the linked sample to provide evidence against the possibility that the Great Migration triggered selective flight, with more racist immigrants or natives leaving MSAs (Tables D.21 and D.23).

Finally, we show that neither first stage nor 2SLS results change in direction or significance when excluding individual MSAs from the sample (Figures D.6 and D.7).

4.3 Effects on other immigrant outcomes

Table 3 examines a number of additional outcomes that capture different dimensions of immigrant assimilation. Column 1 shows that the share of immigrants employed in manufacturing dropped significantly in response to Black inflows. One standard deviation increase in the number of African Americans lowered the share of immigrants working in the manufacturing sector by 2.5 percentage points, or 10% relative to the 1910 mean. Column 2 confirms this by showing that Black inflows had a negative effect on the share of immigrants working in unskilled occupations. Besides indicating economic mobility, these results also reflect economic assimilation, with immigrants leaving the "immigrant intensive" sectors (see also Figure A.2).

We find no effect on immigrants' ability to speak English (column 3), though the estimated coefficient is positive and not far from conventional statistical significance in the linked sample results of Panel B (p-value=0.121). Baseline rates of English proficiency among the foreign-born during the period were very high, partly owing to the fact that census enumerators coded as English-speaking all immigrants with basic language knowledge. These effects are further attenuated in the repeated cross-section because, after 1920, new immigrants were significantly more likely to speak English. The literacy test of 1917 restricted access to the US to immigrants who could read and write, and the Immigration Acts of the 1920s

[17] To even more directly tackle the potential issue of serial correlation, in Table D.16, we implement the procedure described in Jaeger et al. (2018), instrumenting both actual and lagged Black population. Our results go through.

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 even smaller for immigrants who arrived in the US after 1915, and that are part of the repeated cross-section, but not the linked sample.

Column 4 turns to naming patterns – another proxy of social assimilation, which captures a choice of foreign-born parents for the identity of their offspring. Immigrants in MSAs with more Black migrants were significantly less likely to give their children names distinctive of their ethnic origin. The magnitude of the effect is substantial. 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 that from Stanislav to Morris or Max.

Finally, in columns 5 and 6, we estimate regressions at the MSA-nationality level, using as dependent variable an index of residential segregation.[18] We follow Logan and Parman (2017) in using the census schedules to identify households' next door neighbors, computing the segregation measure along ethnic, rather than racial lines. Ethnic segregation becomes significantly lower in MSAs with more Black migrants. That is, European immigrants living in MSAs receiving more African Americans became more likely to have non-immigrant neighbors. In line with results on intermarriage, this is additional evidence of increased acceptance of the foreign-born by native-born whites.

Taken together, these results confirm our findings in Section 4.1 that the Great Migration increased immigrant assimilation in a number of domains. In the next section, we provide evidence on the mechanisms behind these findings.

5 Mechanisms

Immigrant assimilation is consistent with Southern Black migrants changing the perceived distance of European immigrants from white natives, leading to the recategorization of the former as white Americans. Here, we provide evidence for three implications of such a mechanism. These are: *i*) Black inflows changed the perceptions of native whites and reduced the perceived distance of European immigrants from the native white majority; *ii*) the effect of Black inflows was highest for immigrant groups of intermediate distance from native whites; and, *iii*) assimilation effort responded most for relatively distant immigrants, while successful assimilation was highest for relatively less distant ones. We then discuss other plausible

[18]To approximate our individual-level regressions, we compute the segregation measures at the MSA-level for the 16 European countries that had more than 10,000 migrants living in the US in 1900 and stack the data, so that the level of observation in our final dataset is an MSA-nationality-year cell. This allows us to control for MSA by nationality and nationality by decade fixed effects. In choosing nationalities, we follow Abramitzky et al. (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 the Russian Empire.

pathways through which the Great Migration could have facilitated immigrant assimilation, such as economic and political channels.

5.1 Social channels

5.1.1 Changed attitudes of native whites

We use the historical press to provide evidence that Black inflows reduced the perceived distance of immigrants from native-born whites (Proposition 1). 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 71 of the 108 MSAs in our sample.[19] 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. 2SLS 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 effect on generic terms related to immigration ("immigration"; "immigrants"; "aliens"), suggesting that Black arrivals lowered the salience of the immigration issue (column 1). Next, we turn to terms that either reflect concerns over immigration or capture cultural and ethnic prejudice. The Great Migration lowered the relative frequency of the term "quotas" (column 2). Consistent with our previous findings, this result indicates that demand for immigration restrictions fell as European immigrants became increasingly perceived as less distant from native-born 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.

Mentions of both "Catholic" or "Catholic threat" (columns 4 and 5) and the Ku Klux Klan (column 6) also 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 (e.g. Higham, 1998). 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 to 6 as consistent with the idea that the Great Migration softened the widespread anti-immigration sentiment (in particular

[19]Table A.3 presents the list of the 71 MSAs for which newspaper data are available. Table A.4 compares MSAs with newspapers to the full sample. As expected, MSAs with newspapers are larger (though not significantly so), faster growing, and have more Black residents (though not as a share of total population), but are overall comparable to the full sample. Table A.5 (columns 2 and 4) replicates our baseline results restricting attention to the MSAs for which newspaper data are available. Reassuringly, they remain very similar to our findings for the full sample (columns 1 and 3).

towards Catholic immigrants) prevailing at the time.[20]

Alongside prejudice against immigrants, the inflow of African Americans also reduced stereotyping. We provide evidence for this by searching for the co-occurrence of words stereotypically associated with two of the largest European immigrant groups of the time, the Irish and the Italians. Figure 4 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 4, 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. In the right panel, we instead normalize it by the frequency of stereotypical words, to account for the possibility that terms like "crime" or "alcohol" may have become less widely discussed topics over time. Black inflows reduced the probability that the Irish and the Italians were mentioned jointly with the term "Catholic". This lends support to our theoretical mechanism, which posits that immigrants were reclassified as ingroup members, and were thus no longer associated with features that previously distinguished them from the native-born whites, such as religion. 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.

According to our theoretical framework, recategorization of white immigrants happened because Black Americans were perceived as more distant from the native majority relative to white Europeans, and were thus subject to more discrimination. Evidence from the press confirms that. Table 5 displays the frequency of joint mentions of the word "Black" (Panel A) or "Negro" (Panel B), with a number of words stereotypically associated with Black Americans in the early 20th century. We compile the list of negative terms referring to Blacks from a variety of sources. Terms such as "lazy" or "dirty", come from a 1933 Princeton study quantifying the stereotypes of college students in the US (Katz and Braly, 1993). The association between Black men and rape first became prevalent in the sensationalist press of the post-Bellum South, and has been emphasized by Woodward (2002) and Glaeser (2005). Black migrants to the North in particular were thought of as "unstable" and "inefficient" workers, as expressed through interviews of white foremen who considered Blacks for industrial jobs (Bodnar et al., 1979).

Table 5 shows that, as negative stereotyping for European immigrants became less prevalent, stereotypical associations for Blacks increased in MSAs with larger inflows of Black migrants. Consistent with a cognitive mechanism reliant on the meta-contrast principle, the perceived distance of immigrants fell, exactly because the distance of the new outgroup (African Americans) was high.

[20]The decline in KKK mentions reported in column 6 likely represents a lower bound for the effects of Black inflows on the reduction in native-born whites' prejudice towards European immigrants, since the arrival of African Americans may have increased KKK activity directed against Blacks.

5.1.2 Heterogeneity across immigrant groups

Our model predicts that the effect of the Great Migration on immigrant assimilation effort should be non-monotonic in distance (Proposition 3). Assimilation outcomes should have been most positively affected for groups "at the margin", who were excluded before, but who were given the opportunity to integrate once Black inflows changed native whites' perceptions on ingroup boundaries.

We proxy for the distance between immigrants and native-born whites 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 5 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. Using either measure of distance, we observe an inverted U-shaped pattern for naturalization rates.[21] Results are less clear in the case of intermarriage rates, though still present in the case of genetic distance.[22]

In Table A.7, we present results for our main outcomes for non-white immigrants (Mexicans and Chinese), who were viewed as even further away in distance from native-born whites. For these very distant groups there is no effect of Black inflows on either assimilation effort or actual assimilation.

A related, but more subtle, prediction of the model is that assimilation effort and success should peak at different points of the distribution of immigrant distance (Proposition 4). In response to changed attitudes of native whites, immigrants of relatively low distance should experience a large increase in assimilation, with no need to adjust their effort. Immigrants further away from the native whites, instead, may have had to increase their effort more. Given any degree of idiosyncratic uncertainty in the outcome of such effort, not all immigrants who exert effort would end up becoming accepted by the ingroup. Thus, for relatively distant immigrants, the increase in effort should be on average higher than the increase in assimilation.

Figure 6 illustrates a pattern consistent with this prediction. We report 2SLS coefficients for the effects of the Great Migration on the probability of naturalization (left panel) and intermarriage (right panel), under the assumption, justified in Section 3.1, that naturalization

[21] To deal with the potential correlation between genetic or linguistic distance and fixed national origin characteristics that may vary across MSAs or over time, in Table 6, we control for national origin by MSA and national origin by decade fixed effects (as in column 5 of Table 2). The table also reports p-values from a Wald test assessing the fit of the quadratic model against the linear alternative. The model fails to reject the null hypothesis that the quadratic interaction term is zero in three out of the four columns of Table 6.

[22] Genetic and linguistic distance may correlate with group size, but group size is not the driver of the observed U-shaped pattern. Controlling for the share of each origin group in 1900 interacted with year fixed effects leaves these results unchanged.

rates better capture immigrant effort than rates of intermarriage.[23] With the exception of the UK, immigrants from new source regions (the Russian Empire and Eastern and Southern Europe) exhibit the highest increase in naturalization rates. On the contrary, immigrants from old sending countries (Germany and Northern and Western Europe) experience smaller or negative changes. Social assimilation, as proxied by intermarriage rates, follows instead the opposite pattern. While coefficients are positive for most groups, with the notable exception of the Irish, acceptance by the native-born group increases the most for old source country nationals.

Taken together, the patterns of heterogeneity across immigrant groups are consistent with the theoretical model. The effects of Black migration peaked for racially ambiguous immigrants who became incorporated into the American society by virtue of their whiteness. More distant immigrants with a chance to assimilate in light of natives' changed attitudes showed higher increases in effort; instead, culturally proximate immigrants directly benefited from the change in racial boundaries and assimilated at higher rates despite lower effort provision.

5.2 Economic and other channels

Black migration to the US North brought major transformations – both social and economic. We provided evidence for changes due to a social mechanism, but the Great Migration may have fostered immigrant assimilation through more than one channels. In this section, we investigate mechanisms complementary to changes in racial perceptions that could have reinforced the effects estimated above.

We start by examining the role of economic competition between immigrants and African Americans. Between 1910 and 1930, Black migrants were disproportionately low-skilled and were absorbed in sectors that were until then immigrant-dominated. The ensuing competition between low-skilled workers and Blacks has been emphasized in the historical literature (Collins and Wanamaker, 2015; Boustan, 2016). Conflict between ethnic minorities and African Americans was common, and even predated the Great Migration (Rieder, 1987; McDevitt et al., 2002; Cho, 1993). 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 (Ignatiev, 1995).

Competition with Blacks might have induced immigrants to either invest in skill acquisition or actively try to differentiate themselves from their competitors, perhaps by signaling their Americanization in order to become more attractive to employers. This latter channel has been highlighted by the historical literature (Olzak and Shanahan, 2014). In his book *How the Irish became white*, Ignatiev (1995) documents how the Irish before the civil war facilitated their assimilation by emphasizing their differences from African Americans. Roediger (1999)

[23]Regression results underlying Figure 6 are reported in Table A.8.

shows how "immigrants in dirty and disease-ridden cities countered nativist assertions of racial difference with a determined focus on their own whiteness, on [Blacks], and on slavery" (Guterl, 2001).

To proxy for the degree of competition, in Table 7, we augment our baseline specification with interactions between Black inflows and the 1900 share of an immigrant group at the MSA-level that was employed, respectively, in manufacturing (columns 1 and 2) and in unskilled occupations (columns 3 and 4). 2SLS estimates indicate that the positive effect of Black inflows on naturalization rates seems to stem entirely from groups of immigrants employed in these sectors, who were disproportionately exposed to Black competition. This confirms the origin region-level results of Figure 6, and suggests that labor market exposure to Black migrants induced immigrants to naturalize at higher rates. The effect of employment sector on intermarriage rates on the other hand is muted.

Table 7 and Figure 6 suggest that some of our findings may be explained by labor market pressure, which led immigrants to invest in skill acquisition favoring assimilation. It is perhaps puzzling that groups experiencing the highest labor market pressure displayed large changes in naturalization rates, but not intermarriage patterns. One possibility is that all immigrants experienced the same, assimilation-inducing, labor market shock, but that assimilation was reflected in different outcomes, depending on the skills (and pre-existing integration) of different immigrant groups. For those already naturalized (disproportionately Western and Northern Europeans), assimilation may have manifested in increased intermarriage rates. For those with lower skill levels, naturalization constituted the first step in the assimilation process. Such economic forces would have reinforced the effects that Black in-migration had through the changes in native whites' perceptions of immigrants.

Next, we turn to the potential role of labor market complementarities between immigrants and Blacks. If immigrants exhibited some degree of complementarity with African Americans, Black arrivals may have not constituted direct competition, but may have instead led to immigrants' occupational upgrading (Peri and Sparber, 2009; Foged and Peri, 2016). Such economic advancement could then have fostered Europeans' social incorporation. Figure A.2 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 native-born whites and immigrants from old source countries. It is thus possible that the heterogeneous patterns observed for social assimilation, with Northern and Western Europeans exhibiting the highest increases in intermarriage, resulted from Blacks (positively) affecting immigrants' economic status.

We investigate the role of complementarities by examining the heterogeneity in economic outcomes in Figure 7.[24] In the left panel, we consider the employment share of immigrants in manufacturing. In the right panel, we construct the native-immigrant gap in log occupational

[24]Underlying regressions are shown in Table A.9. As for Figure 6, the results in Figure 7 are robust to controlling for the share of each immigrant origin group in 1900 interacted with year fixed effects.

scores.[25] 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. The reduction in the share of immigrants employed in manufacturing is rather uniform across groups, and 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 – a relatively skilled immigrant group. All in all, labor market complementarities seem less plausible than increases in the skill premium as an economic channel driving assimilation.[26]

We conclude by exploring two potential alternative mechanisms, and find little evidence for either of them. First, we consider the possibility that Black arrivals changed the electoral landscape and altered politicians' incentives, inducing them to increasingly cater to immigrant voters, through patronage and targeted redistribution. We provide evidence against this mechanism by showing that Black inflows did not increase immigrants' employment in the public sector (Table A.10) – either overall (column 1) or in specific occupations, such as policemen or firemen (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 benefiting from patronage jobs.

Second, Black arrivals might have 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. In Table A.11 we 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, 2015), are available between 1906 and 1917. For this reason results in Table A.11 include only 1910 and 1920. This result rules out the possibility that Black migration drove labor and social mobilization, which might have in turn favored the assimilation of European immigrants.[27]

[25]Occupational 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 et al., 2014).

[26]The lack of heterogeneity in economic outcomes may mask multiple labor market channels at work. Black inflows could have simultaneously induced upskilling of low-skilled workers and occupational upgrading due to complementarities with high-skilled ones. Yet, this would be hard to reconcile with the findings in Table 7. If this were the channel, one would have expected relatively more skilled groups in 1900 to display higher increases in intermarriage rates, which is not what we find.

[27]We also verify that the local presence of IWW unions in 1910 is uncorrelated with the 1910 to 1920 change in predicted Black population predicted (Figure A.3).

6 Conclusion

What are the implications of migration movements for social relations between existing groups in racially and ethnically diverse societies? We study this question in the context of US history. We explore the interaction of the Great Migration of African Americans and the mass migration of Europeans – two processes that critically contributed to the formation of the modern American racial and ethnic landscape.

We develop a simple conceptual framework that builds on insights from social psychology to formalize the hypothesis – consistent with historical accounts – that the arrival of 1.5 million African Americans to the US North during the First Great Migration changed perceptions of native whites toward European immigrants, making the latter seem "closer" to them than before. We predict Black in-migration using a version of the shift-share instrument, and show that the arrival of African Americans facilitated the assimilation of the foreign-born in Northern and Western MSAs. We consider different proxies for immigrant assimilation, focusing in particular on naturalization rates and intermarriage between immigrants and natives of native parentage.

Using measures of anti-immigrant sentiment from local historical newspapers, we provide evidence that Black in-migration reduced national stereotyping and lowered concerns about immigration among native-born whites, while simultaneously increasing negative stereotyping of Black Americans. Consistent with theoretical predictions, immigrant assimilation had an inverted U-shape; immigrant groups that benefited the most were those of intermediate cultural distance to native-born whites. These groups were considered racially distinct prior to the Great Migration, but were recategorized as "white" after the arrival of Black migrants. Though consistent with economic mechanisms, such as labor market pressure that induced skill upgrading among immigrants, our results cannot be entirely explained by these forces.

Our framework and results are consistent with mechanisms emphasizing salience of particular attributes for social identity and categorization (Shayo, 2009). An alternative version of the paper's central takeaway is that the appearance of African Americans increased the salience of skin color as ingroup identifier, simultaneously lowering the salience of ethnicity. While we do not explicitly model multidimensional attributes and their salience, this mechanism is implied in our unidimensional measure of distance.[28]

Do lessons from this historical episode apply to more recent contexts? While the First Great Migration and white-Black relations in the US have unique characteristics, evidence suggests that the social and cultural distance of new groups matters for the assimilation of existing minorities also in other contexts. Hainmueller and Hangartner (2013) find that discrimination against Italian immigrants in Switzerland – one of the European countries

[28]More specifically, we view the Great Migration as an exogenous change in the distance of the outgroup, rather than as an exogenous change in the salience of an attribute. Fouka et al. (2020) more explicitly show that, when categorization follows the meta-contrast principle, changes in the size of a distant group endogenously change the salience of attributes on the basis of which the group differs from the ingroup.

with the highest share of foreign-born residents – decreases simultaneously with an increase in discrimination against Turks and nationals of former Yugoslavia. Fouka et al. (2020) show that Mexican immigration to the US between 1970 and 2010 improves whites' attitudes towards Blacks, as the salience of immigrant status increases at the expense of race. These studies suggest that the growth of an out-group perceived as distant by the majority has the potential to change the majority's attitudes towards other out-groups in a society. Our framework thus has applicability that extends beyond the interaction of earlier and later immigrant arrivals, and can be used more broadly to study intergroup relations in diverse societies.

References

Abramitzky, Ran and Leah Platt Boustan, "Immigration in American Economic History," *Journal of Economic Literature*, 2017, *55* (4), 1311–1345.

— **, Leah Boustan, and Katherine Eriksson**, "Do Immigrants Assimilate More Slowly today than in the Past?," *American Economic Review: Insights*, 2020, *2* (1), 125–41.

— **,** — **, and Myera Rashid**, *Census Linking Project: Version 1.0 [dataset]*, https://censuslinkingproject.org, 2020.

— **, Leah Platt Boustan, and Katherine Eriksson**, "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration," *Journal of Political Economy*, 2014, *122* (3), 467–506.

— **, Philipp Ager, Leah Platt Boustan, Elior Cohen, and Casper W Hansen**, "The Effects of Immigration on the Economy: Lessons from the 1920s Border Closure," 2019. NBER Working Paper No. 26536.

Adao, Rodrigo, Michal Kolesár, and Eduardo Morales, "Shift-share Designs: Theory and Inference," *The Quarterly Journal of Economics*, 2019, *134* (4), 1949–2010.

Ager, Philipp and Casper Worm Hansen, "Closing Heaven's Door: Evidence from the 1920s U.S. Immigration Quota Acts," 2017. Working Paper.

Alesina, Alberto and Eliana La Ferrara, "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, 2005, *43* (3), 762–800.

Algan, Yann, Thierry Mayer, and Mathias Thoenig, "The Economic Incentives of Cultural Transmission: Spatial Evidence from Naming Patterns across France," 2013. CEPR Discussion Paper No. DP9416.

Amior, Michael, "The Contribution of Immigration to Local Labor Market Adjustment," 2020. Working Paper.

Angrist, Josh, "How Do Sex Ratios Affect Marriage and Labor Markets? Evidence from America's Second Generation," *Quarterly Journal of Economics*, 2002, *117* (3), 997–1038.

Angrist, Joshua D. and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton: Princeton University Press, 2008.

Bandiera, Oriana, Imran Rasul, and Martina Viarengo, "The Making of Modern America: Migratory Flows in the Age of Mass Migration," *Journal of Development Economics*, 2013, *102*, 23–47.

— **, Myra Mohnen, Imran Rasul, and Martina Viarengo**, "Nation-Building through Compulsory Schooling During the Age of Mass Migration," *Economic Journal*, 2019, *129* (617), 62–109.

Bazzi, Samuel, Arya Gaduh, Alexander D Rothenberg, and Maisy Wong, "Unity in Diversity? How Intergroup Contact Can Foster Nation Building," *American Economic Review*, 2019, *109* (11), 3978–4025.

Berry, John W. and Rudolf Kalin, "Reciprocity of Inter-Ethnic Attitudes in a Multicultural Society," *International Journal of Intercultural Relations*, 1979, *3* (1), 99–111.

Biavaschi, Costanza, Corrado Giulietti, and Zahra Siddique, "The Economic Payoff of Name Americanization," *Journal of Labor Economics*, 2017, *35* (4), 1089–1116.

Bisin, Alberto and Giulia Tura, "Marriage, Fertility, and Cultural Integration in Italy," 2019. NBER Working Paper No. 26303.

Black, Dan A., Seth G. Sanders, Evan J. Taylor, and Lowell J. Taylor, "The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South," *American Economic Review*, 2015, *105* (2), 477–503.

Bodnar, John, Michael Weber, and Roger Simon, "Migration, Kinship, and Urban Adjustment: Blacks and Poles in Pittsburgh, 1900-1930," *The Journal of American History*, 1979, *66* (3), 548–565.

Bonomi, Giampaolo, Nicola Gennaioli, and Guido Tabellini, "Identity, Beliefs, and Political Conflict," 2020. Working Paper.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer, "Stereotypes," *Quarterly Journal of Economics*, 2016, *131* (4), 1753–1794.

Borjas, George J., "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 1987, *77* (4), 531–553.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-Experimental Shift-Share Research Designs," 2018. NBER Working Paper No. 24997.

Boustan, Leah Platt, "Was Postwar Suburbanization "White Flight"? Evidence from the Black Migration," *Quarterly Journal of Economics*, 2010, *125* (1), 417–443.

_ , *Competition in the Promised Land: Black Migrants in Northern Cities and Labor Markets*, Princeton: Princeton University Press, 2016.

_ , Matthew E. Kahn, and Paul W. Rhode, "Moving to Higher Ground: Migration Response to Natural Disasters in the Early Twentieth Century," *American Economic Review*, 2012, *102* (3), 238–244.

Campante, Filipe R and David Yanagizawa-Drott, "The Intergenerational Transmission of War," 2015. NBER Working Paper No. 21371.

Card, David, "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, 2001, *19* (1), 22–64.

Carneiro, Pedro Manuel, Sokbae Lee, and Hugo Reis, "Please Call me John: Name Choice and the Assimilation of Immigrants in the United States, 1900-1930," 2016. IZA Discussion Paper No. 9792.

Charnysh, Volha, "Diversity, Institutions, and Economic Outcomes: Post-WWII Displacement in Poland," *American Political Science Review*, 2019, *113* (2), 423–441.

Chiswick, Barry R. and Paul W. Miller, "Linguistic Distance: A Quantitative Measure of the Distance between English and other Languages," *Journal of Multilingual and Multicultural Development*, 2005, *26* (1), 1–11.

Cho, Sumi K., "Korean Americans vs. African Americans: Conflict and Construction," in Robert Gooding-Williams, ed., *Reading Rodney King: Reading Urban Uprising*, New York: Routledge, 1993.

Collins, William J., "When the Tide Turned: Immigration and the Delay of the Great Black Migration," *Journal of Economic History*, 1997, *57* (3), 607–632.

_ , "The Great Migration of Black Americans from the US South: A Guide and Interpretation," 2020. NBER Working Paper No. 27268.

_ and Marianne H. Wanamaker, "Selection and Income Gains in the Great Migration of African Americans," *American Economic Journal: Applied Economics*, 2014, *6* (1), 220–252.

_ and _ , "The Great Migration in Black and White: New Evidence on the Selection and Sorting of Southern Migrants," *Journal of Economic History*, 2015, *75* (4), 947–992.

Derenoncourt, Ellora, "Can you Move to Opportunity? Evidence from the Great Migration," 2019. Working Paper.

Duckitt, John H., *The Social Psychology of Prejudice*, Westport: Greenwood Press, 1994.

Dumenil, Lynn, "The Tribal Twenties: "Assimilated" Catholics' Response to Anti-Catholicism in the 1920s," *Journal of American Ethnic History*, 1991, *11* (1), 21–49.

Dustmann, Christian, Kristine Vasiljeva, and Anna Piil Damm, "Refugee Migration and Electoral Outcomes," *The Review of Economic Studies*, 2019, *86* (5), 2035–2091.

Echenique, Federico and Roland G. Fryer, Jr., "A Measure of Segregation Based on Social Interactions," *Quarterly Journal of Economics*, 2007, *122* (2), 441–485.

Edin, Per-Anders, Peter Fredriksson, and Olof Åslund, "Ethnic Enclaves and the Economic Success of Immigrants – Evidence from a Natural Experiment," *Quarterly Journal* *of Economics*, 2003, *118* (1), 329–357.

Eriksson, Katherine, "Ethnic Enclaves and Immigrant Outcomes: Norwegian Immigrants during the Age of Mass Migration," *European Review of Economic History*, 2020, *24* (3), 427–446.

Feigenbaum, James J, Soumyajit Mazumder, and Cory B Smith, "When Coercive Economies Fail: The Political Economy of the US South After the Boll Weevil," 2020. NBER Working Paper No. 27161.

Ferenczi, Imre and Walter F. Willcox, *International Migrations Volume I: Statistics*, New York: National Bureau of Economic Research, 1929.

Ferrara, Andreas and Price V Fishback, "Discrimination, Migration, and Economic Outcomes: Evidence from World War I," 2020. NBER Working Paper No. 26936.

Ferrie, Joseph, "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*, 1996, *29* (4), 141–156.

Foged, Mette and Giovanni Peri, "Immigrants' Effect on Native Workers: New Analysis on Longitudinal Data," *American Economic Journal: Applied Economics*, 2016, *8* (2), 1–34.

Fouka, Vasiliki, "How do Immigrants Respond to Discrimination: The Case of Germans in the US during World War I," *American Political Science Review*, 2019, *113* (2), 405–422.

_ , "Backlash: The Unintended Effects of Language Prohibition in U.S. Schools after World War I," *The Review of Economic Studies*, 2020, *87* (1), 204–239.

_ , Soumyajit Mazumder, and Marco Tabellini, "Changing In-Group Boundaries: The Effect of Immigration on Race Relations in the US," 2020. Harvard Business School BGIE Unit Working Paper No. 20-100.

Fryer, Roland and Matthew O. Jackson, "A Categorical Model of Cognition and Biased Decision Making," *BE Journal of Theoretical Economics*, 2008, *8* (1), 1–44.

Fryer, Roland G. and Steven D. Levitt, "The Causes and Consequences of Distinctively Black Names," *Quarterly Journal of Economics*, 2004, *119* (3), 767–805.

Gentzkow, Matthew and Jesse M. Shapiro, "What Drives Media Slant? Evidence from U.S. Daily Newspapers," *Econometrica*, 2010, *78* (1), 35–71.

Glaeser, Edward L, "The Political Economy of Hatred," *The Quarterly Journal of Economics*, 2005, *120* (1), 45–86.

Goldin, Claudia, "The Political Economy of Immigration Restriction in the United States, 1890 to 1921," in Claudia Goldin and Gary D. Libecap, eds., *The Regulated Economy: A Historical Approach to Political Economy*, Chicago: University of Chicago Press, 1994.

Goldin, Claudia D, "The Role of World War II in the Rise of Women's Employment," *The American Economic Review*, 1991, pp. 741–756.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik Instruments: What, When, Why, and How," *American Economic Review*, August 2020, *110* (8), 2586–2624.

Gordon, Milton A., *Assimilation in American Life: The Role of Race, Religion, and National Origins*, New York: Oxford University Press, 1964.

Gregory, James N., *The Southern Diaspora: How the Great Migrations of Black and White Southerners Transformed America*, Chapel Hill: University of North Carolina Press, 2005.

_ , "IWW History Project," http://depts.washington.edu/iww/ 2015.

Guglielmo, Thomas A., *White on Arrival: Italians, Race, Color, and Power in Chicago, 1890-1945*, New York: Oxford University Press, 2003.

Guterl, Matthew Pratt, *The Color of Race in America, 1900-1940*, Cambridge: Harvard University Press, 2001.

Hagendoorn, Louk, "Intergroup Biases in Multiple Group Systems: The Perception of Ethnic Hierarchies," *European Review of Social Psychology*, 1995, *6* (1), 199–228.

Hainmueller, Jens and Daniel J Hopkins, "Public Attitudes toward Immigration," *Annual Review of Political Science*, 2014, *17*, 225–49.

_ **and Dominik Hangartner**, "Who Gets a Swiss Passport? A Natural Experiment in Immigrant Discrimination," *American Political Science Review*, 2013, *107* (1), 159–187.

Halla, Martin, Alexander F Wagner, and Josef Zweimüller, "Immigration and Voting for the Far Right," *Journal of the European Economic Association*, 2017, *15* (6), 1341–1385.

Haslam, S. Alexander, Penelope J. Oakes, John C. Turner, and Craig McGarty, "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*, 1995, *34* (2), 139–160.

Hatton, Timothy J. and Jeffrey G. Williamson, *The Age of Mass Migration: Causes and Economic Impact*, New York: Oxford University Press, 1998.

Higham, John, *Strangers in the Land: Patterns of American Nativism, 1860-1925*, New Brunswick: Rutgers University Press, 1998.

Hochschild, Jennifer L. and Brenna Marea Powell, "Racial Reorganization and the United States Census 1850-1930: Mulattoes, Half-Breeds, Mixed Parentage, Hindoos, and the Mexican Race," *Studies in American Political Development*, 2008, *22* (1), 59–96.

Hornbeck, Richard and Suresh Naidu, "When the Levee Breaks: Black Migration and Economic Development in the American South," *American Economic Review*, 2014, *104* (3), 963–990.

Ignatiev, Noel, *How the Irish Became White*, New York: Routledge, 1995.

Jacobson, Matthew Frye, *Whiteness of a Different Color*, Cambridge: Harvard University Press, 1999.

Jaeger, David A., Joakim Ruist, and Jan Stuhler, "Shift-Share Instruments and the Impact of Immigration," 2018. NBER Working No. 24285.

Kahneman, Daniel and Amos Tversky, "Subjective Probability: A Judgment of Representativeness," in Carl-Axel S. Staël Von Holstein, ed., *The Concept of Probability in Psychological Experiments*, Dordrecht: Springer Netherlands, 1974.

Katz, Daniel and Kenneth W. Braly, "Racial Stereotypes of One Hundred College Students," *The Journal of Abnormal and Social Psychology*, 1993, *28* (3), 280–290.

Katz, Lawrence F and Robert A Margo, "Technical Change and the Relative Demand for Skilled Labor: The United States in Historical Perspective," in "Human capital in history: The American record," University of Chicago Press, 2014, pp. 15–57.

Lange, Fabian, Alan L. Olmstead, and Paul W. Rhode, "The Impact of the Boll Weevil, 1892-1932," *Journal of Economic History*, 2009, *69* (3), 685–718.

Lee, Everett S., Ann Ratner Miller, Carol P. Brainerd, and Richard A. Easterlin, "Methodological Considerations and Reference Tables," in Simon Kuznets and Dorothy Swaine, eds., *Population redistribution and economic growth: United States, 1870-1950*, Philadelphia: American Philosophical Society, 1957.

Lleras-Muney, Adriana and Allison Shertzer, "Did the Americanization Movement Succeed? An Evaluation of the Effect of English-Only and Compulsory Schooling Laws on Immigrants," *American Economic Journal: Economic Policy*, 2015, *7* (3), 258–290.

Logan, Trevon D. and John M. Parman, "The National Rise in Residential Segregation," *Journal of Economic History*, 2017, *77* (1), 127–170.

Massey, Douglas S. and Nancy A. Denton, *American Apartheid: Segregation and the Making of the Underclass*, Cambridge: Harvard University Press, 1993.

Mazumder, Soumyajit, "Becoming White: How Mass Warfare Transformed Immigrants into Americans," 2018. Working Paper.

McConnell, Brendon and Imran Rasul, "Contagious Animosity in the Field: Evidence from the Federal Criminal Justice Crime System," *Journal of Labor Economics*, 2020, *Forth-*

coming.

McDevitt, Jack, Jack Levin, and Susan Bennett, "Hate Crime Offenders: An Expanded Typology," *Journal of Social Issues*, 2002, *58* (2), 303–317.

McGarty, Craig, *Categorization in Social Psychology*, London: Sage Publications, 1999.

Moser, Petra, "Taste-Based Discrimination. Evidence from a Shift in Ethnic Preferences after WWI," *Explorations in Economic History*, April 2012, *49* (2), 167–188.

Myrdal, Gunnar, *An American Dilemma: The Negro Problem and Modern Democracy*, New York: Harper & Bros, 1944.

Olzak, Susan and Suzanne Shanahan, "Prisoners and Paupers: The Impact of Group Threat on Incarceration in Nineteenth Century American Cities," *American Sociological Review*, 2014, *79* (3), 392–411.

Peri, Giovanni and Chad Sparber, "Task Specialization, Immigration, and Wages," *American Economic Journal: Applied Economics*, 2009, *1* (3), 135–69.

Pettigrew, Thomas F., "Social Distance Attitudes of South African Students," *Social Forces*, 1960, *38* (3), 246–253.

Putnam, Robert D., "E Pluribus Unum: Diversity and Community in the Twenty-first Century The 2006 Johan Skytte Prize Lecture," *Scandinavian political studies*, 2007, *30* (2), 137–174.

Qian, Nancy and Marco Tabellini, "Discrimination, Disenfranchisement and African American WWII Military Enlistment," 2020. Working Paper.

Ray, Debraj and Joan Esteban, "Conflict and Development," *Annual Review of Economics*, 2017, *9*, 263–293.

Rieder, Jonathan, *Canarsie. The Jews and Italians of Brooklyn against Liberalism*, Cambridge: Harvard University Press, 1987.

Roediger, David R., *The Wages of Whiteness: Race and the Making of the American Working Class*, London: Verso, 1999.

Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek, *Integrated Public Use Microdata Series: Version 5.0 [Machine-Readable Database]*, University of Minnesota, 2010.

_ **, Katie Genadek, Ronald Goeken, Joasiah Grover, and Matthew Sobek**, "Integrated Public Use Microdata Series: Version 6.0 [dataset].," *Minneapolis: University of Minnesota*, 2015.

Sequeira, Sandra, Nathan Nunn, and Nancy Qian, "Immigrants and the Making of America," *The Review of Economic Studies*, 2020, *87* (1), 382–419.

Shayo, Moses, "A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution," *American Political Science Review*, 2009, *103* (2), 147–174.

_ , "Social Identity and Economic Policy," *Annual Review of Economics*, 2020, *12*, 355–389.

Shertzer, Allison, "Immigrant Group Size and Political Mobilization: Evidence from European Migration to the United States," *Journal of Public Economics*, 2016, *139*, 1–12.

_ **and Randall P. Walsh**, "Racial Sorting and the Emergence of Segregation in American Cities," *The Review of Economics and Statistics*, 2019, *101* (3), 415–427.

Spiro, Jonathan P., *Defending the Master Race: Conservation, Eugenics, and the Legacy of Madison Grant*, Burlington: University of Vermont Press, 2008.

Spolaore, Enrico and Romain Wacziarg, "The Diffusion of Development," *Quarterly Journal of Economics*, 2009, *124* (2), 469–529.

Stuart, Bryan A and Evan J Taylor, "The Effect of Social Connectedness on Crime: Evidence from the Great Migration," *Review of Economics and Statistics*, 2017, pp. 1–46.

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, *Database of [United States] Congressional Historical Statistics, 1789-1989*, Ann Arbor:

Inter-university Consortium for Political and Social Research [distributor], 1989.

Tabellini, Marco, "Racial Heterogeneity and Local Government Finances: Evidence from the Great Migration," 2018.

_ , "Gifts of the Immigrants, Woes of the Natives: Lessons from the Age of Mass Migration," *The Review of Economic Studies*, 2020, *87* (1), 454–486.

Tajfel, Henri and Alan L. Wilkes, "Classification and Quantitative Judgment," *British Journal of Psychology*, 1963, *54* (2), 101–114.

_ **, Michael G. Billig, Robert P. Bundy, and Claude Flament**, "Social Categorization and Intergroup Behaviour," *European Journal of Social Psychology*, 1971, *1* (2), 149–178.

Tolnay, S. and E. M. Beck, "Black Flight: Lethal Violence and the Great Migration, 1900 to 1930," *Social Science History*, 1990, *14* (3), 347–370.

Turner, John C., Michael A. Hogg, Penelope J. Oakes, Stephen D. Reicher, and Margaret S. Wetherell, *Rediscovering the Social Group: A Self-Categorization Theory*, Cambridge: Basil Blackwell, 1987.

_ **, Penelope J. Oakes, S. Alexander Haslam, and Craig McGarty**, "Self and Collective: Cognition and Social Context," *Personality and Social Psychology Bulletin*, 1994, *20* (5), 454–463.

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

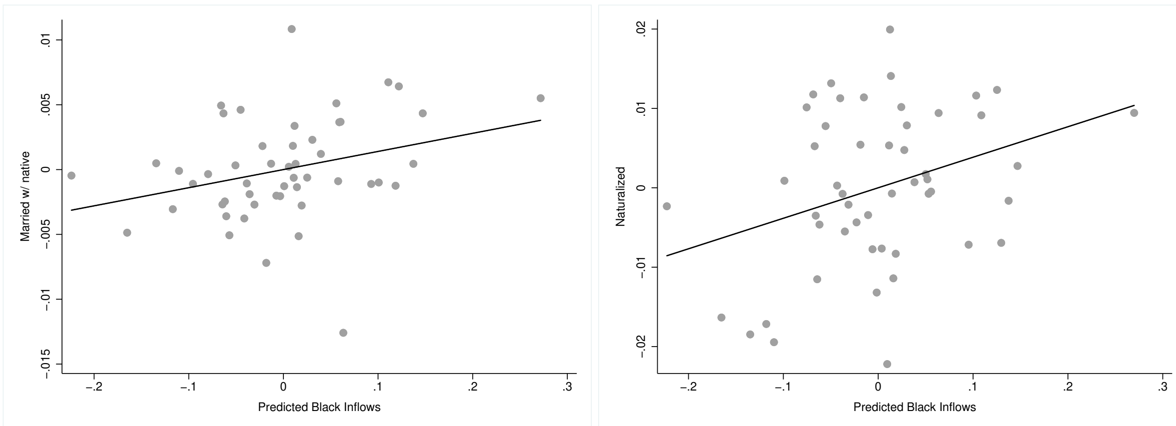
Wilkerson, Isabel, *The Warmth of Other Suns: The Epic Story of America's Great Migration*, New York: Random House, 2010.

Woodward, C. Vann, *The Strange Career of Jim Crow*, Oxford: Oxford University Press, 2002.

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

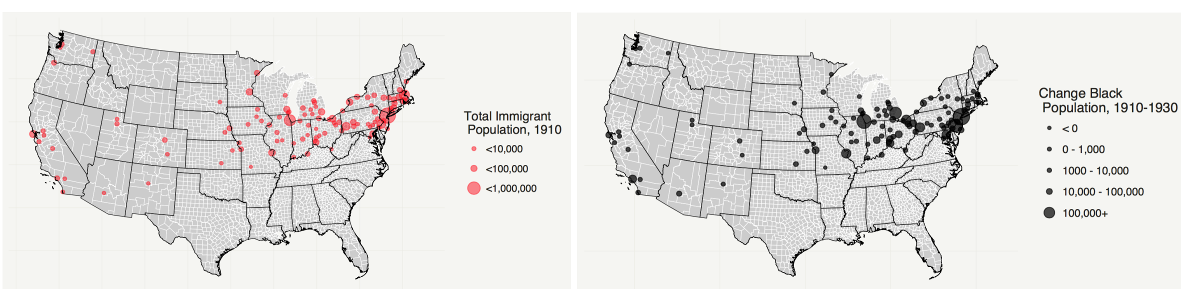
Figures and Tables

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



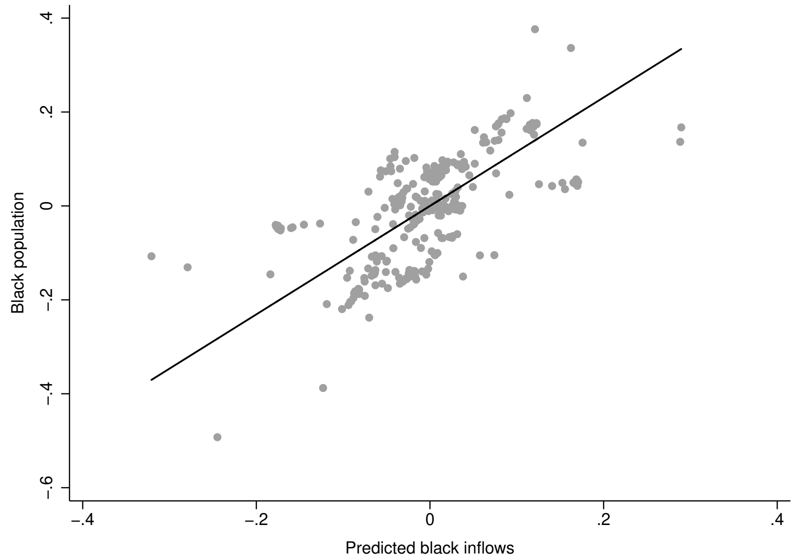
Notes: Binned scatterplot of the relationship between the change in the probability of marrying a native-born person of native-born parents (left panel) or of being a naturalized citizen (right panel) and the change in predicted Black population for the years 1910-1930. Variables on the x- and y-axis represent residual changes, after partialling out total MSA population, MSA and region by decade 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



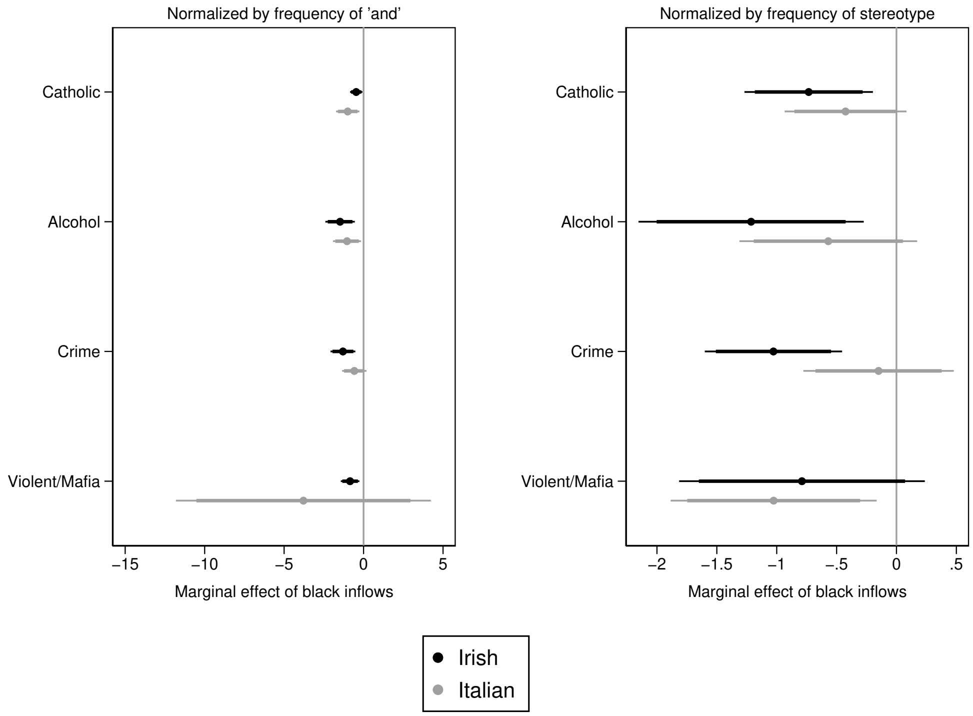
Notes: The map depicts the total number of foreign-born residents in 1910 (left panel) and the change in the number of African Americans between 1910 and 1930 (right panel) in the 108 MSAs in our sample.

Figure 3. First stage



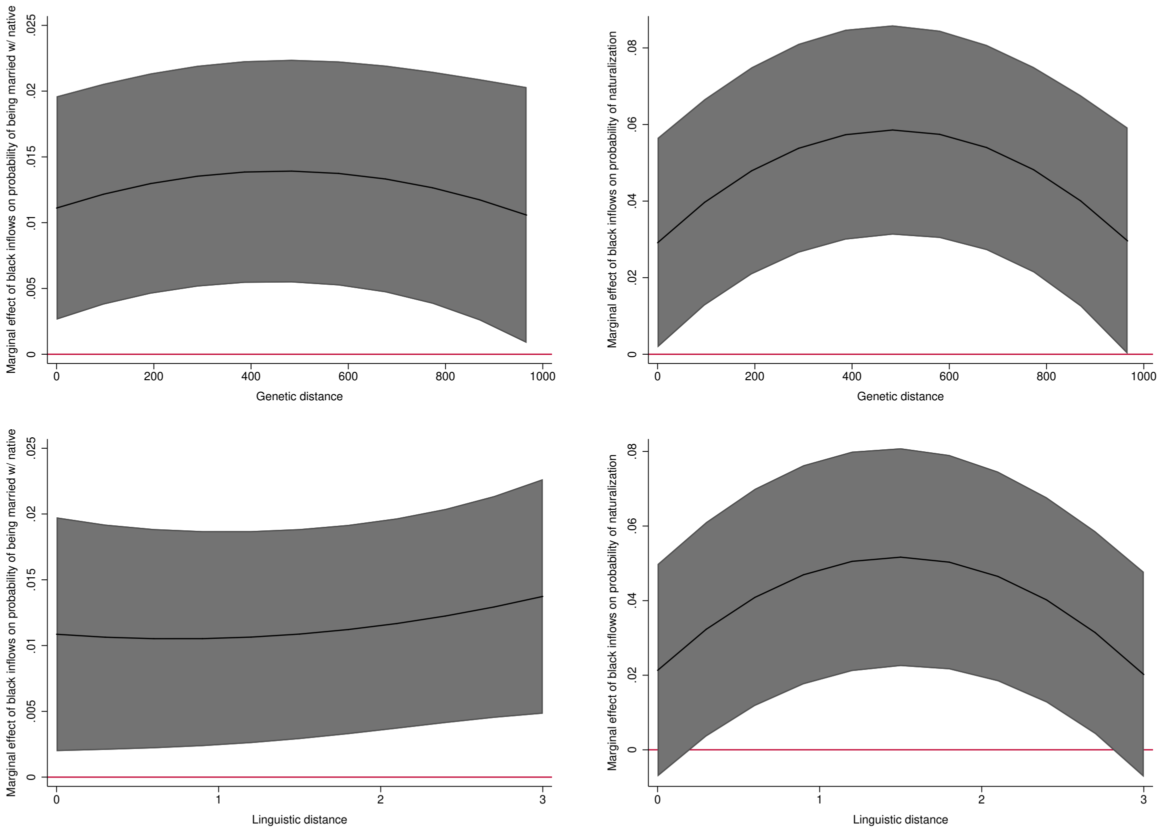
Notes: The figure shows the relationship between actual and predicted Black population for the years 1910 to 1930. Each point represents the residual change in an MSA's actual (y-axis) and predicted (x-axis) number of Blacks after partialling out total MSA population and MSA and region by decade fixed effects.

Figure 4. Effects on the frequency of national stereotypes



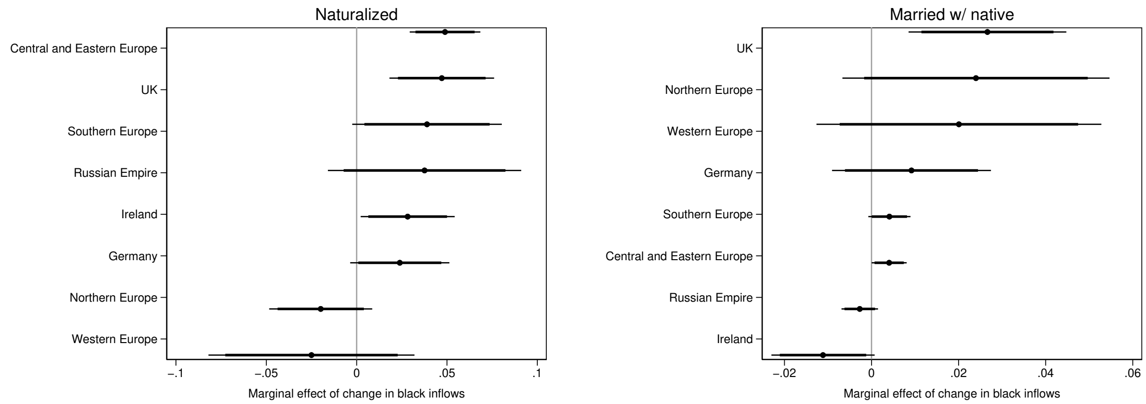
Notes: The figure plots estimates of the marginal effect of Black inflows on the relative frequency of national stereotypes. Each line corresponds to a separate regression. The dependent variable is the frequency of the co-occurrence of each term on the y-axis with the word "Irish" (Black lines) or "Italian" (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 5. Heterogeneity by distance from native-born whites



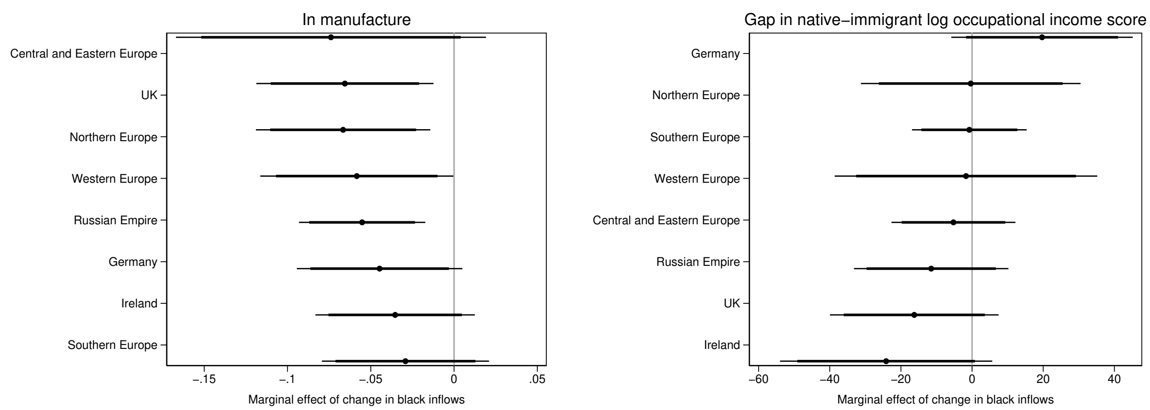
Notes: The figure plots the marginal effect of Black inflows on the probability of being married with a native-born person of native-born parents (left) and the probability of being a US citizen (right) against values of genetic (upper panel) and linguistic distance (lower panel). The underlying regressions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

Figure 6. Heterogeneity by national origin



Notes: The figure plots 2SLS estimates of the marginal effect of Black inflows on the probability of being a naturalized citizen (left panel) or being married to a native-born person of native-born parents (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.

Figure 7. Heterogeneity by national origin – Economic outcomes



Notes: 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 native-born whites and immigrants (right panel) from regressions separately run by national origin group that control for age, years in the US, MSA and region by decade fixed effects. The sample is restricted to men aged 15 to 65. Thick and thin lines represent 90% and 95% confidence intervals, respectively.

Table 1. Summary statistics

Variables	Mean	Median	S.D.	Min	Max	N
MSA characteristics						
Total population	420,640	159,581	992,887	23,606	10,900,000	324
Foreign-born population	94,648	21,478	308,618	781	3,338,862	324
Black population	15,294	2,349	44,830	29	485,750	324
Predicted Black population	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 Black	0.029	0.013	0.039	0.001	0.293	324
Main outcomes						
Married with native	0.150	0.132	0.085	0.036	0.498	324
Naturalized	0.568	0.569	0.141	0.125	0.916	324
Additional outcomes						
In manufacture	0.246	0.229	0.129	0.026	0.551	324
Unskilled	0.400	0.380	0.136	0.102	0.865	324
Speaks English	0.868	0.904	0.117	0.399	1	324
Foreign name index	54.571	54.594	6.689	36.185	75.235	324
Ethnic segregation (both neighbors)	0.399	0.400	0.181	0.049	0.778	324
Ethnic segregation (one neighbor)	0.427	0.428	0.178	0.075	0.775	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-born parentage. Share naturalized refers to the share of immigrant men who are US citizens. For the construction of additional outcomes see Appendix Section C.1.

Table 2. Effects of Black inflows on intermarriage and naturalization rates

	(1) OLS	(2) OLS	(3) 2SLS	(4) 2SLS	(5) 2SLS	(6) 2SLS
	Panel A: Married w/ native *(1910 mean: 0.071)*					
Black population	0.007	0.005	0.012**	0.013***	0.012**	0.019
	(0.006)	(0.004)	(0.005)	(0.005)	(0.005)	(0.017)
Observations	9,323,126	9,323,126	9,323,126	9,323,109	9,323,109	88,892
F-stat			23.33	23.71	23.83	32.75
	Panel B: Naturalized *(1910 mean: 0.491)*					
Black population	0.056**	0.030*	0.033**	0.038***	0.033**	0.144***
	(0.025)	(0.016)	(0.015)	(0.013)	(0.014)	(0.031)
Observations	15,267,846	15,267,846	15,267,846	15,267,844	15,267,844	80,866
F-stat			24.23	24.38	24.51	32.74
Individual controls	X	X	X	X	X	
Region × Decade		X	X	X	X	
MSA × Origin region				X	X	
Origin region × Decade					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-born spouse of native-born parentage. Columns 1 to 5 present results obtained from the repeated cross-sections, while column 6 shows results from the linked panel of men who always remained in the same MSA in the three census years. Columns 1-2 (resp. 3-6) 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. Regressions in column 6 include individual fixed effects. In column 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). F-stat refers to the KP F-stat for weak instruments. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

Table 3. Effects of Black inflows on other assimilation outcomes

Dep. variable	In manufacture	Unskilled	Speak English	Log FNI	Ethnic segregation	
					Both neighbors	One neighbor
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Repeated cross-section			
Black population	-0.054**	-0.021***	0.013	-0.050**	-0.047*	-0.050**
	(0.025)	(0.007)	(0.023)	(0.022)	(0.025)	(0.026)
Observations	14,055,929	14,055,929	15,365,318	4,499,394	4,904	4,918
F-stat	24.82	24.82	24.68	27.04	18.57	18.55
			Panel B: Linked sample			
Black population	-0.020**	-0.014**	0.053	-	-	-
	(0.010)	(0.006)	(0.034)	-	-	-
Observations	261,867	261,867	18,298	-	-	-
F-stat	28.56	28.56	56.56	-	-	-

Notes: In columns 1 to 3 the 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. It is further restricted to individuals aged 15-65 (columns 1 and 2) and to individuals who did not speak English in the previous decade (column 3 of Panel B). In column 4, the sample consists of US-born men of foreign-born fathers who were born in the 10 years before each census year. Columns 5 and 6 present results from a dataset at the MSA-nationality-decade level. Regressions in columns 1 to 4 correspond to the specification of column 3 in Table 2. Individual-level controls in column 4 include year of birth fixed effects and father's region of origin indicators interacted with decade and MSA dummies. Regressions in columns 5 and 6 control for MSA total population, region by decade and nationality by decade fixed effects, and are weighted by MSA population in 1900. Regressions in Panel B include individual fixed effects. F-stat refers to the KP F-stat for weak instruments. 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 the press

Dep. variable	Relative frequency of word over frequency of word "and"					
	Immigration	Quotas	Dago	Catholic	Catholic Threat	KKK
	(1)	(2)	(3)	(4)	(5)	(6)
Black population	-1.005*	-2.000**	-1.382**	-0.923***	-1.283***	-3.837**
	(0.557)	(0.819)	(0.689)	(0.305)	(0.364)	(1.878)
Observations	195	195	195	195	195	195
F-stat	17.67	17.67	17.67	17.67	17.67	17.67
MSAs	71	71	71	71	71	71

Notes: The analysis is restricted to 71 non-southern MSAs for which the instrument and the word frequency measure could be constructed, and to years 1910 to 1930. *Immigration* refers to the joint frequency of the terms "immigration", "immigrants" and "aliens." Dependent variables are standardized by subtracting the mean and dividing by the standard deviation in 1910. All regressions control for MSA and region by decade fixed effects and MSA total population and are weighted by MSA population in 1900. F-stat refers to the KP F-stat for weak instruments. Robust standard errors clustered at the MSA level in parentheses. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

Table 5. Frequency of Black stereotypes in the press

Dep. variable	Frequency of joint mentions						
	Violent	Dirty	Rapist	Crime	Ignorant	Slovenly	Vagrant
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Joint occurrence with "Negro"						
Black population	0.725***	0.553*	1.245***	0.379	0.904***	1.087***	0.821**
	(0.212)	(0.334)	(0.466)	(0.332)	(0.303)	(0.210)	(0.359)
Observations	193	193	193	193	193	193	193
F-stat	17.12	17.12	17.12	17.12	17.12	17.12	17.12
	Panel B: Joint occurrence with "Black"						
Black population	1.186***	0.806*	1.071***	0.663***	0.817**	1.102***	1.113***
	(0.233)	(0.429)	(0.351)	(0.252)	(0.324)	(0.367)	(0.242)
Observations	193	193	193	193	193	193	193
F-stat	17.12	17.12	17.12	17.12	17.12	17.12	17.12
MSAs	71	71	71	71	71	71	71

Notes: The analysis is restricted to 71 non-southern MSAs for which the instrument 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 region by decade fixed effects and MSA total population and are weighted by 1900 MSA total population. F-stat refers to the KP F-stat for weak instruments. Robust 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 native-born whites

Dep. variable	Married w/ native (1)	Naturalized (2)	Married w/ native (3)	Naturalized (4)
Black population	0.011**	0.029**	0.011**	0.021
	(0.004)	(0.014)	(0.005)	(0.015)
Black population × Genetic distance	0.012**	0.121***		
	(0.005)	(0.026)		
Black population × Genetic distance^2	-0.013**	-0.125***		
	(0.006)	(0.031)		
Black population × Linguistic distance			-0.001	0.041***
			(0.002)	(0.006)
Black population × Linguistic distance^2			0.001	-0.014***
			(0.001)	(0.002)
Observations	9,220,664	14,978,414	9,219,280	15,012,424
F-stat	8.145	8.371	7.969	8.211
F-stat (Wald)	4.42	16.70	1.23	32.79
p-value	0.0378	0.0001	0.2698	0.000

Notes: The analysis is restricted to 108 non-southern MSAs for which the instrument could be constructed, and to years 1910, 1920, and 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 region by decade, and control for MSA total population. F-stat refers to the KP F-stat for weak instruments. The last two rows report the F statistic and corresponding p-values from a Wald test, testing the quadratic model against the restricted linear alternative. Robust standard errors clustered at the MSA level in parentheses. Significance levels: *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

Table 7. Heterogeneity by sector of employment and skill in 1900

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

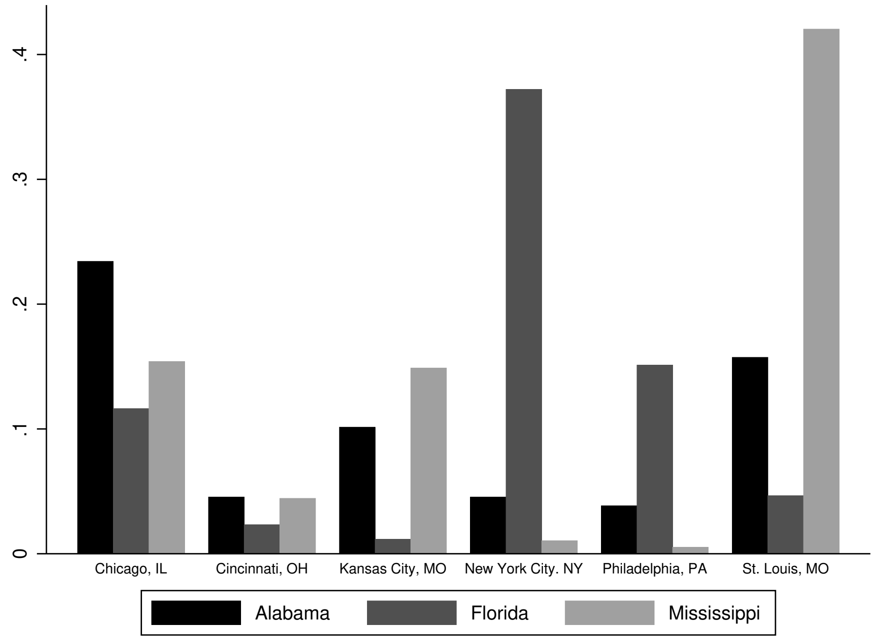
Notes: The analysis is restricted to 108 non-southern MSAs for which the instrument could be constructed, and to years 1910, 1920, and 1930. The sample is restricted to foreign-born men. In columns 1 and 3 it is further restricted to married individuals. All regressions include fixed effects for age, years in the US, MSA by nationality, nationality by year and region by decade, and control for MSA total population. F-stat refers to the KP F-stat for weak instruments. Robust 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

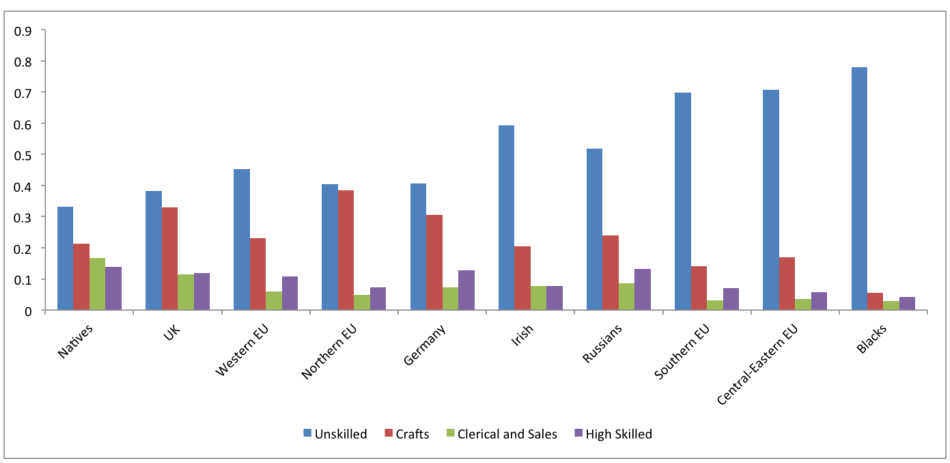
A Additional Figures and Tables

Figure A.1. Share of Blacks from selected Southern states in Northern MSAs, 1900



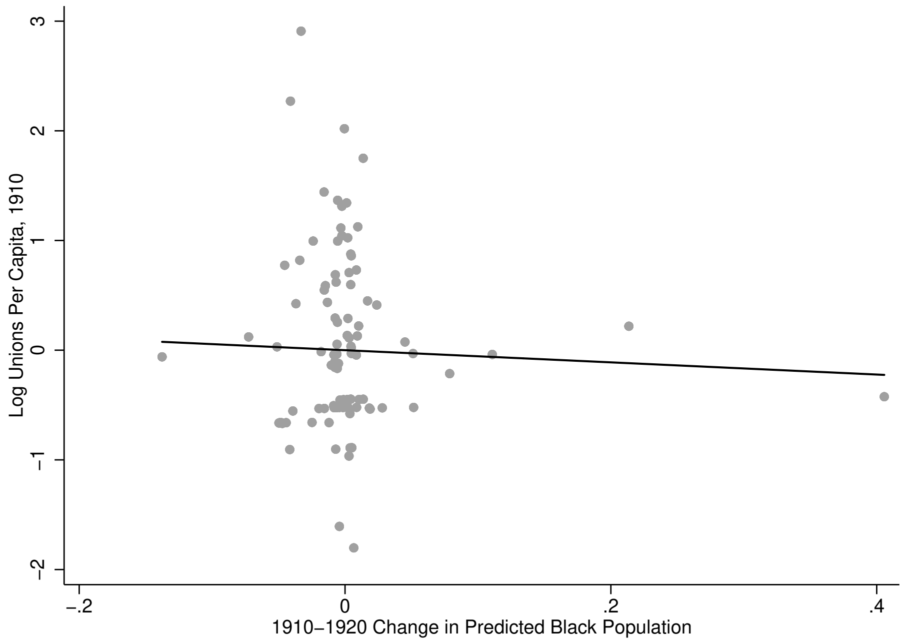
Notes: The figure shows the fraction of Southern-born Blacks from a given state who resided in the North that lived in one of the selected Northern MSAs in 1900. Authors' calculations from 1900 US Census (Ruggles et al., 2015).

Figure A.2. Skill distribution for men aged 15-65 in 1910



Notes: 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 (2014). Authors' calculations from the 1910 full count Census (Ruggles et al., 2015).

Figure A.3. Correlation between local labor unions in 1910 and change in predicted Black inflows



Notes: The figure represents the residual scatterplot for a regression of the logarithm of local IWW chapters in 1910 (y-axis) against the 1910 to 1920 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
Cincinatti-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	0.071	0.265	9,323,128	0.105	0.306	228,677
Naturalized	0.491	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
Unskilled	0.411	0.492	14,055,931	0.278	0.448	269,657
Speaks English	0.871	0.335	15,365,327	0.951	0.216	302,619
Foreign name index	59.065	28.342	4,499,505	-	-	-

Notes: The repeated cross-section 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 consists of foreign-born men who could be linked across census years (as described in Section C.2) always living in one of the 108 non-southern MSAs in the three census years, 1910, 1920, and 1930. When computing the share in manufacture and share unskilled, the sample is restricted to men aged 15-65.

Table A.3. List of MSAs with newspaper data

Akron, OH	Detroit, MI	New Haven-Meriden, CT	Wilkes-Barre-Hazleton, PA
Easton, PA/MA	Duluth-Superior, MN/WI	New York, NY	Seattle-Everett, WA
Altoona, PA	Evansville, IN/KY	Omaha, NE/IA	Sioux City, IA/NE
Baltimore, MD	Fort Wayne, IN	Philadelphia, PA/NJ	Sioux Falls, SD
Binghamton, NY	Green Bay, WI	Phoenix-Mesa, AZ	Spokane, WA
Boston, MA	Hamilton-Middletown, OH	Pittsburgh, PA	Springfield, MO
Bridgeport, CT	Harrisburg-Leban-Carlisle, PA	Pittsfield, MA	Syracuse, NY
Buffalo, NY	Hartford, CT	Portland-Vancouver, OR/WA	Tacoma, WA
Canton, OH	Indianapolis, IN	Racine, WI	Topeka, KS
Cedar Rapids, IA	Kansas City, MO/KS	Reading, PA	Trenton, NJ
Chicago, IL	Lansing-East Lansing, MI	San Bernardino, CA	Washington DC, MD/VA/WV
Cincinatti-Hamilton, OH/KY/IN	Lima, OH	Rochester, NY	Waterloo-Cedar Falls, IA
Cleveland, OH	Lincoln, NE	St. Joseph, MO	Wichita, KS
Columbus, OH	Lorain-Elyria, OH	St. Louis, MO/IL	Wilmington, DE/NJ/MD
Davenport-Moline-Rock Island, IA/IL	Los Angeles-Long Beach, CA	Salt Lake City-Ogden, UT	Worcester, MA
Dayton-Sprinfield, OH	Milwaukee-Waukesha, WI	Ogden, UT	York,
Decatur, IL	Minneapolis-St. Paul, MN	San Francisco, CA	Youngstown-Warren, OH
Des Moines, IA	Muncie, IN	Scranton-Wilkes-Barre-Hazleton, PA	

Notes: This table lists the subset of MSAs of Table A.1 for which newspaper data are available from Newspapers.com.

Table A.4. Comparing MSAs in the newspapers sample with full sample

Variable	Newspapers Sample	Full Sample	Difference	p-value
1900 total population	340,894	251,974	88,920	0.121
1900 foreign-born population	85,867	61,172	24,695	0.185
1900 Black population	21,929	15,737	6,192	0.099
1900 share foreign-born	0.179	0.181	-0.002	0.759
1900 share Black	0.055	0.048	0.007	0.282
1910-1930 change MSA population	265,897	186,924	78,973	0.09
1910-1930 change Black population	21,149	13,812	7,337	0.099

Notes: The table reports the average of each variable in the newspapers and full sample in columns 2 and 3 respectively. The last column reports the p-value from a t-test for the equality of means between the two samples of MSAs.

Table A.5. Baseline results for MSAs with newspapers

Dep. variable	Married w/ native		Naturalized	
	(1)	(2)	(3)	(4)
Black population	0.012**	0.014***	0.033**	0.042***
	(0.005)	(0.005)	(0.015)	(0.014)
Observations	9,323,126	8,061,855	15,267,846	13,194,260
F-stat	23.33	26.49	24.23	27.58
Sample	Baseline	Newspapers dataset	Baseline	Newspapers dataset

Notes: Columns 1 and 3 of the table report the baseline results in column 3 of Table 2. Columns 2 and 4 of the table replicate the specification in column 3 of Table 2 for the sample of MSAs for which newspaper data are available. F-stat refers to the KP F-stat for weak instruments. 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

Dep. variable	Frequency of joint mentions							
	Irish& Catholic	Irish& alcohol	Irish& crime	Irish& violent	Italian& Catholic	Italian& alcohol	Italian& crime	Italian& mafia
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Normalized by frequency of word "and"							
Black population	-0.464**	-1.477***	-1.298***	-0.848***	-0.996***	-1.039**	-0.581	-3.786
	(0.199)	(0.472)	(0.399)	(0.298)	(0.376)	(0.450)	(0.391)	(4.092)
Observations	197	197	197	197	197	197	197	197
F-stat	17.51	17.51	17.51	17.51	17.51	17.51	17.51	17.51
MSAs	71	71	71	71	71	71	71	71
	Panel B: Normalized by frequency of the stereotype							
Black population	-0.733***	-1.214**	-1.027***	-0.789	-0.425	-0.569	-0.149	-1.026**
	(0.274)	(0.480)	(0.293)	(0.523)	(0.260)	(0.379)	(0.320)	(0.439)
Observations	192	188	194	192	192	188	194	187
F-stat	17.66	17.75	17.63	17.82	17.66	17.75	17.63	18.24
MSAs	71	71	71	71	71	71	71	71

Notes: The analysis is restricted to 71 non-southern MSAs for which the instrument 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 region by decade fixed effects. All regressions control for MSA total population and are weighted by MSA total population in 1900. F-stat refers to the KP F-stat for weak instruments. Robust standard errors clustered at the MSA level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.7. Effects for non-Europeans

Dep. variable	Married w/ native			Naturalized		
	Mexico (1)	Canada (2)	China (3)	Mexico (4)	Canada (5)	China (6)
Black population	-0.032	0.028**	0.012	-0.012	0.052*	-0.003
	(0.026)	(0.013)	(0.062)	(0.016)	(0.028)	(0.016)
Observations	81,415	686,760	6,252	211,618	1,137,272	99,490
F-stat	66.81	11.05	79.43	73.53	12.20	84.24

Notes: The analysis is restricted to 108 non-southern MSAs for which the instrument could be constructed, and to years 1910 to 1930. Regressions replicate the specification in column 3 of Table 2. In columns 1-3 the sample is restricted to married men. F-stat refers to the KP F-stat for weak instruments. Standard errors clustered at the MSA level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.8. Heterogeneity by national origin

	Northern Europe	UK	Ireland	Western Europe	Southern Europe	Central and Eastern Europe	Germany	Russian Empire
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Married w/ native							
Black population	0.024	0.027***	-0.011*	0.020	0.004*	0.004**	0.009	-0.003
	(0.015)	(0.009)	(0.006)	(0.017)	(0.002)	(0.002)	(0.009)	(0.002)
Observations	640,387	835,193	674,360	278,255	1,528,411	2,004,635	1,295,258	1,292,126
F-stat	27.72	22.85	25.67	18.32	28.44	23.61	23.25	26.68
	Panel B: Naturalized							
Black population	-0.020	0.047***	0.028**	-0.025	0.039*	0.049***	0.024*	0.038
	(0.014)	(0.015)	(0.013)	(0.029)	(0.021)	(0.010)	(0.014)	(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
F-stat	27.09	23.13	26.59	18.41	28.60	24.19	23.06	27.04

Notes: The analysis is restricted to 108 non-southern MSAs for which the instrument could be constructed, and to years 1910 to 1930. All regressions include fixed effects for age, years in the US, MSA, and region by decade, and control for MSA total population. The sample is restricted to married men in Panel A. F-stat refers to the KP F-stat for weak instruments. Robust standard errors clustered at the MSA level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.9. Heterogeneity by national origin – Economic outcomes

	Northern Europe	UK	Ireland	Western Europe	Southern Europe	Central and Eastern Europe	Germany	Russian Empire
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: In manufacture							
Black population	-0.067**	-0.066**	-0.035	-0.058**	-0.029	-0.074	-0.045*	-0.055***
	(0.026)	(0.027)	(0.024)	(0.029)	(0.025)	(0.047)	(0.025)	(0.019)
Observations	1,003,153	1,138,247	1,020,246	393,204	2,563,303	3,102,183	1,585,529	1,988,789
F-stat	26.70	23.47	27.33	18.28	28.23	24.91	23.32	26.79
	Panel B: Difference in native-immigrant log occupational score							
Black population	-0.347	-11.876	-20.327	4.762	-1.231	-5.430	20.115	-10.778
	(14.987)	(11.235)	(14.970)	(19.698)	(8.023)	(9.153)	(13.770)	(11.380)
Observations	1,015,515	1,157,404	1,048,064	399,180	2,578,615	3,121,770	1,621,207	2,003,156
F-stat	26.78	23.44	27.16	18.32	28.24	24.90	23.31	26.76

Notes: The analysis is restricted to 108 non-southern MSAs for which the instrument could be constructed, and to years 1910 to 1930. All regressions include fixed effects for age, years in the US, MSA, and region by decade, and control for MSA total population. The sample is restricted to men aged 15 to 65. F-stat refers to the KP F-stat for weak instruments. Robust standard errors clustered at the MSA level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.10. Public sector employment

Dep. variable	Employed in		
	Public administration	Police	Fire protection
	(1)	(2)	(3)
	Panel A: All immigrant men of working age		
Black population	-0.165	0.005	0.005
	(0.134)	(0.024)	(0.012)
Observations	14,599,200	14,599,200	14,599,200
F-stat	25.07	25.07	25.07
	Panel B: Immigrant men in the labor force		
Black population	-0.130	0.017	0.010
	(0.117)	(0.027)	(0.014)
Observations	13,005,817	13,005,817	13,005,817
F-stat	24.89	24.89	24.89
	Panel C: Immigrant men employed		
Black population	-0.128	0.025	0.008
	(0.126)	(0.027)	(0.013)
Observations	12,958,064	12,958,064	12,958,064
F-stat	25.85	25.85	25.85

Notes: The table presents results for immigrant men living in the 108 non-southern MSAs for which the instrument could be constructed, and to years 1910, 1920, and 1930. The dependent variable is an indicator for individuals employed in the sector indicated in each column heading. Regressions replicate the specification in column 3 of Table 2. F-stat refers to the KP F-stat for weak instruments. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.11. Local labor unions

Dep. Variable	Log(IWW Locals)			
	(1)	(2)	(3)	(4)
Black population	0.075	0.436	0.175	0.209
	(1.177)	(0.529)	(0.746)	(0.672)
Local unions		Per capita	Over 1900 Pop.	Over 1910 Pop.
Observations	216	216	216	216
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, and to years 1910, 1920, and 1930. The dependent variable is the log of chapters of the Industrial Workers of the World (IWW) in the MSA in a decade. Columns 2 to 4 scale the number of chapters by, respectively contemporaneous, 1900, and 1910 MSA population. Data are from Gregory (2015). All regressions control for MSA total population and region by decade fixed effects and are weighted by 1900 population. Results are estimated for the period 1910-1920 only due to data availability. F-stat refers to the KP F-stat for weak instruments. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B 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-born whites' attitudes and immigrant outcomes. Our goal is not to explain the cognitive processes that reduced native-born prejudice towards immigrants after the Great Migration. Instead, building on existing work in psychology, we assume that a cognitive mechanism led native-born whites to re-categorize immigrants as members of the ingroup, and explore the theoretical and empirical implications of that assumption.

Setup

The population consists of two groups, an ingroup and an outgroup. For our purposes we define the ingroup to be the group containing native-born whites of Anglo-Saxon origin, and the outgroup to be everyone else. Prior to the Great Migration, the outgroup 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, 1994; Hagendoorn, 1995), we define h to be the social distance of outgroup members from the group of native-born 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 $h = 0$. Outgroup members are a share $1 - n$ of the population and are distributed on the line $[0, H^{max}]$.

We assume that native-born 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-born whites. This cost depends on *perceived distance* $\lambda(h, \bar{h})$, where $\bar{h}$ is the average distance of the outgroup from the ingroup. $\bar{h}$ can be thought of as the degree of difference between a native-born white and the members of the outgroup he is faced with *on average* in his daily life or in his neighborhood. As such, it captures the degree of native-born whites' 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 outgroup members makes an outsider of any given distance appear more similar to native-born whites. 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 ingroup if the difference in social distance between the immigrant and the native-born whites is smaller than the difference between the average distance of the ingroup and the outgroup ($\bar{h}$).[29]

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

Discrimination. With each interaction, native-born 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-born 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 outgroup. We denote the decision to discriminate or not by $d \in [0, 1]$, so that utility for an ingroup 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 outgroup members. Outgroup members provide assimilation effort in the hope of becoming part of the native-born white ingroup. Membership in the native-born white 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 outgroup 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-born white group is implicitly defined by

$$F = W(\lambda, \underline{e}(h)) \tag{3}$$

[29] A 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, 1974).

outgroup members can decide between exerting effort $\underline{e}$ and enjoying the benefits of assimilation at cost $c(e)$, or avoiding the costs of effort and foregoing the assimilation benefit. An outgroup 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 $c(e(H^{max})) > B$.

Proposition 0. *There exists a threshold h^*, given by $B = (\underline{e}(h^*))$, such that individuals with a distance below h^* optimally choose to provide sufficient assimilation effort to ensure integration in the native-born white group. Those with a distance exceeding h^* provide zero effort and do not become assimilated. Formally,*

$$e^* = \begin{cases} \underline{e}(h), & \text{for } h \leq h^* \\ 0, & \text{for } h > h^* \end{cases} \tag{4}$$

Proof. 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 (3), the implicit function theorem gives $e_h = -\frac{W_h}{W_e} > 0$, so that the minimum effort necessary for assimilation, as well as its associated costs, are increasing in an outgroup member's distance from the native-born white group. □

The appearance of a more distant outgroup

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 outgroup member. While no study measures the precise social distance hierarchy in the early 20th century United States, works 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, 1994; Pettigrew, 1960). Our assumption builds on this empirical literature. To facilitate correspondence between the model and our empirical results, we start by stating the following trivial proposition, which follows directly from the definition of λ.

Proposition 1. *An increase in H^{max} reduces $\lambda(h, \bar{h})$.*

Proof. An increase in H^{max} raises $\bar{h}$ and $\lambda_{\bar{h}} < 0$ by assumption. □

The increase in the average distance of the outgroup as a result of an increase in H^{max} then changes the threshold for acceptance into the group of native whites:

Proposition 2. *An increase in H^{max} increases h^*.*

Proof. Writing 3 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 outgroup 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 outgroup that is more distant than existing ones will necessarily increase the share of existing outgroup members that are allowed membership into the native-born white ingroup.[30] □

The increase in distance of the most distant outgroup member has implications for the optimal level of effort exerted by outgroup 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.

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 native-born whites is decreasing in the average distance of the outgroup. 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 outgroup 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*

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

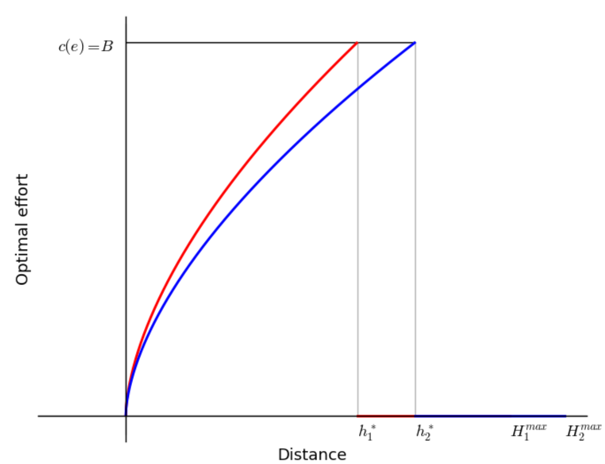
Proof. If $h \leq h_1^*$ an immigrant provides the minimum necessary effort both before and after the increase. Since for any h $\underline{e}$ goes down when H^{max} increases, immigrants who provide $\underline{e}$ decrease their optimal level of provided effort. An immigrant with $h > h_1^*$ has provided zero

[30] The *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 ingroup when African Americans arrive and increase the variance of the outgroup.

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. □

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 native-born. This result is illustrated by Figure B.1. The red line shows optimal effort levels for different values of h before an increase in H^{max}. Individuals who are sufficiently close to native-born whites ($h \leq h_1^*$) 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. The group with $h \leq h_1^*$ was already considered part of the ingroup before the arrival of African Americans. They remain assimilated, but due to lower requirements from the side of the native-born, they are allowed to somewhat decrease their assimilation efforts. The second group, with intermediate levels of h, used to be unable to achieve membership to the ingroup, 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.

Figure B.1. The effects of an increase in H^{max}



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

The testable prediction deriving from Proposition 3 is that there is an inverted U-shaped relationship between immigrant assimilation effort (or successful assimilation) and the distance of immigrants from the native-born white ingroup.

In sum, the 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-born 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 ingroup in significant numbers.

Stochastic assimilation

So far, assimilation was assumed to be a deterministic process. We now generalize the model and include stochastic individual level characteristics that alter an outgroup member's assimilation prospects. Let the disutility that an ingroup member suffers from interacting with an outgroup 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 when effort provision is decided. This may be due to the fact that outgroup members cannot properly predict how the ingroup will respond to their personal characteristics in specific interactions.

The outgroup 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 native-born whites. 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

$$W = \omega_1 \lambda - \omega_2 e$$
$$c = ce^2$$

with ω_1, ω_2 and $c > 0$. Finally we will assume that $\lambda_{h,\overline{h}} = 0$, implying that a change in $\overline{h}$

has the same immediate effect on all outgroup members.[31] 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 $\overline{h}$ on e^*, given by

$$\frac{\partial e^*}{\partial \overline{h}} = \frac{g'(.)B(-\lambda_{\overline{h}})\omega_1\omega_2}{2c - g'(.)B\omega_2^2} \tag{5}$$

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

Equation (5) also shows that the response in effort has an inverted U-shape. The effect of $\overline{h}$ on e^* is increasing in $g'(.)$ and thus maximal for individuals with $\tilde{h}$ such that $F - W(\lambda(\tilde{h}, \overline{h}), e^*(\tilde{h},)) = -\sigma$. 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 ingroup. Assimilation rates of individuals with distance h are given by

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

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

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

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 $F = W(\lambda(\tilde{\tilde{h}}, \overline{h}), e^*(\tilde{\tilde{h}}, \overline{h}))$. 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 $\overline{h}$ 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.

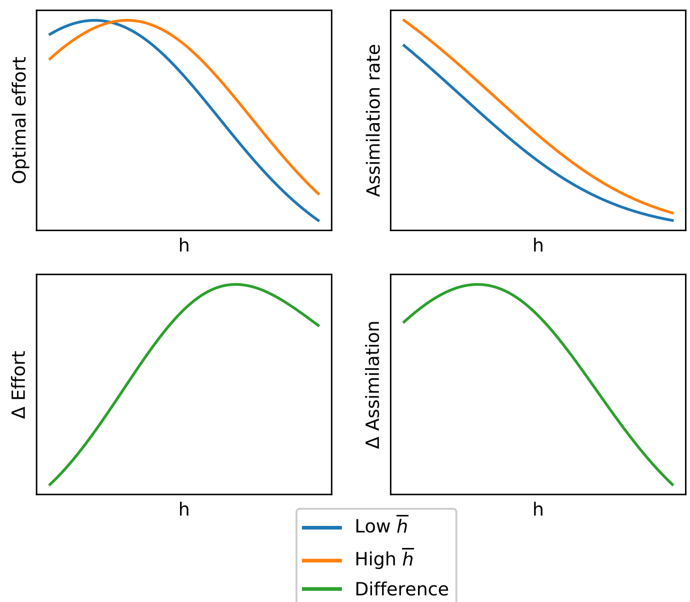
Proposition 4. *The responses of optimal effort and of assimilation rates to an increase in H^{max} are both unimodal with respect to distance h. The peak of assimilation rates occurs at a lower level of h than the peak of optimal effort.*

We illustrate this result with a specific example. We assume the functional form $W = \omega_1 h - \omega_3 \overline{h} - \omega_2 e$, with 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 B.2 plots optimal effort (left panel) and assimilation

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

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 B.2. Response to the inflow of a distant outgroup



Notes: Response of optimal effort (left panel) and assimilation rates (right panel) to an inflow of a distant outgroup (increase in $\overline{h}$, as a function of individual distance h. The response of both measures has an inverted U-shape, but effects on assimilation rates peak at lower values of distance h.

Intuitively, relaxed requirements by the native-born whites allow for the immediate assimilation of the closest (still unassimilated) outgroup 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 native-born whites.

C Data Appendix

C.1 Other measures of assimilation

Manufacturing employment and skill level. European immigrants in the early twentieth century were disproportionately employed in the manufacturing sector and in occupations requiring low levels of skill (see, for example Figure A.2, and the discussion in Abramitzky and Boustan, 2017). We construct an indicator equal to 1 if a foreign-born man was employed in either manufacturing or in unskilled occupations. We interpret a reduction in the probability of working in either sector as evidence of economic assimilation. Consistent with the literature (Abramitzky et al., 2019; Tabellini, 2020), we restrict attention to men in working age (15-64). When defining the unskilled sector, we follow the classification in Katz and Margo (2014) to map occupations to skill levels.

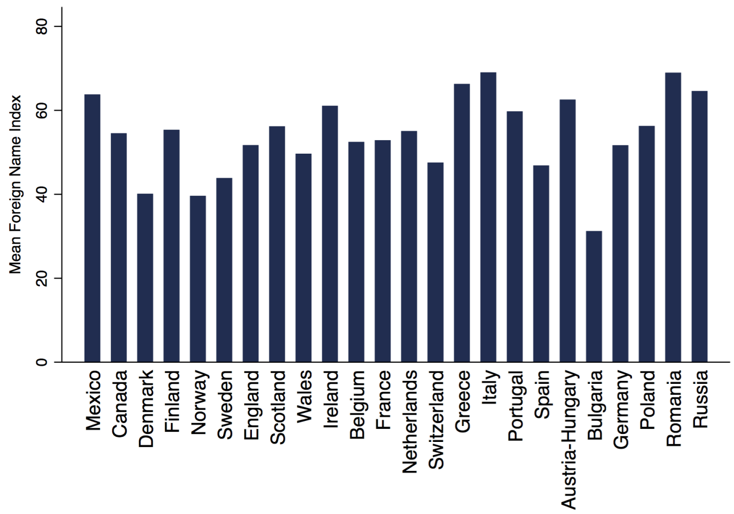
Foreign name index. As an auxiliary measure of assimilation we consider the decision of immigrants to give a foreign-sounding name to their children. Since it involves their offspring and not immigrants themselves, naming captures second-order effects of Black migration on assimilation, and could follow from other assimilation outcomes, such as intermarriage with native-born spouses. Nonetheless, to the degree that parents are attached to their culture, choosing a non-ethnic name for one's children is a costly assimilation decision. Several studies show that there is a labor market penalty associated with foreign-sounding names (Biavaschi et al., 2017; Algan et al., 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 et al., 2017; Carneiro et al., 2016) – 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 was first used by Fryer and Levitt (2004), and more recently by Fouka (2019, 2020) and Abramitzky et al. (2020a).

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 native-born whites). 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 C.1 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 the Russian Empire having some of the most distinctive first names.

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



Notes: 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.

In the empirical analysis, we use the logarithm of the foreign name index as dependent variable to facilitate the interpretation of coefficients in percentage terms. We do not construct the index in the panel dataset, which consists of foreign-born individuals whose names were decided in their country of origin.

Residential segregation. Geographic segregation is a good proxy for the degree of social interactions between immigrants and natives, and a commonly used measure of assimilation (Echenique and Fryer, Jr., 2007; Vigdor, 2010). We calculate residential segregation along ethnic lines building on the method developed by Logan and Parman (2017). This method utilizes the fact that census enumerators traversed around neighborhoods in a sequential fashion. The ordering of respondents (specifically, the relative location of household heads) in the full count US Census manuscript files then allows one to induct the identity of a given household's neighbors. From this, one can calculate a neighbor-based index of segregation along any binary dimension. Differently from Logan and Parman (2017), who construct an index based on racial classification, we use nativity and parentage to define members of the majority and minority group. We define as part of the "majority group" native-born individuals with both native-born parents. Members of the minority group, instead, include first-generation immigrants from European countries.

In what follows, we briefly explain the procedure used to construct the index.[32] Let X_n be the number of households of foreign nationality n (as defined by the household head) with native-born neighbors with native parentage in the MSA. Relying on this observed measure of neighborhood composition, it is possible to compute the expected number that one would obtain under complete integration, $E(\overline{X_n})$, i.e. a situation in which individuals were randomly assigned within neighborhoods and the expected number that one would obtain under complete segregation, $E(\underline{X_n})$, i.e. a situation in which immigrants living next to a native would be only the two individuals on either end of the immigrant neighborhood. The index of residential segregation is computed as

$$Segregation = \frac{E(\overline{X_n}) - X_n}{E(\overline{X_n}) - E(\underline{X_n})}$$

The index is constructed so that higher values indicate more segregated areas. We compute this measure 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 observation in our final dataset is an MSA-nationality-year cell.[33] We calculate two versions of the index: one includes households for which both neighbors are observed on the census schedules, the other uses all households with at least one observed neighbor.

C.2 Linked sample construction

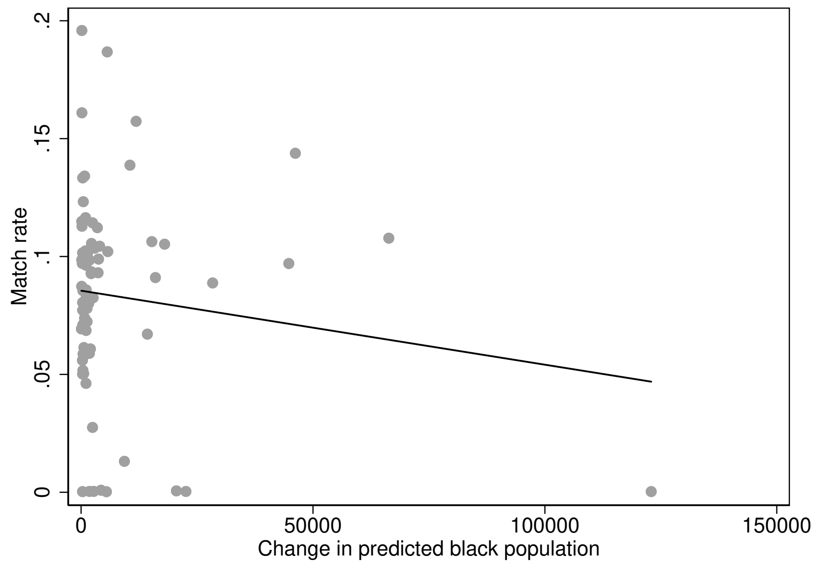
To construct the linked panel dataset, we start with the universe of foreign-born men in 1910 who are unique by first and last name, birthplace and year of birth. Following standard automated census-linking procedures used in the economic history literature (Ferrie, 1996; Abramitzky et al., 2014), we match men to a target census year using the NYSIIS Soundex 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). We discard any records with multiple matches.

Overall, our match rate is 9.4% across three decades. This is comparable to the 12% match rate reported by Abramitzky et al. (2014), who also link immigrants to the US across three census years during a similar time period. The additional discrepancy can be readily explained by the fact that, unlike Abramitzky et al. (2014), our starting dataset is restricted to residents of the 108 Northern MSAs of our sample. Urban residents are younger, and have higher rates of return migration, thus making them harder to successfully link.

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

[33]In choosing these nationalities, we follow Abramitzky et al. (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 the Russian Empire.

Figure C.2. Correlation between match rate and change in predicted Black inflows



Notes: 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.

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 C.2.[34] 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 C.1 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. Finally, the English and Western Europeans are over-represented in the linked panel compared to other immigrant origin regions.

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

Table C.1. Comparison of linked and non-linked records

	Linked	Not-Linked	Difference
Demographic characteristics			
Age	32.716	35.113	-2.397***
	(11.440)	(12.470)	(0.014)
Years in the US	17.686	16.177	1.509***
	(11.295)	(12.951)	(0.025)
Northern Europe	0.039	0.058	-0.019***
	(0.195)	(0.235)	(0.000)
UK	0.067	0.044	0.022***
	(0.249)	(0.206)	(0.000)
Ireland	0.021	0.040	-0.020***
	(0.142)	(0.196)	(0.000)
Western Europe	0.032	0.015	0.017***
	(0.176)	(0.123)	(0.000)
Southern Europe	0.044	0.076	-0.032***
	(0.205)	(0.265)	(0.000)
Central and Eastern Europe	0.032	0.080	-0.048***
	(0.176)	(0.271)	(0.000)
Germany	0.065	0.088	-0.022***
	(0.250)	(0.284)	(0.000)
Russian Empire	0.036	0.064	-0.028***
	(0.185)	(0.245)	(0.000)
Socioeconomic characteristics			
Married with native	0.220	0.224	-0.004***
	(0.414)	(0.417)	(0.001)
Naturalized	0.612	0.491	0.121***
	(0.487)	(0.500)	(0.001)
In manufacture	0.227	0.191	0.036***
	(0.419)	(0.393)	(0.000)
Unskilled	0.311	0.379	-0.068***
	(0.463)	(0.485)	(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.

C.3 Details on instrument construction

Forward survival method. In the absence of data on internal migration prior to 1940, we estimate net migration rates using the forward survival method (Gregory, 2005). Specifically, using data for the United States as a whole, we first compute survival ratios for each age-sex-race group, and then, relying on the latter, we estimate net migration from each Southern state. We compute the net migration rate of African Americans for state s and decade t as

$$\text{Population}_{st} - (\sum_g (\text{Male population}_{gt-1} \times \text{Male survival rate}_{gt} + \text{Female population}_{gt-1} \\ \times \text{Female survival rate}_{gt}) + \text{Population}_{st-1} \times \text{Birth rate}_t)$$

where g denotes a five-year age group. We compute the birth rate and (sex and age group specific) survival rate at the national level using data from the full count of the 1900 to 1930 censuses. For robustness, we compare our measure of estimated out-migration with that computed in Lee et al. (1957). The correlation between the two measures is 0.93. We also validate our estimates, comparing them with those obtained in Boustan (2016).

First stage. Table C.2 verifies the predictive power of the instrument, reporting first stage results for the relationship between the actual and the predicted number of Blacks, as constructed in equation (3).[35] Column 1 controls for total MSA population and includes MSA and decade fixed effects. Column 2 presents our most preferred specification, augmenting the set of controls included in column 1 by interacting decade dummies with region dummies. There is a strong and positive relationship between the instrument and the number of Blacks, and the F-stat for weak instruments is above conventional levels. The coefficient 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 (2019) and in Tabellini (2018) for the same historical period for neighborhoods and cities, respectively. Subsequent columns of Table C.2 explore the robustness of results reported in column 2. In column 3 we show that our estimates are not sensitive to running unweighted regressions. In columns 4 to 7 we augment our baseline specification by interacting decade 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. In all cases, the coefficient remains quantitatively close to that reported in column 2, and its statistical significance is not affected.

[35] All regressions are weighted by 1900 MSA population. Standard errors are clustered at the MSA level.

Table C.2. First stage

Dep. variable	Black population						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Predicted Black population	1.064***	1.155***	0.899***	1.125***	1.101***	1.090***	1.152***
	(0.313)	(0.268)	(0.284)	(0.249)	(0.279)	(0.260)	(0.268)
Observations	324	324	324	324	324	324	324
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 × Decade FE		X	X	X	X	X	X
Weighted	X	X		X	X	X	X
Fr. Black 1900 × Decade FE				X			
Fr. foreign-born 1900 × Decade FE					X		
Log Value manuf. output 1900 × Decade FE						X	
Fr. working age males in manuf. 1900 × Decade FE							X

Notes: The analysis is restricted to 108 non-southern MSAs for which the instrument could be constructed, and to years 1910 to 1930. All regressions control for MSA total population and MSA fixed effects and are weighed by MSA population in 1900. Column 1 includes decade fixed effects. F-stat refers to the KP F-stat for weak instruments. Robust standard errors clustered at the MSA level in parentheses. Significance levels: *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

Push instrument. Even if early Black settlements were as good as randomly assigned, one concern with our baseline IV 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 used in the main text by replacing O_{jt} with predicted (rather than actual) out-migration. In a "zeroth stage", we start by estimating:

$$mig_{sjt} = \alpha_j + \gamma Push_{sjt-1} + e_{sjt} \tag{6}$$

where mig_{sjt} is the net Black migration rate from county s in Southern state j between t and $t-1$. We use beginning of decade county "push factors", $Push_{sjt-1}$, since contemporaneous variables are likely to be themselves affected by out-migration. In our most preferred specification, we include state fixed effects, α_j, and collect in the vector $Push_{sjt-1}$ 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 C.3 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 C.3.

Table C.3. Zeroth stage

Dep. variable	Net Black migration rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Share Black	-0.170***	-0.041	-0.215***	-0.165***	-0.075*	-0.184***
	(0.048)	(0.045)	(0.052)	(0.058)	(0.045)	(0.052)
Rural share	-0.257***	-0.163***	-0.333***	-0.230***	-0.143***	-0.302***
	(0.064)	(0.045)	(0.052)	(0.059)	(0.044)	(0.048)
Share cotton	0.292***	0.295***	-0.105	-0.167	0.291**	-0.233**
	(0.105)	(0.100)	(0.085)	(0.171)	(0.128)	(0.099)
1[Boll Weevil]	-0.034	0.030	-0.052**	-0.205***	0.053	-0.000
	(0.051)	(0.019)	(0.020)	(0.073)	(0.036)	(0.043)
Observations	1,002	989	937	1,002	989	937
State FE				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: *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

In line with the historical evidence (e.g. Boustan, 2016), a higher share of Blacks and a higher fraction of the population living in rural areas are associated with more 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.[36]

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 the baseline instrument to derive the (push-factors induced) predicted number of Blacks moving to city c in decade t. By construction, this predicted measure of Black out-migration 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 patterns between sending and receiving areas – a possible concern for standard shift-share instruments (Jaeger et al., 2018). We show that both first stage and 2SLS results are robust to using this, instead of our baseline version of the instrument.

[36]Results 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.

D Robustness Checks

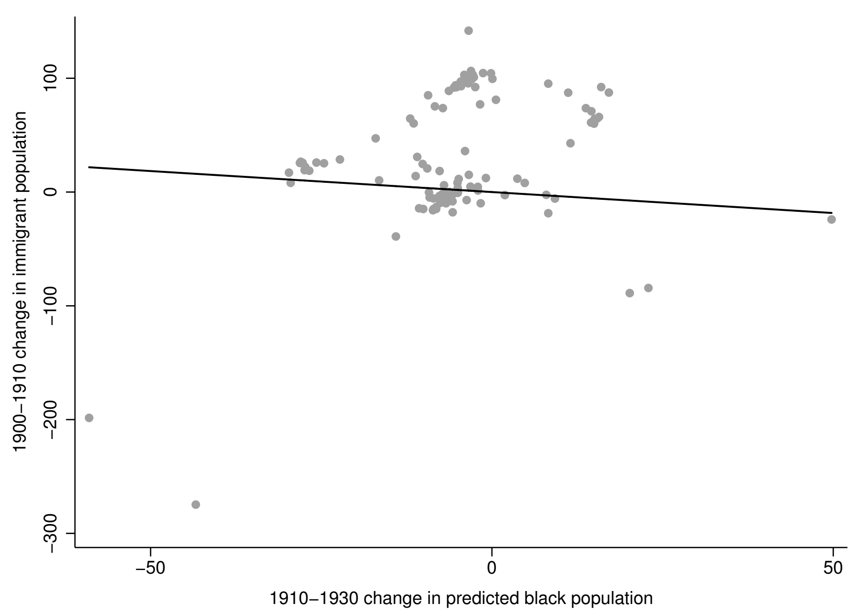
D.1 Addressing threats to identification and instrument validity

D.1.1 Initial shares and local time-varying confounders

The validity of our identification strategy rests on the assumption that the composition (in terms of settlers' Southern states of origin) of 1900 Black settlements is uncorrelated with characteristics of Northern locations that vary within census regions and that may affect immigrant assimilation. It is possible, for instance, that larger and more industrialized urban centers attracted more African Americans from specific Southern states in the 19th century, and that these areas kept growing more in subsequent decades, spurring assimilation of the foreign-born. If, in the same decades, the Southern states that had sent more Black migrants to those same Northern urban centers in 1900 experienced stronger out-migration, a spurious correlation would be introduced between the instrument and immigrant assimilation. 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 D.1). This result is important, as it 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 D.1. Correlation between immigrants' and Blacks' migration patterns



Notes: 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 D.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 2) 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 D.1. These results indicate that, before 1900, Southern-born Blacks did not systematically settle in MSAs that were already undergoing immigrant social or economic assimilation.

Table D.1. Pre-trends

Dep. variable	Married w/ native (1)	Naturalized (2)	Married w/ native (3)	Naturalized (4)
	Baseline specification		Years 1900-1910	
Black population	0.012**	0.033**	-0.024	-0.034
	(0.005)	(0.015)	(0.034)	(0.030)
Observations	9,323,126	15,267,846	2,919,024	4,218,398
F-stat	23.33	24.33	13.83	14.45

Notes: The analysis is restricted to 108 non-southern MSAs for which the instrument could be constructed. 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 2 in the original dataset (years 1910-1930). Columns 3 and 4 replicate the same specification for years 1900-1910. F-stat refers to the KP F-stat for weak instruments. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

We then 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, the share of workers in manufacturing, and shares of immigrants from each origin region (Table D.2). In the first row of Table D.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. 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 systematically correlate with the economic and social assimilation of immigrants. In particular, controlling for the interaction between the 1900 share of Blacks and decade dummies (second row of Table D.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.

Table D.2. Robustness

Dep. variable	Married w/ native (1)	Naturalized (2)
	2SLS coefficient on number of Blacks	
Baseline	0.012***	0.033**
	(0.005)	(0.015)
F-stat	23.33	24.23
Share Blacks 1900 × Decade FE	0.014***	0.032**
	(0.005)	(0.016)
F-stat	24.75	24.83
Share foreign-born 1900 × Decade FE	0.014***	0.046***
	(0.005)	(0.016)
F-stat	19.27	19.92
Log Value manufacturing output 1900 × Decade FE	0.010*	0.033**
	(0.006)	(0.015)
F-stat	22.58	24.23
Share in manufacture 1900 × Decade FE	0.012**	0.031**
	(0.005)	(0.014)
F-stat	23.2	24.34
All 1900 controls × Decade FE	0.014**	0.029*
	(0.006)	(0.017)
F-stat	19.08	19.69
Share origin regions 1900 × Decade FE	0.015***	0.028
	(0.005)	(0.019)
F-stat	22.82	23.29
Predicted industrialization	0.014**	0.046***
	(0.006)	(0.015)
F-stat	17.88	18.58
Observations	9,323,126	15,267,846

Notes: The analysis is restricted to 108 non-southern MSAs for which the instrument could be constructed, and to years 1910 to 1930. The sample consists of foreign-born men. It is further restricted to married individuals in column 1. All regressions replicate the specification reported in column 3 of Table 2 in the main text. F-stat refers to the KP F-stat for weak instruments. Robust standard errors clustered at the MSA level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In the last row of Table D.2, we replicate our baseline results by separately controlling for a measure of predicted industrialization. To construct this variable, we restrict attention to non-southern MSAs. We compute the 1900 share of employment in each 1-digit industry in each MSA, and interact these initial shares with the national growth rate of employment in that industry. Reassuringly, also in this case, our results are largely unchanged.[37]

One concern raised by Amior (2020) is that the enclaves used for the construction of the shift-share instrument may contribute to determining the persistence of labor demand shocks. As a result, historical enclaves may be spuriously correlated with future employment growth. To address this concern, we use the 1900 and 1910 full count US Census to construct the share of men who are, respectively, in the labor force and employed. We create these variables separately for all men of working age (15-64), and for white non-immigrant men, so as to ensure that our baseline measure is not unduly affected by immigrants' or Blacks' employment.[38] In Table D.3, we replicate our baseline specification by adding interactions between decade dummies and the 1900 share of men in the labor force (columns 2 and 3), the 1910 share of men in the labor force (columns 4 and 5) or share of men employed (columns 6 and 7). Reassuringly, results remain always close to the baseline, which is reported in column 1 of Table D.3 to facilitate comparisons.

[37]A similar approach is also used in Sequeira et al. (2020) and in Tabellini (2020). To more precisely proxy for labor demand shocks in non-southern industries, we compute industry national growth rates for the non-South only. Results are unchanged when including the US South to compute national demand growth.

[38]We experimented with both 1900 and 1910 because the 1900 Census does not directly include a question on labor force participation, and altogether lacks a question on employment status. These variables were instead present in the 1910 Census. To construct a proxy for the labor force share in 1900 we considered those men (in the 15-64 age range) who reported an occupational income score, i.e. the proxy for labor income in census years before 1940 (see also Abramitzky et al., 2014), different from zero.

Table D.3. 1900-1910 change in years in the US

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A: Married w/ native						
Black population	0.012**	0.012**	0.011**	0.012**	0.012**	0.013**	0.014**
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
Observations	9,323,126	9,323,126	9,323,126	9,323,126	9,323,126	9,323,126	9,323,126
F-stat	23.33	26.58	27.53	21.73	24.49	20.84	21.76
	Panel B: Naturalized						
Black population	0.033**	0.033**	0.032**	0.031*	0.038**	0.037***	0.033**
	(0.015)	(0.014)	(0.014)	(0.017)	(0.015)	(0.014)	(0.013)
Observations	15,267,846	15,267,846	15,267,846	15,267,846	15,267,846	15,267,846	15,267,846
F-stat	24.23	27.46	28.19	22.84	25.45	21.46	22.72
Group		Everyone	Native white	Everyone	Native white	Everyone	Native white
Year		1900	1900	1910	1910	1910	1910
Employment measure		Share in labor force	Share in labor force	Share in labor force	Share in labor force	Share employed	Share employed

Notes: Column 1 replicates the specification in column 3 of Table 2. All regressions include MSA and region by decade fixed effects, and control for age, years in the US, and nation of origin fixed effects. Columns 2 to 7 augment that specification with interactions between decade dummies and different proxies for employment in the MSA at baseline. 1900 share in the labor force is constructed as the number of men 15-64 reporting a gainful occupation, divided by all men 15-64 in the MSA. In 1910, both labor force and employment status were reported in the full count Census. Thus, the share in labor force (resp. employed) is constructed (for each group) as the number of men 15-64 in the labor force (resp. employed) over all men 15-64. F-stat refers to the KP F-stat for weak instruments. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

Next, we augment our baseline specification by separately controlling for the number of European immigrants in the MSA in each decade. This check is important to rule out the possibility that both 1900 Northern settlements of African Americans and patterns of Black emigration from the South were spuriously correlated with those of specific European immigrant groups.[39] To do so, we first predict the number of European immigrants following the same logic as for the instrument used for Black population (see equation 3). Then, we re-estimate equation (1) with 2SLS, instrumenting the actual number of European immigrants with its predicted version. To account for the possibility that the patterns of European immigration as well as their effects on immigrant assimilation differed between "old" and "new" sending countries (Abramitzky and Boustan, 2017), we construct two separate instruments – one for each group.[40]

Table D.4 reports the results, focusing on intermarriage in Panel A and on naturalization in Panel B. Column 1 presents our baseline specification (column 3 of Table 2). Columns 2 and 3 add, respectively, the instrumented number of European immigrants from "new" and "old" sending countries, whereas column 4 simultaneously controls for both. Columns 2 to 4 also report the AP F-stats for weak instruments in each separate first stage – the relevant F-statistics in the case of multiple endogenous regressors and multiple instruments (Angrist and Pischke, 2008). In all cases, the AP F-stats are above conventional levels, indicating that all instruments are strong. Reassuringly, the point estimate on Black population remains close to that estimated in our baseline specification, and, if anything, tends to become larger when controlling for European immigration, especially from new sending countries. When simultaneously controlling for both immigrant groups, in the case of naturalization, the coefficient becomes slightly smaller and is no longer statistically significant at conventional levels (column 4, Panel B). However, it remains quantitatively similar to – and is not statistically different from – the coefficient reported in column 1 of Panel B.

More broadly, Goldsmith-Pinkham et al. (2020) show that shift-share instruments can be expressed in terms of the "initial shares" (i.e. the enclaves of Black individuals born in different Southern states and living across Northern MSAs in 1900, or the α_{jn}^{1900} terms in equation 3), and that the exogeneity condition can be stated in terms of the shares. Following the approach suggested in Goldsmith-Pinkham et al. (2020), we compute the Rotemberg weights associated to each of the 15 Southern states considered in our sample.[41] The largest of these weights (for

[39]For instance, one may be worried that a large share of Blacks born in Alabama settled in Chicago before 1900 – where a large German enclave existed (Abramitzky and Boustan, 2017) – and that emigration from Alabama was particularly strong in the 1920s, when also immigration from Germany increased significantly (Tabellini, 2020). These or similar situations might create a spurious correlation between our instrument for Black population and the inflow of European immigrants.

[40]We follow Tabellini (2020), and classify as "old" immigrants from: the UK; Northern and Western Europe; Ireland; and Germany. We classify as "new" immigrants from: Southern Europe, the Russian Empire, and Central-Eastern Europe.

[41]In our context, the Rotemberg weights are obtained from the initial shares of African Americans from each Southern state.

Table D.4. Controlling for European immigrants

	(1)	(2)	(3)	(4)
	Panel A: Married w/ native			
Black population	0.012**	0.020***	0.012**	0.026**
	(0.005)	(0.005)	(0.005)	(0.010)
European immigrants (New origins)		0.010***		0.020*
		(0.004)		(0.011)
European immigrants (Old origins)			-0.001	0.008
			(0.002)	(0.005)
Observations	9,323,126	9,323,126	9,323,126	9,323,126
KP F-stat	23.33	4.503	11.22	1.247
AP F-stat (Blacks)		30.3	634.9	180.1
AP F-stat (new)		121.9		133.8
AP F-stat (old)			825.4	410.7
	Panel B: Naturalized			
Black population	0.033**	0.039**	0.035**	0.027
	(0.015)	(0.019)	(0.014)	(0.016)
European immigrants (New origins)		0.008		-0.014
		(0.011)		(0.018)
European immigrants (Old origins)			-0.012***	-0.019**
			(0.003)	(0.008)
Observations	15,267,846	15,267,846	15,267,846	15,267,846
KP F-stat	24.23	4.366	11.71	1.308
AP F-stat (Blacks)		31.33	733.5	220.3
AP F-stat (new)		105.3		114.3
AP F-stat (old)			832.2	419.0

Notes: This table replicates the results reported in column 3 of Table 2 in the main text by separately controlling for the (instrumented) number of European immigrants in the MSA in each decade. Column 1 replicates the baseline specification of Table 2. Columns 2 to 3 include, respectively, the number of European immigrants from "new" and "old" sending regions. Column 4 simultaneously includes both. We follow Tabellini (2020), and classify as "old" immigrants from: the UK; Northern and Western Europe; Ireland; and Germany. We classify as "new" immigrants from: Southern Europe, the Russian Empire, and Central-Eastern Europe. KP F-stat is the Kleibergen-Paap Wald rk statistics for weak instruments, and AP F-stats refer to the robust F-stat for the Angrist and Pischke weak identification test for each individual endogenous regressor. F-stat refers to the KP F-stat for weak instruments. Standard errors clustered at the MSA level in parentheses. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

the state of Mississippi) is equal to 0.257. For the top 5 Southern states in terms of Rotemberg weights (Mississippi, Alabama, Georgia, Tennessee, and Louisiana), we report the correlation between the 1900 Black share from each state (i.e., the α_{jn}^{1900} terms in equation 3) and the ratio of immigrants to native whites in 1910. Reassuringly, the correlation between these two variables is always low, and never statistically significant, suggesting that the instrument is not relying more heavily on variation coming from MSAs with systematically high or low immigrant presence (Table D.5). This is important because high immigrant presence may trigger native backlash, in turn reducing the prospects of immigrants' assimilation (Dustmann et al., 2019; Tabellini, 2020).

Table D.5. Initial Black shares and immigrant to native white ratio

Dep. variable	Ratio immigrants to native-born whites (1910)				
	(1)	(2)	(3)	(4)	(5)
1900 share Blacks from State	0.068	0.398	-0.021	0.291	0.169
	(0.275)	(0.345)	(1.441)	(0.441)	(0.341)
Observations	108	108	108	108	108
Rotemberg weight	0.257	0.253	0.166	0.124	0.099
State	MS	AL	GA	TN	LA

Notes: The analysis is restricted to the 108 non-southern MSAs for which the instrument could be constructed, and to years 1910 to 1930. The dependent variable is the ratio of immigrants to native-born whites in the MSA in 1910. It is regressed against the share of Blacks living in the MSA as of 1900 and born in the Southern state indicated at the bottom of the table, for each of the 5 Southern states corresponding to the top 5 Rotemberg weights. All regressions are weighted by 1900 MSA total population, and control for MSA population and for region fixed effects. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

We then examine the relationship between the top 5 Rotemberg weights and three additional key MSA-level characteristics: the 1900 fraction of Blacks; the 1900 fraction of immigrants; and, the 1900 share of manufacturing employment.[42] We report results for this exercise in Table D.6. Reassuringly, only for the case of Mississippi is there a statistically significant correlation between the Rotemberg weights and the 1900 fraction of Blacks. In all other cases, the 1900 fraction of Blacks is positively associated with the Rotemberg weights, but never statistically significant at conventional levels. As shown in Table D.2, our results are robust to controlling for the 1900 fraction of Blacks interacted with decade dummies, mitigating any additional concerns resulting from the imprecise positive correlation. In the case of the fraction of immigrants and of the manufacturing share, coefficients are unstable and far from significance throughout. Notably, when considering the variation in the data of the

[42]Manufacturing was both the most important sector of employment for European immigrants in urban areas (Abramitzky and Boustan, 2017; Tabellini, 2020), and one of the main "pull" forces for African Americans during the Great Migration (Boustan, 2016).

share of Blacks from different states, it becomes obvious that the magnitude of the estimates in Table D.6 is fairly small. For instance, the average share of Blacks from Mississippi (resp. Alabama and Georgia) is as low as 0.006 (resp. 0.005 and 0.007).

Table D.6. Correlation between initial Black shares and 1900 MSA characteristics

Dep. variable	1900 MSA Characteristics				
	(1)	(2)	(3)	(4)	(5)
	Panel A: Share Black				
1900 share Blacks from State	0.120**	0.078	0.064	0.170	0.121
	(0.051)	(0.087)	(0.384)	(0.107)	(0.073)
	Panel B: Share immigrant				
1900 share Blacks from State	0.064	0.252	-0.135	0.188	0.127
	(0.167)	(0.192)	(0.869)	(0.256)	(0.202)
	Panel C: Share manufacturing employment				
1900 share Blacks from State	-0.039	-0.244	-0.730	-0.234	-0.124
	(0.171)	(0.206)	(0.764)	(0.263)	(0.213)
Observations	108	108	108	108	108
Rotemberg weight	0.257	0.253	0.166	0.124	0.099
State	MS	AL	GA	TN	LA

Notes: The analysis is restricted to the 108 non-southern MSAs for which the instrument could be constructed, and to years 1910 to 1930. Dependent variables indicated in panel titles are regressed against the share of Blacks living in the MSA as of 1900 and born in the state indicated at the bottom of the table, for each of the 5 Southern states corresponding to the top 5 Rotemberg weights. All regressions are weighted by 1900 MSA population, and control for MSA total population and for region fixed effects. F-stat refers to the KP F-stat for weak instruments. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

D.1.2 Accounting for the impact of the immigration quotas and WWI

During the period considered in our analysis, two major shocks affected immigrants in the United States. The first one was the introduction of the immigration restrictions of the 1920s (Abramitzky and Boustan, 2017; Goldin, 1994), which drastically reduced the number of immigrants moving to America. The second was the eruption of World War I (WWI), which both temporarily halted European immigration and increased animosity against specific immigrant groups like the Germans (Moser, 2012; Fouka, 2019, 2020; Ferrara and Fishback, 2020). Although there is no ex-ante reason to expect the local impact of these shocks to be correlated with our instrument, we now formally test that there is no relationship between either of them and the predicted change in Black population across the MSAs in our sample.

Immigration quotas. We start from the concern 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 enclaves 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

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

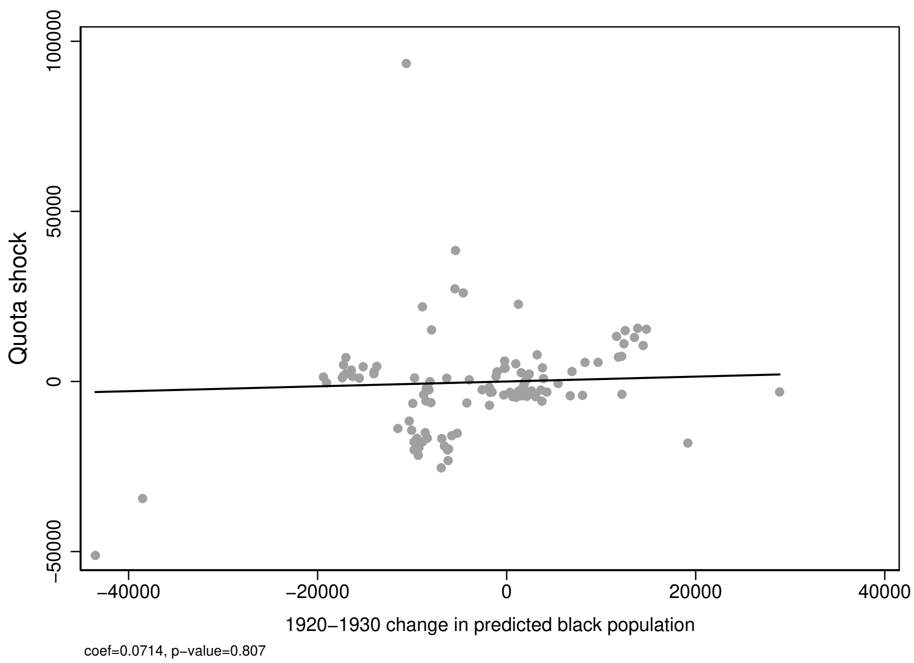
the share of immigrants (relative to 1920 MSA population) "lost" by each MSA between 1921 and 1930 due to the quotas. In particular, $\hat{M}_{k,22-30}$ is the predicted number of immigrants from k that would have entered the US, had the quota system not been introduced. As in Ager and Hansen (2017), $\hat{M}_{k,22-30}$ 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.[43] $Q_{k,22-30}$ is the total number of immigrants from k that were allowed to enter the US according to the yearly quotas. 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 $\left(\hat{M}_{k,22-30} - Q_{k,22-30}\right)$ 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

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

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 D.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.

Figure D.2. Correlation of quota shock with instrument



Notes: 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 1920 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 column 1 of Table D.7 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 D.7 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.

In Table D.8, we verify that also the 1900-1910 change in years spent in the US was uncorrelated with the 1910-1930 instrumented change in Black population. Column 1 pools all immigrant origin groups and shows that there are no pre-trends. The point estimate is not only not statistically significant, but also quantitatively small. As in Table D.7, in columns 2 to 9 we consider each region of origin separately. With the exception of immigrants from the Russian Empire (column 2), no pattern of correlation emerges between Black in-migration after 1910 and the pre-1910 change in immigrants' length of stay.

Table D.7. Effect on average years in the US

Dep. variable	Average years in the US								
	All	Northern Europe	UK	Ireland	Western Europe	Southern Europe	Central & Eastern Europe	Germany	Russian Empire
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Black population	0.097	-0.411	1.672*	0.330	0.783	-0.251	-0.048	-0.648	-0.062
	(0.406)	(0.447)	(0.903)	(0.671)	(0.552)	(0.179)	(0.522)	(0.683)	(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
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 instrument could be constructed, and to years 1910 to 1930. Column 1 replicates the specification of column 3, Table 2, without indicators for years in the US. Columns 2-9 replicate the specification in column 3 of Table 2 separately for each region of origin. F-stat refers to the KP F-stat for weak instruments. Robust standard errors clustered at the MSA level in parentheses. Significance levels: *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

Table D.8. 1900-1910 change in years in the US

Dep. variable	1910-1900 Change in the number of years in the US								
	All (1)	Northern Europe (2)	UK (3)	Ireland (4)	Western Europe (5)	Southern Europe (6)	Central & Eastern Europe (7)	Germany (8)	Russian Empire (9)
1930-1910 Change	-0.199	1.374	2.468	0.258	2.930	0.767	0.583	0.473	2.647*
Black Population	(1.188)	(1.191)	(1.597)	(0.605)	(2.103)	(1.061)	(1.656)	(0.393)	(1.396)
Observations	4,376,705	4,376,705	4,376,705	4,376,705	4,376,705	4,376,705	4,376,705	4,376,705	4,376,705
F-stat	13.96	13.96	13.96	13.96	13.96	13.96	13.96	13.96	13.96
1910 Mean	16.16	17.27	22.30	26.60	19.69	8.674	9.389	24.98	9.408
1900 Mean	18.13	15.27	21.74	25.04	19.61	9.752	11.40	21.88	10.59

Notes: The analysis is restricted to the immigrants living in 1910 in the 108 non-southern MSAs for which the instrument could be constructed. The dependent variable is the 1910-1900 change in the average length of stay in the US. We estimate individual level regressions, even though the outcome is at the MSA level in order to keep the weighting scheme consistent with that of other tables in the paper. Column 1 considers the 1910-1900 change in the length of stay for all immigrants. Columns 2 to 9 construct the same variable focusing on each immigrant group reported at the top of the column. All regressions control for region fixed effects and include fixed effects for nationality and age. F-stat refers to the KP F-stat for weak instruments. Robust standard errors clustered at the MSA level in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

World War I. We now turn to the potentially confounding effects of WWI. There are two potential concerns. First, that the war differentially impacted Northern MSAs, for instance through its effects on military enlistment or casualties, in a way that is spuriously correlated with our instrument. Such differential impact could, in turn, have affected immigrant assimilation. Second, that it affected immigrant assimilation directly, through changed attitudes towards specific groups (e.g. German immigrants) and that such changes were correlated with predicted Black migration.

We address these concerns in a variety of ways. First, we rely on a a specific question in the 1930 Census that asked individuals whether they served in WWI. Restricting attention to immigrant men who, given their age in 1930, were eligible to serve in WWI (i.e. those who were between 31 and 58 in 1930), we create a dummy equal to 1 for WWI veterans.[44] In Table D.9 we estimate individual (column 1) and MSA (column 2) level regressions, to test the relationship between immigrant WWI veteran status and the instrumented change in Black population between 1910 and 1930.[45] Reassuringly, the coefficient on the change in the Black population is small and never statistically significant, indicating that there is no relationship between Black in-migration and immigrants' WWI veteran status. Figure D.3 confirms these patterns, presenting visually the estimates reported in column 2 of Table D.9. In columns 3 and 4, we show that our main results of column 3 in Table 2 remain unchanged when including the interaction between decade dummies and the MSA WWI veteran share among immigrants.

These results weigh against the possibility that the instrument is spuriously correlated with the local impact of WWI. However, two caveats should be highlighted. First, WWI participation may have been an outcome of Black in-migration, if enlistment is interpreted as a measure of assimilation into American society. Yet, since the first Great Migration took off only after 1915 (Boustan, 2016), i.e. precisely with WWI, this possibility seems unlikely. Second, the analysis described above rests on the assumption that the Great Migration was not systematically associated with selective internal migration of WWI veteran immigrants across MSAs in our sample. We acknowledge that this might be a possibility. Yet, our analyses provide no reason to believe that this mechanism was operating in practice. As we demonstrate in Section D.3 below, the inflow of Black migrants was not associated with any change either in the immigrant share (relative to natives) or in the composition of immigrants across MSAs.

[44] We define the age range following Campante and Yanagizawa-Drott (2015) and Qian and Tabellini (2020).

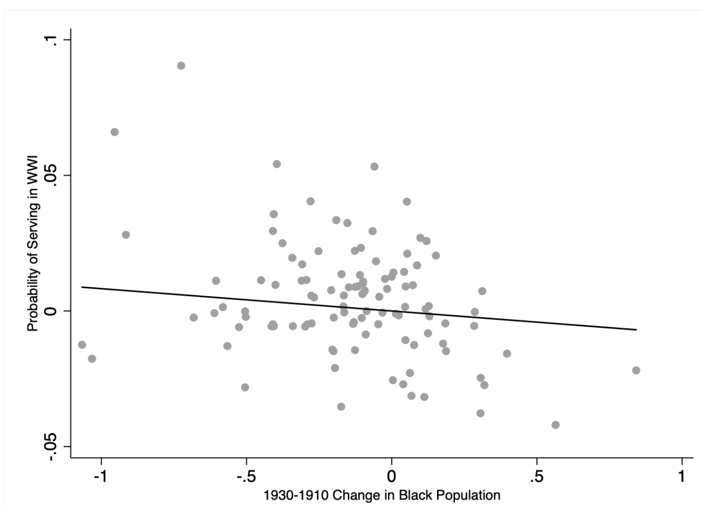
[45] We always control for the change in MSA total population and for census region fixed effects. Individual level regressions also include fixed effects for age, years in the US, and nation of origin. The MSA level regression is weighed by the number of observations in the individual level dataset. Equivalent results are obtained when either estimating unweighted regressions or when weighing by MSA population. Standard errors are always clustered at the MSA level.

Table D.9. Black migration and WWI immigrant veterans

Dep. variable	1[WWI Veteran]		Married w/ native	Naturalized
	(1)	(2)	(3)	(4)
Δ Black population (1930-1910)	-0.002	-0.008		
	(0.004)	(0.008)		
Black population			0.014**	0.037**
			(0.006)	(0.016)
Observations	3,499,511	108	9,323,126	15,267,846
F-stat	17.05	26.65	27.43	29.40
Specification	Individual level	MSA level	Individual level	Individual level

Notes: In columns 1 and 2 the dependent variable is a dummy equal to one for veteran immigrant men living in an MSA in 1930. The sample is restricted to men who, given their age in 1930, were eligible to serve in WWI (i.e. those who were between 31 and 58 in 1930). The MSA-level regression in column 2 is weighted by the number of observations in column 1. Columns 3 and 4 replicate the specification in column 3 of Table 2 including interactions between decade dummies and the 1930 share of WWI immigrant veterans in the MSA (i.e. the dep. variable in columns 1 and 2). Regressions in columns 1-2 include region fixed effects. Individual-level regressions include indicators for age, years in the US, and nation of origin dummies. F-stat refers to the KP F-stat for weak instruments. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

Figure D.3. Black migration and share of WWI immigrant veterans



Notes: The figure represents the residual scatterplot from a regression of the share of WWI immigrant veterans (y-axis) against the (instrumented) 1910 to 1930 change in Black population (x-axis), after partialling out region dummies and changes in total MSA population. The corresponding estimates are reported in Table D.9.

To partly address the two concerns just described as well as the potential issue of selective mortality during the war, we complement the previous exercise by relying on data on casualties and enlistment rates constructed by Ferrara and Fishback (2020) using historical sources.[46] Tables D.10 and D.11 estimate MSA-level regressions using the enlistment and the casualty rate as dependent variables, respectively. In columns 1 and 2 (resp. 3 and 4) of Table D.10 the denominator for the enlistment rate is constructed by focusing on men who were in the age range 17-45 and 18-45 in the 1920 (resp. 1910) full count US Census. In Table D.11, we follow the same structure for columns 2 to 5, and also present results using the casualty rate among those enlisted in column 1. In all cases, Black in-migration is uncorrelated with both enlistment and casualty rates, regardless of how the latter are constructed.

Table D.10. Black migration and WWI enlistment rate

Dep. variable	Share enlisted			
	(1)	(2)	(3)	(4)
Δ Black pop. (1930-1910)	1.117	1.148	-2.911	-2.997
	(0.935)	(0.963)	(3.223)	(3.315)
Observations	108	108	108	108
F-stat	9.536	9.536	9.536	9.536
Mean dep. variable	11.21	11.60	12.65	12.25
Sample	Men 17-45	Men 18-45	Men 17-45	Men 18-45
Census year	1920	1920	1910	1910

Notes: The dependent variable is the enlistment rate computed as the number of individuals enlisted in the MSA according to the Ferrara and Fishback (2020) data divided by: i) the number of men who, by the 1920 Census, were 17-45 (resp. 18-45) in 1917 for column 1 (resp. column 2); ii) the number of men who, by to the 1910 Census, were 17-45 (resp. 18-45) in 1917 for column 3 (resp. column 4). The age restriction used to construct the denominator is imposed in order to restrict attention to men who were eligible to serve in WWI (see Campante and Yanagizawa-Drott, 2015, for more details). All regressions include the 1910 to 1930 change in MSA population and region fixed effects. The 1910 to 1930 change in Black population is instrumented with the corresponding change in the instrument. All regressions are weighed by 1900 MSA total population. F-stat refers to the KP F-stat for weak instruments. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

[46] We thank the authors for kindly sharing the data with us.

Table D.11. Black migration and WWI casualty rate

Dep. variable	Casualty rate				
	(1)	(2)	(3)	(4)	(5)
Δ Black pop. (1930-1910)	-1.043	-0.015	-0.016	-0.053	-0.055
	(1.410)	(0.080)	(0.082)	(0.103)	(0.106)
Observations	108	108	108	108	108
F-stat	9.536	9.362	9.381	9.252	9.242
Mean dep. variable	2.989	0.295	0.305	0.316	0.327
Sample		Men 17-45	Men 18-45	Men 17-45	Men 18-45
Census Year		1920	1920	1910	1910

Notes: The dependent variable is the death rate computed as the number of individuals who died during WWI in the MSA according to the Ferrara and Fishback (2020) data divided by the number of men who were enlisted (from Ferrara and Fishback, 2020) in column 1. In columns 2 to 5, we scale the number of casualties in the MSA by: i) the number of men who, by the 1920 Census, were 17-45 (resp. 18-45) in 1917 for column 2 (resp. column 3); ii) the number of men who, by the 1910 Census, were 17-45 (resp. 18-45) in 1917 for column 4 (resp. column 5). Columns 2 to 5 also include the corresponding enlistment rate. The age restriction used to construct the denominator is imposed in order to restrict attention to men who were eligible to serve in WWI (see Campante and Yanagizawa-Drott, 2015, for more details). All regressions include the 1910 to 1930 change in MSA population and region fixed effects. The 1910 to 1930 change in Black population is instrumented with the corresponding change in the instrument. All regressions are weighted by 1900 MSA total population. F-stat refers to the KP F-stat for weak instruments. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

Table D.12. Effects after omitting Germans

Dep. variable	Married w/ native		Naturalized	
	(1)	(2)	(3)	(4)
Black population	0.012***	0.012**	0.037**	0.033**
	(0.005)	(0.005)	(0.016)	(0.015)
Observations	8,027,867	9,323,126	13,343,913	15,267,846
F-stat	23.56	23.33	24.50	24.23
Omit Germans	X		X	

Notes: Columns 2 and 4 report results from column 3 of Table 2. Columns 1 and 3 replicate the same specification omitting Germans from the sample. F-stat refers to the KP F-stat for weak instruments. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

As documented in existing research, WWI triggered a substantial increase in anti-German sentiment (Moser, 2012; Fouka, 2019, 2020; Ferrara and Fishback, 2020). If the presence of Germans – established or new migrants – is correlated with predicted Black in-migration, WWI may spuriously influence our results. Table D.12 shows that our estimates are unaffected when omitting German immigrants from the data. Next, we exploit the linked sample to test whether Black in-migration was associated with selective attrition among Germans, or other immigrants from countries involved in WWI. Table D.13 presents results from individual level regressions restricted to our sample of immigrant men in 1910, where the dependent variable is an indicator for individuals linked across all three census decades (1910 to 1930). This is regressed against an indicator for German immigrants (resp. Italian and Austrian-Hungarian) in column 1 (resp. columns 2 and 3) and its interaction with the 1910-1930 instrumented change in Black population. All regressions control for MSA fixed effects (absorbing the main effect of Black migration).

Table D.13. Testing for differential attrition

Dep. variable	1[Matched]		
	(1)	(2)	(3)
German	0.003		
	(0.003)		
German × Δ Black pop. (1930-1910)	0.013		
	(0.009)		
Italian		0.014***	
		(0.003)	
Italian × Δ Black pop. (1930-1910)		-0.004	
		(0.009)	
Austrian/Hungarian			-0.053***
			(0.002)
Austrian/Hungarian X × Δ Black pop. (1930-1910)			-0.000
			(0.016)
Observations	3,507,147	3,507,147	3,507,147
F-stat	6.065	6.083	5.564

Notes: The sample consists of all foreign-born men living in the 108 sample MSAs in 1910. The dependent variable is an indicator for immigrant men linked across all three decades. All regressions control for MSA fixed effects and for the interaction between each nationality and the change in MSA population between 1910 and 1930. F-stat refers to the KP F-stat for weak instruments. Robust standard errors, clustered at the MSA level, in parenthesis. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

There is no indication that the inflow of African Americans was associated with differential attrition of immigrants from enemy countries across MSAs in our sample. In Section D.3 below, using the cross-sectional dataset, we further verify that Black in-migration was uncorrelated with the change in the share of German immigrants (as well as the share of all other origin groups).

D.1.3 Other threats to instrument validity

Complementing the analysis in Goldsmith-Pinkham et al. (2020), which focuses on the exogeneity of shock exposure (or "initial shares"), Borusyak et al. (2018) show that identification in shift-share designs can be expressed in terms of the exogeneity of shocks (or "shifts"). In our context, it is possible that the shift component of the instrument, out-migration from each Southern state, is not independent of cross-MSA pull factors systematically related to 1900 settlers' state of origin. To address this concern, we construct an alternative version of the instrument. As described in Section C.3 above, we follow Boustan (2010), and rely on (Southern) county-level push factors to predict net Black migration rates from each Southern state. Table D.14 demonstrates the robustness of the first stage using this version of the instrument. In Table D.15 we replicate the results of Table 2 using the push instrument. This exercise leaves both the magnitude and significance of coefficients practically unchanged.

Results in Table D.15 reduce concerns that Southern out-migration may be correlated with shocks that simultaneously influence the conditions of Northern labor markets. Moreover, they weigh against the possibility that our 2SLS estimates may conflate the long and the short run effects of Black migration, due to serial correlation – a potential concern with shift-share instruments applied to recent data (Jaeger et al., 2018). To even more directly tackle the potential issue of serial correlation, we implement the procedure described in Jaeger et al. (2018), instrumenting both actual and lagged Black population.

We report results of this analysis (which, by necessity omits 1910) in Table D.16. For consistency, in columns 1 and 4 we replicate the baseline specification (Table 2, column 3) focusing on 1920 and 1930 only. Specifications in the rest of the table control for contemporaneous and lagged MSA population respectively. Our results go through. Coefficients for both intermarriage and naturalization become larger in magnitude, though the latter estimate loses in significance when controlling for lagged, rather than contemporaneous population (column 6).

Table D.14. First stage with push instrument

Dep. variable	Black population						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Predicted Black population	0.553**	0.781***	0.717**	0.710***	0.808***	0.659***	0.778***
	(0.266)	(0.252)	(0.284)	(0.233)	(0.250)	(0.202)	(0.256)
Observations	324	324	324	324	324	324	324
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 × Decade FE		X	X	X	X	X	X
Weighted	X	X		X	X	X	X
Fr. Black 1900 × Decade FE				X			
Fr. foreign-born 1900 × Decade FE					X		
Log Value manuf. output 1900 × Decade FE						X	
Fr. working age males in manuf. 1900 × Decade FE							X

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

Table D.15. Main results with push instrument

	(1) OLS	(2) OLS	(3) 2SLS	(4) 2SLS	(5) 2SLS	(6) 2SLS
	Panel A: Married w/ native					
Black population	0.007	0.005	0.013**	0.016***	0.014***	0.025
	(0.006)	(0.004)	(0.005)	(0.004)	(0.004)	(0.024)
Observations	9,323,126	9,323,126	9,323,126	9,323,109	9,323,109	88,892
F-stat			32.22	32.42	32.80	37.25
	Panel B: Naturalized					
Black population	0.056**	0.030*	0.030*	0.038***	0.034**	0.190***
	(0.025)	(0.016)	(0.016)	(0.014)	(0.014)	(0.036)
Observations	15,267,846	15,267,846	15,267,846	15,267,844	15,267,844	80,866
F-stat			33.39	33.41	33.83	35.45
Individual controls	X	X	X	X	X	
Region × Decade		X	X	X	X	
MSA × Origin region				X	X	
Origin region × Decade					X	
Linked sample						X

Notes: The table replicates Table 2, instrumenting actual Black population with Black population predicted by Southern push factors. F-stat refers to the KP F-stat for weak instruments. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

Table D.16. Correcting for potential serial correlation

Dep. variable	Married w/ native			Naturalized		
	(1)	(2)	(3)	(4)	(5)	(6)
Black population	0.026***	0.035**	0.065***	0.034**	0.102**	0.079
	(0.010)	(0.016)	(0.013)	(0.014)	(0.051)	(0.062)
Lagged Black population		-0.014	-0.040**		-0.113*	-0.093
		(0.015)	(0.015)		(0.059)	(0.064)
Observations	6,787,261	6,787,261	6,787,261	10,896,459	10,896,459	10,896,459
KP F-stat	11.49	3.856	8.855	11.63	4.006	10.33
AP F-stat contemp.		7.62	30.53		7.35	34.10
AP F-stat lagged		19.92	47.80		19.83	47.27
Lagged MSA pop.			X			X

Notes: The analysis is restricted to the 108 non-southern MSAs for which the instrument could be constructed, and to years 1910 to 1930. Columns 1 and 4 replicate the specification in column 3 of Table 2 restricting attention to 1920-1930. Columns 2-3 and 4-5 replicate column 3 of Table 2 controlling for instrumented lagged Black population. Columns 3 and 6 replace contemporaneous with lagged MSA population. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Finally, exploiting the fact that the point estimates from the baseline and the "push" version of the shift-share instrument are somewhat different, we follow Derenoncourt (2019) and implement an over-identification test as proposed in Goldsmith-Pinkham et al. (2020). We report results in Table D.17, presenting 2SLS estimates from the baseline and the "push" instruments separately in columns 1-2 and 4-5. Columns 3 and 6 display 2SLS estimates obtained when instrumenting Black population with both instruments simultaneously. Results in these two columns show that *i*) the F-stat is large; *ii*) second stage coefficients remain positive and statistically significant; and, *iii*) the Hansen J statistic p-value is above 0.5 in both cases, implying that we cannot reject the null of over-identifying restrictions.

Table D.17. Over-identification test

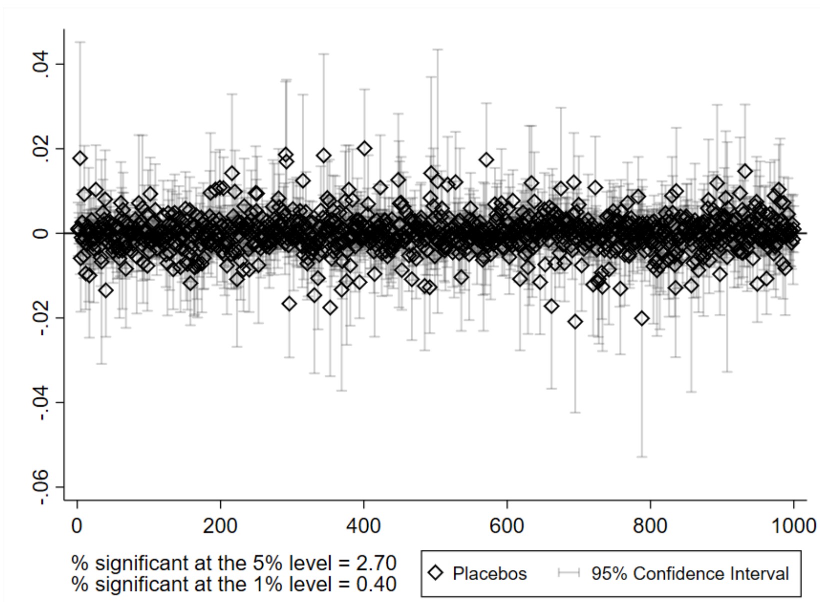
Dep. variable	Married w/ native			Naturalized		
	(1)	(2)	(3)	(4)	(5)	(6)
Black population	0.012**	0.013**	0.011**	0.033**	0.030*	0.035**
	(0.005)	(0.005)	(0.005)	(0.015)	(0.016)	(0.016)
Observations	9,323,126	9,323,126	9,323,126	15,267,846	15,267,846	15,267,846
Instrument(s)	Baseline	Push IV	Both IVs	Baseline	Push IV	Both IVs
F-stat	23.33	32.22	13.18	24.23	33.39	13.25
P-value			0.505			0.651

Notes: The analysis is restricted to the 108 non-southern MSAs for which the instrument could be constructed, ano to years 1910 to 1930. Columns 1 and 4 report 2SLS estimates from column 3 of Table 2. The remaining columns replicate this specification by varying the instruments used, as indicated at the bottom of the table. P-value is the Hansen J statistic p-value for the test of over-identified restrictions. F-stat refers to the KP F-stat for weak instruments. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

D.2 Inference

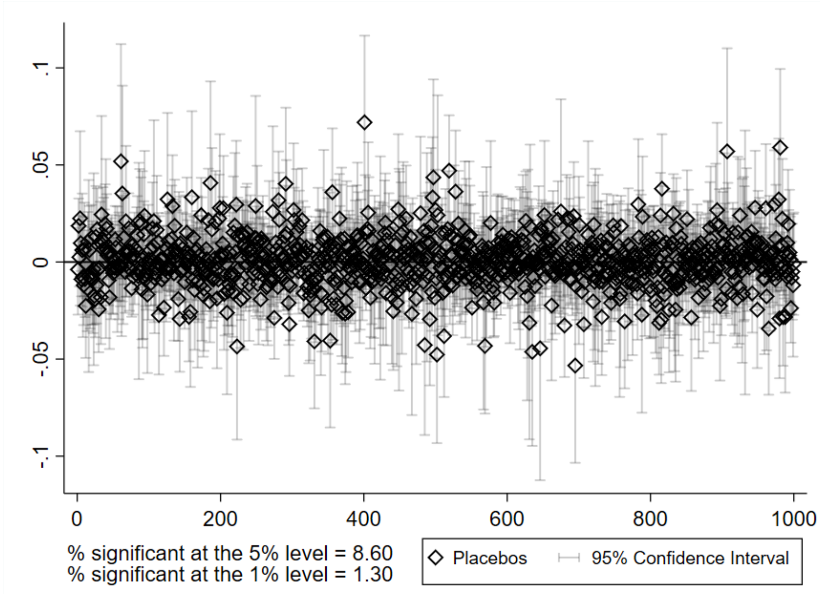
Adao et al. (2019) show that standard errors associated with 2SLS estimates in the case of shift-share instruments may be excessively small. We tackle this issue by performing a placebo exercise similar to that conducted by Derenoncourt (2019) in the context of the Second Great Migration. In particular, we interact the initial shares of Southern Black migrants (i.e. the α_{jn}^{1900} terms of equation 2 with a normally distributed random variable with mean 0 and variance 5. Parameter values were chosen following Adao et al. (2019) and Derenoncourt (2019). We then iterate this procedure 1,000 times, and report the fraction of times estimates are significant at the 5% and 1% level. Results, reported in Figures D.4 and D.5 are reassuring: the coefficient on Black in-migration is significant 2.7% of the times at the 5% level and 0.4% of the times at the 1% for intermarriage, and 8.6% of the times at the 5% level and 1.3% of the time for naturalization.

Figure D.4. Placebo migration shocks: Married w/ native



Notes: The figure plots the coefficient on placebo shocks in 1,000 separate regressions. The placebo shock is defined as the interaction between 1900 Black Southern migration patterns (α_{jn}^{1900} terms in eq.(2)) and a normally distributed random variable with mean 0 and variance 5. Regressions are estimated at the MSA level and control for MSA and decade fixed effects and for total MSA population. Robust standard errors are clustered at the MSA level.

Figure D.5. Placebo migration shocks: Naturalized



Notes: The figure plots the coefficient on placebo shocks in 1,000 separate regressions. The placebo shock is defined as the interaction between 1900 Black Southern migration patterns (α_{jn}^{1900} terms in eq.(2)) and a normally distributed random variable with mean 0 and variance 5. Regressions are estimated at the MSA level and control for MSA and decade fixed effects and for total MSA population. Robust standard errors are clustered at the MSA level.

D.3 Addressing compositional changes

An important set of concerns is related to the possibility that the Great Migration triggered (selective) flight. Most directly, if the least assimilated immigrants left Northern MSAs with many Black migrants, our estimates would be the result of compositional changes rather than changed attitudes of native whites. Relatedly, immigrant flight from – or curbed immigrant in-migration to – Northern MSAs experiencing Black migration could lead to smaller immigrant enclaves, and precipitate assimilation through a pure group size effect.

Using the linked sample mitigates some of these concerns, as it shows that our estimated effects are not entirely driven by out-migration of less assimilated immigrants. However, compositional changes, and the declining size of the immigrant population could indirectly affect the assimilation of immigrants remaining in the same MSA throughout the 1910-1930 period. We perform several checks to address such concerns.

We start by showing that the inflow of African Americans did not have any statistically significant effect on the share of the foreign-born, constructed using different reference populations in the denominator. We estimate MSA-level regressions, controlling for MSA and region by decade fixed effects, as well as for MSA population. As in our baseline specification, the main regressor of interest is the number of Blacks in the MSA in each decade. Regressions are weighed by 1900 MSA population, and standard errors are clustered at the MSA level. In Panel A of Table D.18, the dependent variable is the number of all immigrants (column

1) and European immigrants (column 2) over MSA population. Columns 3 to 10 separately consider each region of origin. The coefficient on Black in-migration is always imprecisely estimated and quantitatively small. In addition, there is no systematic pattern across nationalities. Panels B and C replicate the analysis constructing the share of immigrants relative to all non-Black (Panel B) and to native-born non-Black population (Panel C) in the MSA, respectively. Also in this case, coefficients are always unstable, quantitatively small, and imprecisely estimated. Finally, in Panel D, we examine the share of immigrants from each origin relative to the MSA foreign born population. Once again, Black in-migration has no systematic effect on the composition of the immigrant group.

Table D.18 provides evidence against immigrant flight. It is still possible that, while the fraction of immigrants remained constant, their composition changed in response to Black migration. It is hard to assess this from cross-sectional data, as we only have few baseline variables that are not potentially endogenous to Black inflows. Table D.18 shows that composition in terms of origin region does change differentially in MSAs with more Black migrants. Table D.19 shows that this is also true for immigrant's age, both on average, and separately for each region of origin. In Table D.20, we additionally verify that Black in-migration did not alter the sex ratio within the immigrant group, spuriously driving our results on intermarriage. Reassuringly, there is no statistically significant relationship between Black inflows and sex ratios, either for younger cohorts or for older ones.[47]

To further examine the possibility that Black in-migration was associated with systematic changes in the characteristics of immigrants in Northern MSAs, we turn to evidence from the linked sample. We estimate a series of regressions comparing characteristics between individuals who moved and those who stayed in the same MSA for the entire 1910-1930 period. In Table D.21, we regress the characteristic (measured in 1910) of the individual reported at the top of each column against *i*) a dummy equal to 1 for being a "mover", and *ii*) its interaction with the (instrumented) 1910-1930 change in the Black population in the MSA. All regressions include MSA fixed effects and an interaction between the mover indicator and the 1910-1930 change in total MSA population.[48]

Movers and non-movers are systematically different, generally in the expected direction. Movers are younger, less likely to be married, and more likely to be employed in unskilled occupations. Yet, with the exception of age, there is no systematic difference in that gap by degree of exposure to Black migration. Importantly, we find no differential out-migration of Italian and Irish immigrants – two nationalities frequently documented to exhibit prejudice against African Americans (Ignatiev, 1995). This indicates that more racially intolerant individuals were not more likely to migrate out of MSAs with more Black arrivals.

[47]Following 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 when considering the second generation, and when looking at sex ratios separately for each ethnic group.

[48]The inclusion of MSA fixed effects absorbs the 1910-1930 change in Black and total MSA population.

We next turn to the possibility of selective flight among native-born whites. We start by examining this using a cross-sectional dataset of all native-born whites between 1910 and 1930, focusing on characteristics that are expected to change little, if at all, during the life course. Table D.22 provides no evidence of systematic differences by exposure to Black in-migration in the share of the native-born white population that was male or literate. The average age of native-born whites tends to increase in response to Black inflows, but the effect is not statistically significant at conventional levels. All in all, these results provide no evidence for selective out-migration of whites.

Arguably, selective out-migration driven by racial intolerance is more of a concern for native-born whites than for immigrants, since it is the attitudes of this population that determined immigrants' assimilation outcomes. In order to more directly examine this possibility, we use the samples linked across 1910-1920 and 1920-1930 by Abramitzky et al. (2020b). We focus on native-born white individuals of native-born parents linked across all three census years and residing in one of the MSAs in our sample in 1910. We identify as "movers" those who were not residing in the same MSA in all of the three census decades. In Table D.23, we replicate the analysis of Table D.21. Once again, the difference between movers and non-movers is not systematically correlated with Black migration. In columns 5-7, to better proxy for preferences towards outgroups, we consider intermarriage with a foreign-born spouse (column 5), a native-born spouse of foreign-born parentage (column 6), and a spouse of a different race (column 7). Reassuringly, there is no significant effect of Black inflows on the difference in these outcomes between movers and non-movers.

Table D.18. Black migration and immigrant composition

Dep. variable	Fraction of Immigrants from									
	All	Europe	Russian Empire	Northern Europe	UK	Ireland	Western Europe	Southern Europe	Central & Eastern Europe	Germany
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Panel A. Relative to MSA population									
Black population	-0.005	-0.008	-0.002	0.002	-0.003	-0.003	-0.001	0.000	-0.004	0.002
	(0.011)	(0.012)	(0.003)	(0.006)	(0.002)	(0.003)	(0.001)	(0.002)	(0.005)	(0.007)
1910 Mean	0.191	0.158	0.017	0.020	0.020	0.016	0.007	0.019	0.024	0.035
	Panel B. Relative to non-Black MSA population									
Black population	-0.005	-0.007	-0.002	0.003	-0.003	-0.003	-0.001	0.001	-0.004	0.002
	(0.012)	(0.013)	(0.003)	(0.006)	(0.003)	(0.003)	(0.001)	(0.002)	(0.006)	(0.007)
1910 Mean	0.195	0.161	0.017	0.020	0.021	0.017	0.008	0.019	0.024	0.035
	Panel C. Relative to non-Black native MSA population									
Black population	0.009	0.003	0.001	0.004	-0.003	-0.002	-0.001	0.003	-0.006	0.007
	(0.025)	(0.027)	(0.004)	(0.010)	(0.004)	(0.005)	(0.001)	(0.004)	(0.009)	(0.010)
1910 Mean	0.261	0.216	0.023	0.028	0.027	0.023	0.010	0.026	0.033	0.046
	Panel D. Relative to foreign born MSA population									
Black population		0.012	0.005	-0.007	-0.014	-0.012	-0.003	0.019	0.018	0.005
		(0.010)	(0.013)	(0.005)	(0.009)	(0.007)	(0.003)	(0.013)	(0.017)	(0.021)
1910 Mean		0.843	0.083	0.084	0.110	0.088	0.043	0.091	0.121	0.223
Observations	324	324	324	324	324	324	324	324	324	324
F-stat	17.46	17.46	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 instrument could be constructed and to decades 1910 to 1930. All regressions are weighted by 1900 MSA total population and control for MSA and region by decade 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 D.19. Black migration and immigrants' average age

Dep. variable	Average age									
	All	Europe	Russian Empire	Northern Europe	UK	Ireland	Western Europe	Southern Europe	Central & Eastern Europe	Germany
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Black	0.179	-0.269	0.120	0.406	1.480	1.857	0.606	-0.465	0.432	-0.911
population	(1.038)	(1.125)	(0.428)	(1.111)	(1.275)	(1.956)	(1.140)	(0.527)	(0.439)	(1.097)
Observations	324	324	324	324	324	324	324	324	324	324
F-stat	17.46	17.46	17.46	17.46	17.46	17.46	17.46	17.46	17.46	17.46

Notes: All regressions control for MSA total population, MSA and region by decade fixed effects, and are weighted by 1900 MSA population. F-stat refers to the KP F-stat for weak instruments. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

Table D.20. Sex ratios

Dep. variable	Sex ratio		
	All	Young	Old
	(1)	(2)	(3)
Black population	-0.084	-0.042	-0.102
	(0.102)	(0.113)	(0.089)
Observations	324	324	324
F-stat	17.46	17.46	17.46

Notes: The analysis is restricted to 108 non-southern MSAs for which the instrument could be constructed, and to years 1910 to 1930. All regressions control for MSA and region by decade fixed effects, MSA total population and are weighted by population in 1900. F-stat refers to the KP F-stat for weak instruments. Robust standard errors clustered at the MSA level in parentheses. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

Table D.21. Selective migration among immigrants

Dep. variable	Age	In Manufacturing	Unskilled	Married	Irish	Italian
	(1)	(2)	(3)	(4)	(5)	(6)
Mover	-5.856***	0.010	0.117***	-0.220***	0.027***	0.092***
	(0.219)	(0.007)	(0.007)	(0.007)	(0.003)	(0.006)
Mover × Δ Black pop.	-1.520*	-0.017	-0.003	-0.006	0.006	-0.016
	(0.903)	(0.019)	(0.021)	(0.024)	(0.010)	(0.010)
Observations	224,981	196,723	196,723	224,981	224,981	224,981
F-stat	6.364	6.417	6.417	6.364	6.364	6.364

Notes: The analysis is restricted to linked foreign born men observed in 1910 in one of the 108 sample MSAs. In columns 2 and 3, the sample is restricted to individuals aged 15 to 64. *Mover* is a dummy equal to 1 if the individual did not always reside in the same MSA. All regressions control for MSA fixed effects and for the interaction between the 1910-1930 change in MSA population and the mover indicator. F-stat refers to the KP F-stat for weak instruments. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

Table D.22. Compositional changes among native whites

Dep. V+variable	Share literate	Share male	Average age
	(1)	(2)	(3)
Black population	-0.004	0.001	1.307
	(0.003)	(0.004)	(0.895)
Observations	324	324	324
F-stat	17.46	17.46	17.46

Notes: All regressions control for MSA total population, MSA and region by decade fixed effects and are weighted by 1900 population. F-stat refers to the KP F-stat for weak instruments. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.

Table D.23. Selective migration among native whites

Dep. variable	Age	In Manufacturing	Unskilled	Married	Married immigrant	Married w/ 2nd gen. immigrant	Married w/ non-white
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mover	-2.747***	0.003	-0.010***	-0.084***	0.004	-0.004	0.0001
	(0.338)	(0.002)	(0.002)	(0.008)	(0.002)	(0.007)	(0.0001)
Mover$\times$ Δ Black pop.	3.540	0.008	-0.011	0.083	-0.011	-0.055	-0.0003
	(2.501)	(0.007)	(0.006)	(0.063)	(0.012)	(0.048)	(0.0002)
Observations	610,691	610,691	610,691	610,691	178,669	178,669	178,669
F-stat	6.182	6.182	6.182	6.182	6.596	6.596	6.596

Notes: The analysis is restricted to linked native-born men of native parentage observed in 1910 in one of the 108 sample MSAs. In columns 2 and 3, the sample is restricted to individuals aged 15 to 64. *Mover* is a dummy equal to 1 if the individual did not always reside in the same MSA. All regressions control for 1900 MSA total population, MSA fixed effects and for the interaction between the 1910-1930 change in MSA population and the mover indicator. F-stat refers to the KP F-stat for weak instruments. Robust standard errors, clustered at the MSA level, in parentheses. Significance levels: *** $p< 0.01$, ** $p< 0.05$, * $p< 0.1$.

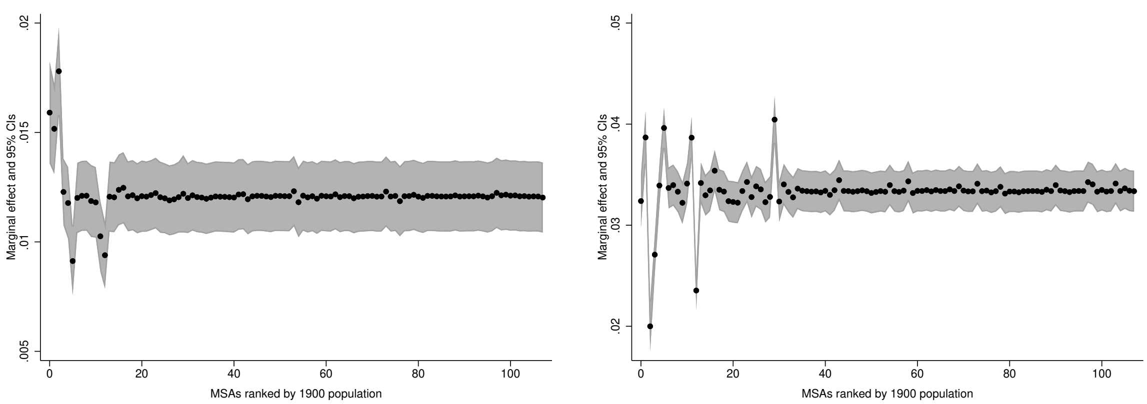
D.4 Additional robustness checks

One remaining concern with a specification in levels is that results might be at least in part driven by outliers. In Figure D.6, we show that, reassuringly, this is not the case. We plot 2SLS 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 do not entirely rely on any one MSA, and always retain their sign and statistically significance. Figure D.7 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.

The particular form of heterogeneity observed in Figure D.6 can be explained based on the composition of the immigrant group in MSAs of different size in 1900. Larger MSAs tended to have relatively more immigrants from "New Source" countries (Central and Eastern Europe, Southern Europe, the Russian Empire). As Figure 6 in the main text indicates, the effect of Black in-migration was highest on naturalization rates for immigrants in this group; instead, the effect on intermarriage rates was highest for immigrants from "Old Source" countries (UK, Northern and Western Europe, Germany, Ireland). Consistent with this heterogeneity, dropping MSAs with relatively more New than Old source country immigrants leads to an increase in estimates for intermarriage rates, and a decrease in estimates for naturalization.[49]

Figure D.6. Robustness to dropping outliers

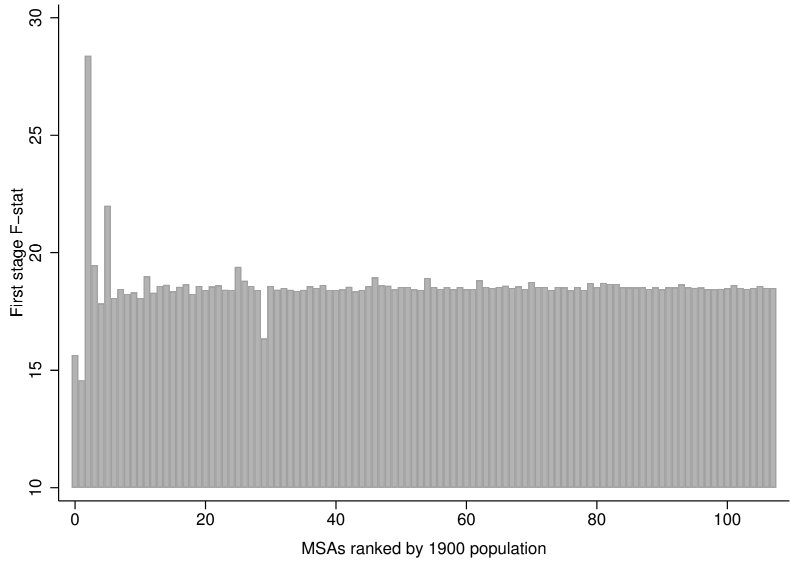
(a) Married w/ native (b) Naturalized



Notes: 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 3 of Table 2 that drop one MSA at a time. MSAs are ranked according to 1900 population.

[49]Replicating Figure D.6 ordering MSAs by the ratio of old to new source country immigrants makes this pattern clear. Results available from the authors.

Figure D.7. First stage robustness to dropping outliers



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