

NBER WORKING PAPER SERIES

THE ETHNIC SEGREGATION OF IMMIGRANTS IN THE UNITED STATES FROM
1850 TO 1940

Katherine Eriksson
Zachary A. Ward

Working Paper 24764
<http://www.nber.org/papers/w24764>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2018

Thanks to Tim Hatton, Laura Panza, John Parman, Allison Shertzer and Dafeng Xu for helpful comments. We thank those at the University of Minnesota Population Center and Ancestry.com for access to historical census files. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2018 by Katherine Eriksson and Zachary A. Ward. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Ethnic Segregation of Immigrants in the United States from 1850 to 1940
Katherine Eriksson and Zachary A. Ward
NBER Working Paper No. 24764
June 2018
JEL No. F22,J61,N31

ABSTRACT

We provide the first estimates of ethnic segregation between 1850 and 1940 that cover the entire United States and are consistent across time and space. To do so, we adapt the Logan-Parman method to immigrants by measuring segregation based on the nativity of the next-door neighbor. In addition to providing a consistent measure of segregation, we also document new patterns such as the high levels of segregation in rural areas, in small factory towns and for non-European sources. Early 20th century immigrants spatially assimilated at a slow rate, leaving immigrants' lived experience distinct from natives for decades after arrival.

Katherine Eriksson
Department of Economics
University of California, Davis
One Shields Avenue
Davis, CA 95616
and NBER
kaeriksson@ucdavis.edu

Zachary A. Ward
25a HW Arndt Building
Research School of Economics
Canberra, ACT 2600
Australia
Zach.A.Ward@gmail.com

It is well known that immigrants are not randomly distributed across a country; rather, they tend to cluster near each other and end up segregated from the native born. This pattern has led to an extensive literature exploring the causes of ethnic segregation and consequences for employment, intermarriage and second-language acquisition.¹ Much of this evidence comes from American history, when the United States received immigrants from a wide variety of sources in the 19th and early 20th centuries (e.g., Lieberman, 1963; Lieberman, 1980; Cutler et al., 2008a). Yet despite the importance of ethnic segregation, there are still issues with the first-order problem of measuring segregation: due to several data limitations, there are still no consistent and comprehensive time series of ethnic segregation during the Age of Mass Migration and beyond (1850-1940).

Most segregation measures are based on how immigrants and natives are allocated across different sub-city areas, such as across city wards, census tracts or enumeration districts. Unfortunately, these measures fail to cover key segments of the migrant population outside of the major urban centers. In particular, rural segregation has been routinely ignored despite rural areas containing half of the migrant population in the 19th century. Even when measures do cover urban areas, comparing segregation across cities and census years can be problematic because the sub-city area is not always consistently sized across time and space. This problem is especially severe with the city ward, the most-used unit in the pre-1940 segregation literature (Shertzer et al., 2016).² Since much of the literature relies on the city ward, we still do not have

¹ There is a long sociology literature on segregation, primarily associated with Duncan and Lieberman (1959), Lieberman (1963), and Massey and Denton (1988). Economists have studied the effects of immigrant networks or ethnic enclaves on various economic outcomes (e.g., Munshi, 2003; Damm, 2009; Edin et al., 2003; Cutler et al., 2008b). A common finding is that, after taking selection into account, living in an enclave, in terms of a higher fraction or total number of foreign born in an area, increases immigrant earnings and wages. Beaman (2011) argues that the effect varies for recent arrivals and longer-established arrivals since recent arrivals are in more direct competition with new arrivals. However, evidence from the 19th and early 20th century Norwegians suggests that enclaves worsened economic outcomes (Eriksson 2018).

² Moreover, the city ward is often too large to detect segregation at local levels, which has led others to use the much smaller enumeration districts or census tracts (Cutler et al., 2008a; Hershberg, 1976; Logan and Zhang, 2012). While district-based measures are a vast improvement over those based on the city ward, census tracts are only available after 1940, and enumeration districts are not available prior to 1880. Therefore, they do not help provide a consistent measure of segregation between 1850 and 1940.

high-quality information on how segregation changed for the key periods of immigration, such as during high inflow years for the Irish following the Great Famine and for Southern and Eastern Europeans prior to World War I.

To address these problems, we take a simple approach to build the first panel of ethnic segregation that both covers the entire United States and is comparable across time and space: we measure segregation based on whether the next-door neighbor was native born. This next-door neighbor method was first used by Trevon D. Logan and John M. Parman (2017a), who applied it to black-white segregation in the 1880 and 1940 full-count Censuses. The key innovation of the measure is to exploit the fact that historical censuses were taken on a line such that neighbors are listed immediately next to each other on the enumeration page (Agresti, 1980). The resulting neighbor-based measure is advantageous relative to other measures in that it is straightforward and intuitive, is consistent across time and space, covers rural and urban areas, and is straightforward to implement. Instead of using race as the basis for the in- and out-group as in Logan and Parman (2017a, 2017b), we use country of birth for the in-group and the native-born for the out-group, and then apply this measure to each full-count census between 1850 and 1940.³

The neighbor-based measure reveals several new insights on ethnic segregation throughout American history. First, the most highly segregated areas in the United States were not the main entry ports of New York, Boston and Philadelphia; rather, they were smaller factory towns and rural areas that were heavily reliant on migrant labor. Some of the highest levels of segregation were for Irish in Lowell, Massachusetts, and Austro-Hungarians in Passaic, New Jersey. Farming communities in the 19th century were also highly segregated, especially for Scandinavians – they even nearly reached the urban segregation levels for

³ We also calculate segregation from those born in a different country (rather than segregation from those born in the United States), and find similar qualitative results. See Appendix D.

Italians and Russians in the early 20th century. The high levels of rural segregation suggest that segregation was not purely an urban phenomenon that reflected industrial composition, anti-immigrant residential policies or city structure; rather, segregation emerged since enclaves provided economic and social benefits for new arrivals.

Since the neighbor-based measure is consistent across time and space, it allows us to compare segregation across well-known enclaves in American history. For example, the Irish in 1850 Boston were slightly more segregated than Italians in 1910 New York. However, both Irish and Italians were less segregated than Russian and Polish immigrants in 1900 Chicago. While the segregation of Europeans has long been of interest, another contribution of the measure is that it covers immigrants from non-European sources, such as those arriving from Mexico and China. The Chinese were among the most highly segregated ethnicities in the 19th century; in fact, Chinese segregation in 1880 San Francisco is the highest segregation level for the entire 1850 to 1940 period. Mexican segregation was also high but was more similar to that of Southern Europeans; therefore, Mexicans were not uniquely segregated despite the substantial discrimination that faced Mexicans in the early 20th century.

While the neighbor-based measure provides a more comprehensive and consistent depiction of segregation than previous work, it does not overturn conclusions from prior studies on major cities; in fact, it confirms a few speculations already in the literature. First, we show that, on average, pre-1870 segregation levels for Western Europeans were high, but they were nowhere near that of Southern and Eastern Europeans in the early 20th century (with the exception of the Irish in mid-19th century Boston). Second, we confirm that Southern and Eastern European segregation steadily decreased between 1910 and 1940, a pattern long expected but never conclusively shown due to the switch from ward-based to tract-based measures in 1940 (Cutler et al., 2008a; Lieberson, 1963). However, the fall in segregation is less steep when one measures segregation from third-generation natives. The downward trend

in ethnic segregation in the early 20th century contrasts with an upward trend in black-white segregation shown by Logan and Parman (2017a, 2017b), a pattern also recognized previously – but here we show it applies to both rural and urban areas across the entire country (Lieberson, 1963; Lieberson, 1980; Cutler et al., 1999; Cutler et al., 2008a).

While the fall in ethnic segregation after 1910 suggests that immigrants spatially assimilated rapidly by quickly moving out of immigrant neighborhoods after arrival, this was not the case. Using linked census data from the 1910-1930 censuses, we show that only 40 percent of 1905-1909 European households had a native-born neighbor at arrival.⁴ This is in comparison with 90 percent of native-born households with a native-born neighbor. After a decade of duration in the United States, this gap between the native- and foreign-born closed slightly from 50 to 41 percentage points, reflecting that immigrants did spatially assimilate, but at a slow rate. A slow rate of spatial assimilation is consistent with a lack of convergence in occupational distributions for many source countries between 1900 and 1920 (Abramitzky, Boustan and Eriksson, 2014); yet it contrasts with the quick rate of social assimilation after arrival in terms of English acquisition, immigrants adopting Anglicized names and immigrants naming their children with Anglicized names (Abramitzky, Boustan and Eriksson, 2016; Biavaschi et al., 2017; Ward, 2018). Therefore, despite the social assimilation of immigrants, immigrants' average lived experience was quite distinct from that of the native born.

I. Overview of literature on historical segregation measures

One of the earliest studies to quantify ethnic segregation also demonstrates a key limitation of the literature. Stanley Lieberson (1963) measured ethnic segregation in ten major cities and showed that segregation fell between 1910 and 1920, and also between 1930 and 1950. However, the problem is that one cannot directly compare the 1910-1920 and 1930-1950

⁴ The linked census data is from Ward (2018), who applied the Feigenbaum (2016) linking method to immigrants between 1910 and 1930.

periods because segregation is calculated with city wards in the earlier period and with census tracts in the later period.⁵ Besides the fact that city wards may be gerrymandered to reflect ethnic neighborhoods, they can also be over 10 times larger than census tracts and therefore hide segregation; indeed, tract-level racial segregation measures yield dissimilarity scores about 15 points higher than ward-level measures (Cutler et al., 1999). Unfortunately, census tracts did not become available for the entire United States until 1940; before this, they were only available in a select group of cities as in the Lieberman (1963) study. Because of this switch from city wards to census tracts, David Cutler, Edward Glaeser and Jacob Vigdor (2008a), who provide a long-run series of ethnic segregation between 1910 and 2000, show no absolute fall in dissimilarity between 1910 and 1940. This may lead a naïve reader to conclude the dissimilarity-based segregation did not fall in the early 20th century.⁶ Cutler et al. and Lieberman are careful to note this measurement issue with city wards and census tracts in text, but the true drop in ethnic segregation in the early 20th century has not been conclusively established.

The problem of using city wards to measure segregation is well known; therefore, many have resorted to census manuscripts to calculate segregation at finer levels of geography. However, this method is quite costly and therefore has been employed by only a few researchers (e.g. Thernstrom, 1973; Kantrowitz, 1979; Zunz, 1982).⁷ The most comprehensive study using this method was the Philadelphia Social History Project, which plotted the addresses of over 2.5 million Philadelphians between 1850 and 1880 (Hershberg, 1976).⁸ After

⁵ Census tracts are not available for all cities until the 1940 census. Lieberman (1963) is not the first to calculate dissimilarity measures, but is the first to do it for several different cities. Duncan and Lieberman (1959) calculate measures for Chicago across time.

⁶ Cutler et al. (2008) show a fall in isolation-based segregation between 1910 and 1940. See Massey and Denton (1988) for a discussion of different segregation measures, including the isolation and dissimilarity index.

⁷ White et al. (1994) use the 1 in 250 sample from the 1910 Census to explore whether sampled households on either side of the immigrant were foreign or native born, under the assumption that individuals 250 people apart was a good proxy for a neighbor. Our study improves on White et al. (1994) by filling in this 250-person gap with the full-count data; furthermore, we use full-count data from multiple censuses to estimate the trend in segregation over time, as opposed to White et al.'s (1994) snapshot of 1910.

⁸ More recently, the work of John Logan and various co-authors have continued this detailed work of mapping addresses, but so far this primarily involves the 1880 census (e.g., Logan and Shin, 2016; Logan and Martinez, 2018; Spielman and Logan, 2013).

projecting 1930 census tract boundaries onto mid-19th century maps, Hershberg et al. (1981) document that dissimilarity levels were low for Irish and German migrants in 1850 at about 0.30, but then then increased slightly to 0.35 in 1880. A small increase in dissimilarity-based segregation may be surprising given the large inflows of Irish and Germans after 1850; unfortunately, evidence on segregation for the years between 1850 and 1880 outside of Philadelphia is scarce.

This detailed evidence from Philadelphia has led to a consensus that segregation levels were lower in the earlier stages of the Age of Mass Migration and were higher for Southern and Eastern European sources – yet Philadelphia may not be representative of the entire country. Recent efforts to digitize entire censuses allow researchers to look beyond Philadelphia; for example, John Logan and Weiwei Zhang (2012) use the full-count 1880 census to estimate segregation measures for 67 cities across the country. They calculate segregation measures using enumeration districts, which are about the size of a census tract – a vast improvement over the city ward due to the enumeration district’s size and comparability with tract-based measures. After exploring cities outside of Philadelphia, they confirm that segregation levels were relatively low for “old” sources in 1880 compared with “new” sources in the early 20th century. However, Logan and Zhang also show that the variation in dissimilarity measures across cities was wide, which suggests that the city by city studies prior to 1880 may not be informative of the national average. While using enumeration districts to measure dissimilarity is promising, unfortunately enumeration districts do not exist prior to 1880, so one cannot use them to extend segregation measures back to 1850.

Even though measurement of segregation improves when researchers exploit census manuscripts, the literature has ignored segregation outside of larger cities. The literature’s focus on cities partially reflects that most immigrants settled there in the 20th century, and also that it is difficult to calculate a dissimilarity index in an area without city wards. However,

about half of immigrants lived in rural areas in the 19th century, leaving a large gap in the literature. Rural settlement was especially common for Northern and Western Europeans in the Midwest, where many small towns today are still connected with the ethnic identity formed in the past – such as for the Dutch in Holland, Michigan, and the Swiss in Berne, Indiana. Immigrants who lived in less populated areas did not just work in agriculture, but also in mining and manufacturing; these industries relied on cheap labor from abroad in both the 19th and 20th centuries. Yet we still do not know the extent of segregation outside the major cities.

Our paper continues the trend of using newly digitized census files to measure ethnic segregation. Since we observe everyone who is enumerated, we can exploit the census manuscripts to fix the major measurement issues in the literature. First, we cover more areas, including rural communities and smaller towns. Second, we measure segregation for decades and cities previously unquantified, particularly during the first major wave of immigration between 1850 and 1880 in cities outside of Philadelphia. Third, we provide measures that are comparable across time and space, and do not depend on inconsistently sized city wards. Finally, we measure segregation for non-European sources from Mexico and China, which has been overlooked in the literature. All of this can be done due to the digitization of full-count census files between 1850 and 1940.

II. Applying the Logan-Parman method to immigrants between 1850 and 1940

We use full-count Census data between 1850 and 1940 to measure ethnic segregation. This data is available from IPUMS at the University of Minnesota Population Center (Ruggles et al., 2017) and was accessed at the National Bureau of Economic Research (NBER).⁹ We measure segregation based on the country of birth of the next-door neighbor’s household head, which we can observe because, starting with the 1850 Census, the census was taken “on a line”

⁹ Currently, the University of Minnesota has cleaned and released versions of the 1850 and 1880 Censuses and preliminary versions of the 1900 to 1940 United States Censuses. We clean the 1860 and 1870 Censuses as described in Appendix A.

such that households listed next to each other on a census page are reasonable proxies for next-door neighbors (Logan and Parman, 2017a). Censuses prior to 1850, while also available from IPUMS, do not record country of birth; moreover, they were not enumerated on a line (Agresti, 1980). We do not use the 1890 Census because most of the original manuscripts were lost in a fire.¹⁰

We measure segregation of the foreign-born following Logan and Parman's (2017a) method for black-white segregation with a few simple modifications: primarily, instead of using race for the in- and out-group, we use a specific country of birth for the in-group (which we refer to as ethnicity) and the native born for the out-group.¹¹ There are several other ways one could create in-groups and out-groups. For example, between 1880 and 1930 we have further information on mother and father's country of birth, so for these censuses we can alternatively define the out-group as US-born to two US-born parents. We could also define the out-group as all others from a different country of birth, not just the native born. We do this in Appendix D, which shows similar qualitative results for most countries as our preferred out-group of the native born. We focus on using the native born as the out-group to be consistent with the ethnic segregation literature (e.g., Lieberson, 1963). The literature focuses on segregation from the native born since this is related to other types of social and economic assimilation, such as moving out of enclaves to take advantage of better economic opportunities or public amenities, or linguistic assimilation through contact with natives. Here we will briefly

¹⁰ One should keep in mind that a fundamental limitation of the data is that we cannot estimate segregation for those not enumerated. Hacker (2013) estimates that under-enumeration in the census was common, where about 4 to 7 percent of the *native-born white* population was not counted between 1850 and 1930. Hacker does not estimate under-enumeration of the *foreign-born* population since one cannot fully separate undercount estimates from return migration estimates; however, the standard assumption is that under-enumeration of immigrants is more severe due to difficulties resulting from language barriers or the more transient nature of the immigrant population. If those who were not enumerated were more segregated than those enumerated, then we would underestimate the true level of segregation between 1850 and 1940.

¹¹ It would be preferable to measure segregation by language group, but this is unavailable across the 1850 to 1940 period.

describe the segregation measure, but those interested in more detailed information should reference Appendix B.

To create the neighbor-based segregation measure, we first keep the household head, dropping those in non-households and other non-heads in the household. In other words, we measure the segregation of households and not the segregation of individuals. This is a non-trivial restriction since immigrants were also non-family members such as boarders or servants, and also lived in non-household institutions such as employee camps. For example, about 90 percent of the migrant population lived in households between 1850 and 1940, leaving 10 percent in non-households.¹² Of those in households, about 10 percent were non-relatives of the head.¹³ Moreover, by keeping the household head, we do not account for the birth place of the spouse. We will explore the robustness of the measure to keeping others in the household and non-households in Appendix C, but we will keep to household heads now so our measures are comparable to racial segregation measures from Logan and Parman (2017a, 2017b). While keeping only household heads may be problematic, we find similar segregation estimates when accounting for others in the household.

After keeping household heads and defining the nativity of the household based on the head, we then identify those on the same census page and sort them by line number such that the households listed next to each other proxy for a next-door neighbor.¹⁴ After this sorting, we create a variable which indicates whether either of the next-door household heads are native born, a variable on the extensive margin rather than the intensive margin of how many neighbors were native born.

¹² This is based on authors' calculation from IPUMS, with a low of 85.9 percent in 1850 and a high of 94.8 percent in 1940.

¹³ This is based on authors' calculation from IPUMS, with a high of 17.4 percent in 1850 and a low of 5.3 percent in 1940.

¹⁴ We sort by the y-coordinate position in the raw full-count census files in 1860 and 1870 when line number is not available.

Given this information on the next-door neighbors, for each county and country of birth we know (1) the number of foreign-born households, (2) the number of native-born households, and (3) the number of foreign-born households with a native-born neighbor. The neighbor-based measure uses these values in a formula to compare the observed level of segregation to extremes of random assignment or complete segregation.¹⁵

$$\eta_c = \frac{E(\overline{native}_c) - native_c}{E(\overline{native}_c) - E(\underline{native}_c)} \quad (1)$$

To calculate the segregation measure η_c for country of birth c , the number of foreign-born individuals with at least one native-born neighbor ($native_c$) is compared with the expected number under the conditions of either random household location ($E[\overline{native}_c]$) or complete segregation from the native born ($E[\underline{native}_c]$). Complete segregation from the native born suggests that the ethnic neighborhood (enumerated on a line) is surrounded by foreign-born households from other countries of birth. Therefore, complete segregation would lead to zero native-born neighbors ($E[\underline{native}_c] = 0$).¹⁶ See Appendix B for the formula for ($E[\overline{native}_c]$). The segregation measure typically ranges from zero to one, where one indicates perfect segregation and zero indicates random assignment of neighbors and thus complete integration. While the measure can be calculated for any level of geography, in this paper we present measures at the county/city-level since we wish to describe the broad trends of segregation.

¹⁵ This measure can be conceptualized as a measure of evenness across households in a county or city, similar to the dissimilarity index (Massey and Denton, 1988; Logan and Parman, 2017a). The neighbor-based measure captures complete segregation and integration well when there are more than ten foreign-born households, but can be noisy with less than ten households (Logan and Parman, 2017a). For this paper we will primarily focus on the national trends in segregation and the level of segregation in cities and rural counties with more than 1,000 ethnic households rather than segregation for very small immigrant communities.

¹⁶ This is not true for counties or cities where the foreign born come entirely from one country of birth and no others; however, this rarely happened. For example, it does not occur in the 1880 full-count census.

To provide an idea of how the segregation measure works, consider the example of Italians in 1910 Manhattan. According to our data, there were 66,428 Italian households, 190,006 native-born households and 576,557 total households in 1910 Manhattan. Under random assignment, one would expect to observe about 32,422 Italian household heads with at least one next-door native-born neighbor (see Appendix B for formula). However, we only observe 9,684 Italians that have at least one native-born neighbor. Under the other extreme of complete segregation, one would expect zero native-born neighbors since Italians would be clustered along a line with non-Italian foreign-born households (such as a German or Polish household) on either side of this counterfactual Italian neighborhood. After plugging these numbers into Equation (1), our segregation measure for Italians in 1910 Manhattan is 0.701.

While the segregation measure commonly ranges between zero and one, we document several important cases when the segregation measure goes below zero. A segregation level below zero indicates that immigrants lived closer to the native born than they would under random assignment. This occurs primarily because first-generation immigrants were more likely to live near second-generation individuals from the same origin rather than first-generation immigrants from a different origin. Due to this issue, we sometimes measure negative levels in major cities for long-established sources such as Germans in New York in 1930; however, we never measure a negative level segregation of immigrants from the *third* generation (i.e., native born to native-born parents). Unfortunately, we cannot observe the third generation for all years, but only between 1880 and 1930. Nevertheless, one should interpret a negative level of segregation as a population living closer to the native born than under random assignment.

III. The Ethnic Segregation of Immigrants between 1850 and 1940

A. Measures by source country

Figure 1 presents trends in segregation levels from 1850 to 1940 after grouping countries of birth into either Western, Northern, Eastern or Southern European.¹⁷ The neighborhood-based measure immediately confirms a few inferences in the literature, although here we show the national trend while the literature has been limited to urban areas. First, the segregation of Western Europeans during their peak immigration period in the mid-19th century was less than that of Southern and Eastern Europeans during their peak period in the early 20th century. Early Western Europeans (i.e., English, Irish and Germans) started in 1850 with a segregation level of 0.34. On the other hand, Southern and Eastern European segregation levels were about 0.53-0.56 between 1900 and 1910. Therefore, the different immigrant waves had distinct experiences in the United States; this may reflect that Southern and Eastern Europeans entered a more highly urbanized country while earlier arrivals often moved to (less segregated) rural areas. We will explicitly measure differences in urban and rural segregation later.

A second lesson from Figure 1 is that segregation trended downward for all sources after 1910, indicating that immigrants became more integrated with the native-born population during the early 20th century. This trend has long been suspected but never confirmed due to the switch from ward-based to tract-based measures in 1940; here, we are able to confirm it with a consistent measure between 1910 and 1940. Declining segregation after 1910 is almost certainly because of the significant drop in inflows due to World War I and the immigration quotas. That is, if new arrivals were the most segregated, then a smaller fraction of new arrivals in the migrant stock would lead to a lower level of segregation. We will later directly show with linked data that new arrivals were indeed the most segregated. Yet a compositional shift

¹⁷ Germany is included in Western Europe. We code countries per IPUMS bpl codes: codes starting with 40 are Northern Europe, 41 or 42 are Western, 43 is Southern, 45 or 46 are Eastern, except for Germany. Note that Austria is included in Eastern Europe. To create the national trends by group (i.e. Western, Northern, Eastern or Southern), we first calculate the segregation for each county, city and country of birth. We then aggregate these scores to the national and group-level after weighting by the number of foreign-born households from that country of birth.

towards fewer recent arrivals does not explain the downfall after 1910 since segregation also fell for groups who stayed more than 10 years (see Appendix Figure A1).

A third insight from Figure 1 is that Northern Europeans, a group often ignored in the segregation literature due to their rural residence, were highly segregated in the mid-19th century. We measure their level of segregation at 0.50 in 1850 – slightly higher than Southern and Eastern European segregation in 1920. High levels of rural segregation imply that ethnic clustering was not purely an urban phenomenon during the Age of Mass Migration; rather, immigrants clustered for cultural and financial benefits. After their high levels of segregation in the mid-19th century, Northern European segregation steadily decreased in the following decades, most rapidly after 1880. Interestingly, Northern European segregation decreased when inflows increased in the 19th century, which is the exact opposite relationship for inflows and segregation for Southern and Eastern Europeans in the late 19th and early 20th centuries. Therefore, the relationship between inflows and overall segregation levels may not be so clear cut.¹⁸

We further split the broad regions of Western, Northern, Eastern or Southern European into 12 selected countries of birth in Figure 2 (see Table A1 for underlying estimates for all countries).¹⁹ This figure reveals starker differences across source countries than for the aggregated regions of Northern, Western, Southern and Eastern Europe. For example, English immigrants were the least segregated of all sources and remained at a low level of segregation throughout the entire period. In fact, English immigrants were perfectly integrated with native born in some decades, perhaps because most native born were descendants of England during

¹⁸ There is a quadratic relationship between fraction of foreign born in a county/city and segregation within our dataset, as shown in Figure A2, suggesting that higher inflows and thus a larger fraction of immigrants in a county is associated with more segregation.

¹⁹ We group Austria, Hungary and Czechoslovakia together to form Asutria/Hungary. We also group Russia, Poland, Estonia, Latvia and Lithuania together to form Russia/Poland. It would be better to group people by mother's tongue, which separates Jewish immigrants from other sources, but this is not available across all decades.

the mid-19th century. Standing out on the opposite end of the segregation spectrum was Norway, which was much more segregated than its Northern European counterparts of Denmark or Sweden.

Mexican and Chinese immigrants were highly segregated, but not much more so than Italians or Russians/Poles during their peak of immigration between 1900 and 1910. The peak of segregation for Mexicans was at 0.44 in 1920, the first decade immediately following the Mexican Revolution when hundreds of thousands fled the country for safety; yet many economic migrants came at the same time and worked in segregated mining towns and farming areas. Mexican segregation, like European segregation, fell following the 1920s, reflecting the mass movement back to Mexico due to the Great Depression and deportations. On the other hand, the peak of Chinese segregation was earlier in 1870 at 0.67 when there were relatively few Chinese household heads (~11,000). The segregation of Chinese fell in the next few decades to a low of 0.24 in 1920, lower than the level for Southern and Eastern Europeans. Therefore, the Chinese were indeed highly segregated, but primarily only in the 19th century.

B. Restricting the out-group to Third-Generation Native born

A common problem for ethnic segregation studies is that measured segregation may decrease as immigrants' native-born children age and live near the first generation; in this case, a next-door neighbor may be native born, but also have the same ancestry. This pattern may explain the strong downward trend in segregation during the 20th century as sources became more established in the United States. To account for this possibility, we use information on mother and father's country of birth when available between 1880 and 1930. Given that we can identify the third generation – that is, US-born to two US-born parents – we recalculate the segregation measures as the first generation's segregation from the third generation.

The results when measuring segregation from the third generation are also plotted in Figure 2.²⁰ The figure confirms that immigrants were less likely to live next to the third generation than to the second generation. For example, German segregation from the second generation was 0.26 in 1880, but 0.51 from the third generation, nearly twice the magnitude – and also higher than 1880 Italian segregation from the second generation (0.40). Italy’s segregation from the second generation was 0.59 in 1910, while segregation from the third generation was 0.76. Another key point when measuring segregation from the third generation is that the low levels of segregation observed for Western Europeans between 1920 and 1940 are somewhat misleading: while we observed a segregation level of near zero for Germans and English in 1930, this masks segregation from the third generation of 0.30 for England and 0.41 for Germany. This suggests that areas that were mixed immigrant-native communities were often communities of people from the same ancestral background.

Not only were immigrants more segregated from the third generation than from the second generation, but also segregation from the third generation trended downward more slowly over time. For example, Italian segregation from the *second* generation fell 0.23 points between 1910 and 1930 (from 0.59 to 0.36), while segregation from the *third* generation fell only 0.10 points (from 0.76 to 0.66). It is possible that the fall in ethnic segregation was even less steep if one could measure segregation from the fourth-plus generation, or those with American-born parents and four American-born grandparents, but grandparent’s country of birth is not observed. Yet at the same time measuring segregation from higher-order generations is not as informative since intermarriage was relatively common and immigrants socially assimilated relatively quickly in terms speaking English and adopting Americanized names (Abramitzky et al., 2016; Alba, 1985; Wildsmith et al., 2003).

C. Comparison to Black-White segregation

²⁰ See Table A2 for the numbers underlying segregation from the third generation.

Segregation levels for immigrants were high for some ethnicities, but how did these levels and trends compare to those of black-white segregation? This is a common question in the literature, dating back to early work from Lieberman (1963, 1980); thus, it is worth quickly reviewing results from dissimilarity and isolation measures already in the literature. First, black-white segregation was about equal to Southern and Eastern European ethnic segregation in 1910 (Cutler et al., 1999; Cutler et al., 2008a).²¹ Following 1910, black-white segregation and ethnic segregation diverged such that black-white segregation increased and Southern and Eastern European ethnic segregation decreased. However, these comparisons come from select cities and miss the large set of African Americans and foreign born who lived in rural areas. Given that we follow Logan and Parman's (2017a) methodology to measure segregation, it is straightforward to compare our estimates of ethnic segregation to their estimates of black-white segregation from 1880 to 1940.

The neighbor-based measure confirms the prior literature in that ethnic and black-white segregation levels started out similar in 1910, but then diverged afterwards. Logan and Parman (2017b) measure black-white segregation at about 0.58 in 1910, which was similar to our measures of ethnic segregation for Southern Europeans (0.56) and Eastern Europeans (0.54). However, if one measures ethnic segregation from the third generation instead of from the second generation, then ethnic segregation for Southern and Eastern Europeans (0.75-0.77) was much higher than black-white segregation in 1910. Of course, these national levels mask significant variation by city and source country where black-white segregation was higher than ethnic segregation, as pointed out by Lieberman (1980).

²¹ Cutler et al. (2008, Figure 4) show that ethnic segregation for Greece, Hungary, Italy and Russia was between 0.45 and 0.55 in 1910, while Cutler et al. (1999, Figure 1) show that black-white segregation was about 0.52 in 1910. Also see Cutler et al. (2005) for a direct comparison on black-white and new immigrant isolation indices. However, note that Lieberman (1963) shows that while black-white segregation was higher than ethnic segregation on average, this did not hold for every city in the early 1900s.

From this roughly equal level of ethnic and black-white segregation in 1910, the neighbor-based measure shows that black-white segregation increased, while ethnic segregation decreased. In 1940, Logan and Parman (2017b) calculate black-white segregation at 0.67, which was much higher than Southern and Eastern European segregation from the second generation at 0.18-0.21. Segregation levels for Asians and Mexicans were also lower than for African Americans in 1940, showing that African Americans were unique among racial and ethnic groups for their high levels of segregation in the middle of the 20th century.

D. Measuring Segregation across Urban and Rural Areas

In this section, we turn to document something which has been routinely ignored in the literature: segregation in rural areas. We are primarily interested in how the magnitude of segregation differed across rural and urban areas. We can further examine the trend in segregation levels across urban and rural areas, which may indicate whether urban phenomenon, such as the rise of mass transit or urban factories, led to increased segregation between 1850 and 1940. If segregation trended similarly in rural and urban areas, then this suggests that cultural or demographic factors were more influential, such as a preference for living in an ethnic community.

Figure 3 plots segregation by rural and urban counties for the same 12 source countries between 1850 and 1940.²² The figure demonstrates two key points. First, rural segregation for Western and Northern Europeans was often higher than urban segregation. This was especially true between 1850 and 1880 when rural Norwegians, Swedes, and the Dutch were more segregated than their urban counterparts. The level of segregation in rural areas could be quite high – Dutch rural segregation in 1850 was higher than Irish urban segregation at the same time, which may be surprising since this included the infamous Irish slums in Boston and New

²² See Table A3 for underlying numbers in Figure 3. Following Logan and Parman (2017a), we define counties to be urban if more than 25 percent of the population lived in an IPUMS-defined urban area, which are cities and incorporated areas with more than 2,500 residents.

York during the Great Famine (Anbinder, 2001; Handlin, 1959). Moreover, Dutch rural segregation in the mid-19th century was near that of Southern and Eastern European urban segregation between 1900 and 1910. Overall, rural America in the 19th century was highly segregated.

Based on the national trends by ethnicity in Figure 3, segregation in urban areas mostly trended with segregation in rural areas. At face value, this suggests that segregation trends during the Age of Mass Migration reflect factors other than just urban phenomenon. Yet Figure 3 also shows that urban segregation increased more rapidly than rural segregation between 1880 and 1910 for many Southern and Eastern European sources. An increase in urban segregation for these groups may reflect factors specific to these ethnicities, such as their higher participation in factory employment or the effects of mass transit allowing immigrants and natives to sort into different areas. Differential trends in rural and urban areas may also simply reflect a mechanical relationship where inflows were more likely to locate in urban areas and a higher proportion of recently arrived immigrants leads to a higher segregation level. Ultimately more research is needed to explore these trends across rural and urban areas; rather the primary insight from Figure 3 is that segregation was often higher in rural areas than in urban areas in the 19th century.

E. *The most segregated areas in America, 1850-1940*

Urban phenomena clearly did play a role in residential patterns since some of the most highly segregated towns in our dataset were factory towns. Table 1 lists the cities with the highest segregation levels by year for ethnicities which had over 1,000 household heads in town – thus, the list includes both major and minor cities. The most highly segregated cities across the entire 1850 to 1940 period were not the major entry points of New York and Boston, but rather textiles towns; for example, the Irish in 1850 Lowell and Lawrence, Massachusetts were highly segregated. One of the most segregated ethnicities and cities in the entire dataset

is Austro-Hungarians in 1900 Passaic, New Jersey (0.908). Other highly segregated manufacturing towns were Bridgeport, Connecticut, and Buffalo, New York. Yet manufacturing hubs do not completely dominate the list of most segregated ethnicities and towns. Chinese immigrants in 1880 San Francisco were also highly segregated—the highest level of segregation of all cities and years in Table 1 (0.919), perhaps indicating that discriminatory factors led to significant segregation given the anti-Chinese sentiment that led to the 1882 Exclusion Act.

The major entry ports are largely absent from the list of the most segregated cities in Table 1. This may be surprising since new arrivals were often the most segregated. To uncover the segregation level of larger cities, which has been the dominant interest of the literature, we limit the sample to cities with a sizeable ethnic population in Table 2.²³ Based on this list of large cities, the Irish in Boston were the most highly segregated ethnicity between 1850 and 1880, reflecting those fleeing the Great Famine and its aftermath. Yet even the Irish in mid-19th century Boston were not as highly segregated as the Irish in the small factory towns outside of Boston, as we saw from Table 1. For example, the level of segregation for Irish in 1850 Boston was 0.692, while it was 0.801 in Lowell in the same year. At the opposite end, the English in New York City had a *negative* level of segregation in 1860, indicating that they were more likely to live next to a US-born household than next to a foreign-born household from Germany or Ireland.

Some of the highest segregation levels in large cities between 1850 and 1940 were for “new” source immigrants in the early 20th century rather than “old” sources during the 19th century. This is consistent with evidence from Philadelphia in the 19th and 20th centuries, but our data broadens the result to other cities (Hershberg et al., 1981). But the high levels of

²³ We keep cities with more than 10,000 immigrant households from an ethnicity, except for 1850, which we limit to more than 8,000 households. We adopt a lower threshold in 1850 since the migrant stock was smaller.

segregation did not persist long into the 20th century; after the immigration quotas were enacted in the 1920s, New York City almost entirely fell off the list of most segregated cities. Instead of the standard Northeastern and Midwestern cities dominating the list, Mexicans in El Paso topped the list in 1930, reflecting the changing composition of arrivals due to the quotas. Besides these entry points of New York and El Paso, several large cities in the Midwest were highly segregated, such as Germans in Cincinnati, Saint Louis and Chicago, and immigrants from Poland/Russia in 1900 Chicago.

The most highly segregated urban areas were smaller towns associated with manufacturing, but how did segregation in these factory towns compare to segregation in rural areas? Table 3 lists the most segregated rural counties between 1850 and 1940 (here a rural county has less than 25 percent of the population in an urban area). Classic ethnic rural communities appear on this list, such as the Dutch in 1860 Ottawa County, Michigan, where the town of Holland is located. Norwegian farming communities in Minnesota and Wisconsin also top the list during Norway's high inflow periods between 1860 and 1880. In fact, the most highly segregated ethnicity in a rural county was Norwegians in Otter Trail County, Minnesota at 0.722 – more segregated from native-born households than the 1850 Irish in Boston.

Not all highly segregated rural counties were associated with farming; in fact, by the turn of the 20th century, the most segregated rural counties were in areas associated with coal mining and steel production in Western Pennsylvania. These counties topped the list between 1900 and 1940, including Somerset, Indiana, Fayette and Westmoreland County. Segregation in these rural counties was so high that it rivalled that of New York and Boston. Besides these mining areas in the Northeast, mining and agriculture in the American Southwest also led to high segregation levels for Mexicans in New Mexico and California.

F. Robustness of segregation trends

In the appendix, we gauge the robustness of these segregation patterns to alternative measures. One potential issue with the segregation measure is that it is based on household heads, which therefore misses those in non-households or non-household heads. This is nontrivial since many immigrants lived as boarders in houses or in mining or railroad camps, or had native-born spouses. In Appendix C, we present alternate national trends based on the proportion of adult native-born on a census page, which includes all individuals older than 18, rather than just household heads. The resulting estimates from this “page-based” measure has a correlation of 0.941 with the main household-based measure. Therefore, the results from the page-based method are consistent with most results from the neighbor-based measure; for example, the relative levels and trends by country of birth are similar, as well as the levels and trends across rural and urban areas.

Another approach one could take is to change the out-group from the native born to those from *any other* country of birth. Our approach of measuring segregation from the native born is like that of Lieberman (1963); however, an immigrant population highly segregated from natives may be more integrated with immigrants from other sources. In Appendix D, we present results on segregation from any other country of birth, which show mostly the same trends as our preferred measure; the correlation between segregation from the native-born and segregation from all other countries is 0.837. However, Eastern European segregation does depend on the out-group since they were highly segregated from the native-born, but relatively integrated with immigrants from other countries. It is possible that there were stages of spatial assimilation for some sources where one first lived near fellow countrymen, then near immigrants from other sources, and then near the native born. It is also possible that for Eastern Europeans, segregation by country of birth is a poor measure of ethnic group since they do not coincide well with linguistic group.

IV. The Spatial Assimilation of Immigrants

The national trends from the aggregated neighbor-based segregation measures show that segregation tended to rise and fall with inflows during the 20th century, especially following World War I and the immigration quotas. This suggests that immigrants arrived highly segregated but then eventually moved out of the enclave as they became more socially assimilated; however, it could also be that those highly segregated returned home and therefore the overall segregation level fell. In this section, we estimate the rate at which immigrants moved closer to the native born with individual-level data that follows 1900-1919 arrivals for up to 20 years after arrival.²⁴ The individual-level data is advantageous since we can simply use the indicator variable for whether a next-door household head is native born, rather than the aggregated measure to the county level; thus we are able to capture spatial assimilation due to movements within the county – key to the Park-Burgess (1925) model of spatial assimilation due to movement from the center to the outer rings of the city. We also use the fraction of adults on the page that are native born, which has the advantage of including both non-household heads and non-relatives in the segregation measure.²⁵ Both measures lead to the same qualitative results.

The longitudinal data takes the population of 18- to 40-year-old European males in the 1910 and 1920 censuses who arrived in the last ten years, and then tracks them ten years later to the next census.²⁶ The final dataset includes 103,392 male immigrants linked from 1910 to

²⁴ Our approach to estimating spatial assimilation closely mirrors that of Vigdor (2010), who shows with 1900-1930 census data that the fraction of foreign born in the city ward lowers as immigrants stay more years in the country. However, this result may be biased if temporary migrants were more likely to live in wards with a high fraction of foreign born. We improve on this work by using longitudinal data rather than repeated cross sections, which eliminates the possibility of selective return migration driving an increase in spatial assimilation (Abramitzky et al., 2014). Further, we can measure spatial integration at a much finer level (at the census page rather than city ward).

²⁵ Another reason why we measure spatial assimilation with the native-born composition on the page instead of our main segregation measure is because our main measure captures segregation from the native-born, which is not applicable to the native born. In other words, the in-group would be the same as the out-group for the native born. Without a segregation outcome for the native born, we cannot account for decadal and aging effects as in a traditional assimilation regression. Nevertheless, we show the raw means of county-level segregation measures for immigrants in the longitudinal data in Table A4. The initial segregation level drops by about one-third to one-fourth from decade to decade, which is consistent with the general argument that spatial assimilation was not rapid in the first two decades of stay.

²⁶ Those who arrived in the same year as the census are excluded since it does not capture the full cohort.

1920, and 113,799 linked from 1920-1930. The data was created based on machine-learning techniques from James Feigenbaum (2016) and was first presented in detail in Ward (2018). Given that linking tends to produce non-random samples, the sample is weighted to be representative on observables according to the census (Bailey et al., 2017).²⁷ To estimate the rate of spatial assimilation, or the rate at which immigrants converged to natives in neighborhood composition, we pool the immigrant panel with a one percent random samples of male natives from the 1910 to 1930 censuses and from the same birth cohorts. When calculating the fraction of native born on the census page for an individual, we leave out that individual so that we do not mechanically have a gap in the fraction of native born on the page between immigrants and natives.

Table 4 splits the 1900-1919 arrivals in the panel data into five-year cohorts, and shows that recent arrivals arrived highly segregated from the native-born population. For example with 1905-1909 arrivals, only 34 percent of adults on the same page were native-born. Alternatively, 41 percent of next-door household heads were native born. After starting at this low point, immigrants were more likely to live near native-born neighbors by the next decade; for 1905-1909 arrivals, the fraction of native-born neighbors increased from 34 percent to 45 percent. Importantly, since we have a panel, this increase over time is not driven by the selective return of those with fewer native-born neighbors. Overall the data suggests that immigrants spatially integrated with natives after more years of duration, but not by much. In fact, the increase was far less when measuring segregation from native-born with native-born

²⁷ After linking, the linked sample is weighted to be representative based on the predicted likelihood of being in the linked sample, following the methodology of Bailey et al. (2017). It is weighted to be representative with respect to the ending census since this census contains the same population (i.e., migrants who stayed at least ten years); if we weighted the linked sample to be representative with respect to the first census, then the sample would be weighted to reflect observables of temporary migrants and permanent migrants. This weighting regression is done for each ethnicity and weights by age, region of residence, literacy, marital status, ability to speak English, year of arrival, and occupational group (farmer, professional, sales, semi-skilled, low-skilled service, and laborer). The final weights are scaled to match the country of birth distribution in the cross section, similar to Abramitzky, Boustan and Eriksson (2014).

parents; instead of going from 34 to 45 percent, the fraction of 3rd-generation natives increased from 17 to 23 percent.

Immigrants were more likely to live near native-born individuals in the decades after arrival, but even after 20 years of stay, they still had quite different neighborhood compositions than the average native-born male. Since over 90 percent of native-born males had a native-born neighbor, about double the number for immigrants, the gap between immigrants and natives in the fraction of native-born in the neighborhood was large. The gap between natives and immigrants decreased from 50 percentage points to 41 percentage points for the 1905-1909 cohort between the 1910 and 1920 censuses, or by about 20 percent. Of course, part of the reason why the gap is so large is because immigrants located in different areas of the country than natives, but a sizable gap remains even within county: while the across county gap in fraction of native-born on the page was 50 percentage points for the 1905-1909 arrivals, the within county gap was 36 percentage points.²⁸

We can further use the panel data in combination with repeated cross sections to gauge selection into return migration; since the panel contains only permanent migrants and the cross sections contain both permanent and temporary migrants, the difference between the panel and cross sections recovers characteristics of temporary migrants (Abramitzky et al., 2014).²⁹ Figure 4 plots the assimilation profile for both the panel and for a repeated cross section using a standard assimilation specification in the literature (Borjas, 1985).³⁰ The spatial assimilation

²⁸ See Figure A3 and Table A6.

²⁹ We append a 1 percent sample of foreign-born male *household heads* aged 18 to 50 from the 1910-1930 censuses to measure the likelihood that the next-door neighbor is native born. We also append a 1 percent sample of foreign-born *males* (not just household heads) aged 18 to 50 for measuring the fraction of the page that is native born. We only keep immigrants from the same countries of birth as the panel, which primarily drops immigrants from Asia, Canada, and Mexico.

³⁰ To estimate the assimilation profile, we run the following regression separately for the panel and the repeated cross sections:

$$y_{ict} - \widehat{y}_{ict} = f(yrsusa_{it}) + \gamma_c + \varepsilon_{ict}$$

where y_{ict} is the fraction of adults on the census page that are native born for individual i in arrival cohort c in year t . We also use an indicator variable for whether one of the next-door household heads is native born. The variable \widehat{y}_{ict} is the predicted likelihood of having a native-born neighbor based on an auxiliary regression of y_{ict} on age and year fixed effects using a sample of only native-born individuals. When controlling for geography, we

profile clearly shows the slow convergence in neighborhood composition between immigrants and natives over the first twenty years of stay. Based on the comparison of the panel and cross sectional data, return migrants came from more highly segregated neighborhoods since the cross section estimates a larger gap at arrival than the panel (58.4 percentage points versus 50.1 percentage points). Negative selection into return migration on the native-born composition of the neighborhood is consistent with evidence that return migrants were negatively selected on occupational status, English fluency and skill (Abramitzky et al., 2014; Ward, 2017; Ward, 2018). According to Panel B, the magnitude of negative selection into return migration was roughly constant across the 1900 to 1919 cohorts since the gap between the repeated cross-sectional estimates and panel estimates is also roughly constant. However, immigrants who came in the 1910s arrived more integrated with the native born than for those entering in the 1900s, perhaps because there were lower inflows during the 1910s due to travel interruptions from World War I.

While immigrants on average arrived highly segregated from natives and did not converge at a quick rate, this masks heterogeneity by source country in Figure 5. Consistent with the county-level segregation measure, the individual-level panel data shows that Southern and Eastern Europeans arrived the most highly segregated, while those from Northern Europeans and England arrived less segregated. While the size of the initial gaps varied across source countries, the convergence of gaps was similar across sources such that there was little to no closure after 16 to 20 years of stay. Overall, the evidence from the panel data confirms that immigrants' experience in the United States during the early 20th century was distinct from that of natives, despite immigrants having a similar level of occupational status and

also include state and county fixed effects. Therefore $y_{ict} - \widehat{y}_{ict}$ is the gap between natives and immigrants in neighborhood composition, where a negative number indicates that foreigners have fewer native born on the same page. We model the gap in spatial outcomes as a function of a 4th-order polynomial function of years in the United States, and cohort of arrival as fixed effects for five-year groups (i.e., 1900-1904; 1905-1909; 1910-1914; 1915-1919). We plot an estimated profile for the 1900-1904 cohort in Figure 4 Panel A, and the cohort effects in Panel B.

assimilating quickly in terms of English proficiency, intermarriage and Anglicization of names (Abramitzky et al., 2014; Abramitzky et al., 2016; Biavaschi, 2017; Ward, 2018).

V. Conclusions

In this paper, we document the ethnic segregation of immigrants between 1850 and 1940 based on the nativity of the next-door neighbor. Our measure adapts the method first introduced by Logan and Parman (2017a) to immigrants. The neighbor-based segregation measure reveals several new results, such as the high levels of rural segregation during the 19th century and for Chinese and Mexican immigrants. It also provides several comparisons that are consistent over time and space, such as the result that rural Norwegians were more segregated than the urban Irish in 1850. Further, it shows that the most segregated areas were smaller factory towns, places that were mostly ignored by the broader literature since there were no city wards to calculate a traditional dissimilarity or isolation index. While the neighbor-based measure broadens our knowledge on segregation by covering more areas and time periods, it does not overturn prior results from city ward/census tract-based studies, such as the decrease in segregation during the early 20th century and that “new” sources tended to more segregated during the early 20th century than “old” sources in the 19th century (Cutler et al., 2008a; Hershberg et al. 1981; Lieberson, 1963).

Our primary aim is to present the measure the broad segregation patterns between 1850 and 1940. By limiting ourselves to a birds-eye view of segregation, we do not explore the rich detail for specific ethnicities, cities, counties or time periods. For instance, there is little knowledge about the causes and consequences of segregation during the high immigration period prior to the Civil War, when German and Irish immigrants arrived after fleeing famine and political violence. More research could be done on the effects of segregation; for example, one could estimate how social and economic assimilation depended upon on arriving in a highly segregated neighborhood, or the effect of segregation on subsequent generations’

outcomes.³¹ These effects could vary by rural and urban communities as well. One could also relate the measure to the public economics literature, for example by exploring how public good provision, such as the quality of schools or the implementation of mass transit, was related to ethnic segregation. Given the extensive detail in the newly digitized census manuscripts, there is much to explore.

³¹ For example, Logan and Parman (and co-authors) use their segregation measure to pursue an extensive research agenda on black-white segregation, including estimating the association between segregation and lynching (Cook, Logan and Parman, 2017), home ownership (2017b), mortality (2017c) and present-day intergenerational mobility (Andrews et al., 2017).

References

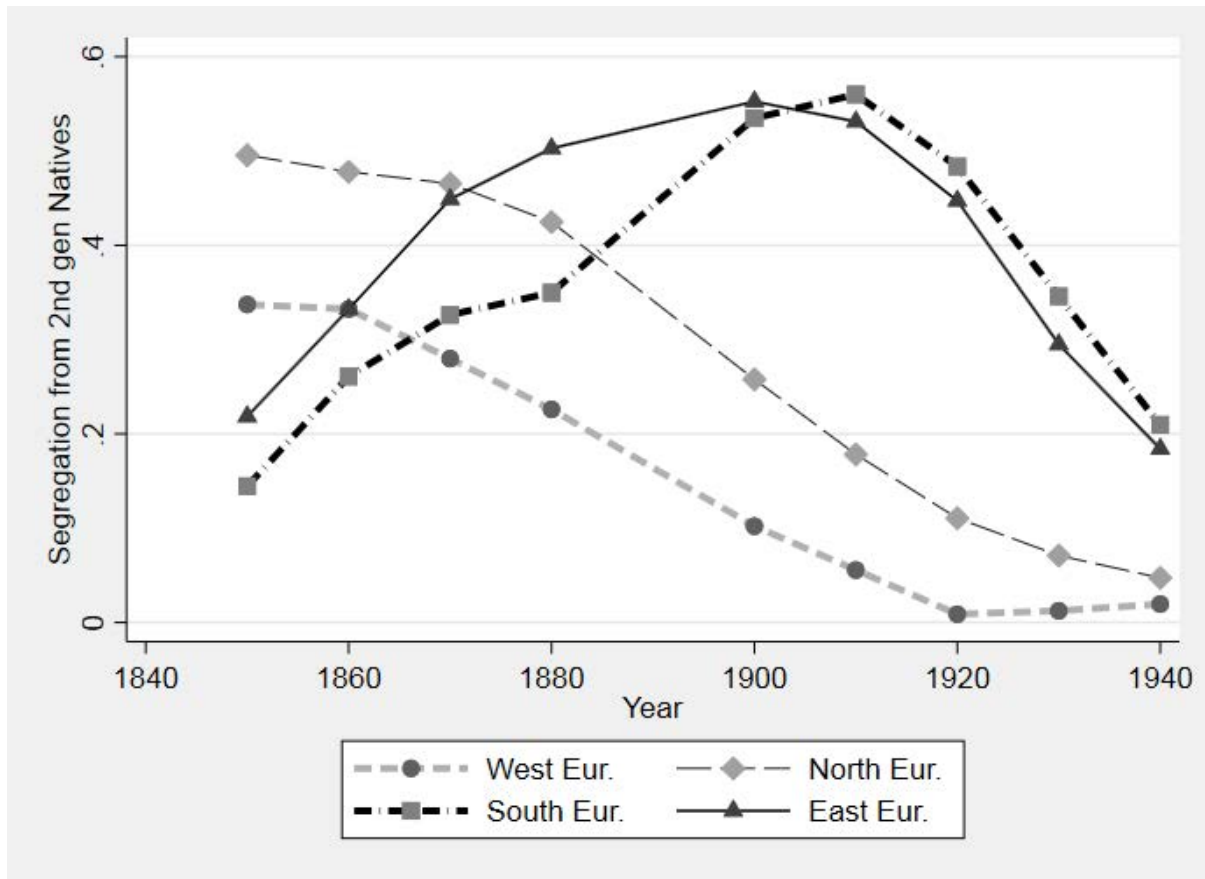
- Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. "A nation of immigrants: Assimilation and economic outcomes in the age of mass migration." *Journal of Political Economy* 122.3 (2014): 467-506.
- Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. *Cultural assimilation during the age of mass migration*. No. w22381. National Bureau of Economic Research, 2016.
- Agresti, Barbara F. "Measuring Residential Segregation in Nineteenth-Century American Cities." *Sociological Methods & Research* 8.4 (1980): 389-399.
- Alba, Richard D. *Italian Americans: Into the twilight of ethnicity*. Prentice Hall, 1985.
- Anbinder, Tyler. *Five Points: The 19th-century New York City neighborhood that invented tap dance, stole elections, and became the world's most notorious slum*. Simon and Schuster, 2001.
- Andrews, R., Casey, M., Hardy, B. L., & Logan, T. D. (2017). Location matters: Historical racial segregation and intergenerational mobility. *Economics Letters*, 158, 67-72.
- Bailey, Martha, et al. *How Well Do Automated Linking Methods Perform in Historical Samples? Evidence from New Ground Truth*. Working Paper, 2017.
- Beaman, Lori A. "Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the US." *The Review of Economic Studies* 79.1 (2011): 128-161.
- Biavaschi, Costanza, Corrado Giulietti, and Zahra Siddique. "The economic payoff of name americanization." *Journal of Labor Economics* 35.4 (2017): 1089-1116.
- Borjas, George J. "Assimilation, changes in cohort quality, and the earnings of immigrants." *Journal of labor Economics* 3.4 (1985): 463-489.
- Cook, Lisa D., Trevon D. Logan, and John M. Parman. *Racial segregation and southern lynching*. No. w23813. National Bureau of Economic Research, 2017.
- Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. "The rise and decline of the American ghetto." *Journal of political economy* 107.3 (1999): 455-506.
- Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. "Is the melting pot still hot? Explaining the resurgence of immigrant segregation." *The Review of Economics and Statistics* 90.3 (2008a): 478-497.
- Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. "When are ghettos bad? Lessons from immigrant segregation in the United States." *Journal of Urban Economics* 63.3 (2008b): 759-774.
- Damm, Anna Piil. "Ethnic enclaves and immigrant labor market outcomes: Quasi-experimental evidence." *Journal of Labor Economics* 27.2 (2009): 281-314.

- Duncan, Otis Dudley, and Stanley Lieberson. "Ethnic segregation and assimilation." *American Journal of Sociology* 64.4 (1959): 364-374.
- Edin, Per-Anders, Peter Fredriksson, and Olof Åslund. "Ethnic enclaves and the economic success of immigrants—Evidence from a natural experiment." *The quarterly journal of economics* 118.1 (2003): 329-357.
- Eriksson, Katherine. "Ethnic Enclaves and Immigrant Outcomes: Norwegian Immigrants during the Age of Mass Migration". 2018.
- Feigenbaum, James J. "Automated census record linking: A machine learning approach." (2016).
- Hacker, J. David. "New estimates of census coverage in the United States, 1850–1930." *Social Science History* 37.1 (2013): 71-101.
- Handlin, Oscar. *Boston's immigrants, 1790-1880: A study in acculturation*. Harvard University Press, 1959.
- Hershberg, Theodore. "The Philadelphia social history project: an introduction." *Historical Methods Newsletter* 9.2-3 (1976): 43-58.
- Hershberg, Theodore, et al. "A Tale of Three Cities: Blacks, Immigrants and Opportunity in Philadelphia." *Philadelphia: Work, Space, Family, and Group Experience in the 19th Century: Essays Toward an Interdisciplinary History of the City* (1981): 461-91.
- Kantrowitz, Nathan. "Racial and ethnic residential segregation in Boston 1830-1970." *The Annals of the American Academy of Political and Social Science* 441.1 (1979): 41-54.
- Lieberson, Stanley. *Ethnic Patterns in American Cities*. New York: The Free Press of Glencoe, 1963.
- Lieberson, Stanley. *A Piece of the Pie: Blacks and White Immigrants since 1880*. Berkeley: University of California Press, 1980.
- Logan, John R., and Matthew J. Martinez. "The spatial scale and spatial configuration of residential settlement: Measuring segregation in the postbellum South." *American Journal of Sociology* 123.4 (2018): 1161-1203.
- Logan, John R., and Hyung-jin Shin. "Birds of a feather: social bases of neighborhood formation in Newark, New Jersey, 1880." *Demography* 53.4 (2016): 1085-1108.
- Logan, John R., and Weiwei Zhang. "White ethnic residential segregation in historical perspective: US cities in 1880." *Social science research* 41.5 (2012): 1292-1306.
- Logan, Trevon D., and John M. Parman. "The national rise in residential segregation." *The Journal of Economic History* 77.1 (2017a): 127-170.

- Logan, Trevon D., and John M. Parman. "Segregation and Homeownership in the Early Twentieth Century" *American Economic Review, Papers and Proceedings* 107.5 (2017b): 410-414.
- Logan, Trevon D., and John M. Parman. "Segregation and mortality over time and space." *Social Science & Medicine* (2017c).
- Massey, Douglas S., and Nancy A. Denton. "The dimensions of residential segregation." *Social forces* 67.2 (1988): 281-315.
- Munshi, Kaivan. "Networks in the modern economy: Mexican migrants in the US labor market." *The Quarterly Journal of Economics* 118.2 (2003): 549-599.
- Park, Robert E., and Ernest W. Burgess. *The City*. Chicago: The University of Chicago Press, 1925.
- Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. *Integrated Public Use Microdata Series: Version 7.0* [dataset]. Minneapolis: University of Minnesota, 2017. <https://doi.org/10.18128/D010.V7.0>.
- Shertzer, Allison, Randall P. Walsh, and John R. Logan. "Segregation and neighborhood change in northern cities: New historical GIS data from 1900–1930." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 49.4 (2016): 187-197.
- Spielman, Seth E., and John R. Logan. "Using high-resolution population data to identify neighborhoods and establish their boundaries." *Annals of the Association of American Geographers* 103.1 (2013): 67-84.
- Thernstrom, Stephan. *The other Bostonians: Poverty and progress in the American metropolis, 1880-1970*. Harvard Univ Pr, 1973.
- Vigdor, Jacob L. *From Immigrants to Americans: The Rise and fall of fitting In*. Rowman & Littlefield, 2010.
- Ward, Zachary. "Birds of passage: Return migration, self-selection and immigration quotas." *Explorations in Economic History* 64 (2017): 37-52.
- Ward, Zachary. "Have language skills always been so valuable? The low return to English fluency during the Age of Mass Migration", 2018.
- Wildsmith, Elizabeth, Myron P. Gutmann, and Brian Gratton. "Assimilation and intermarriage for US immigrant groups, 1880–1990." *The History of the Family* 8.4 (2003): 563-584.
- White, Michael J., Robert F. Dymowski, and Shilian Wang. "Ethnic neighbors and ethnic myths: An examination of residential segregation in 1910." *After Ellis Island: Newcomers and natives in the 1910 Census*: 175-208.

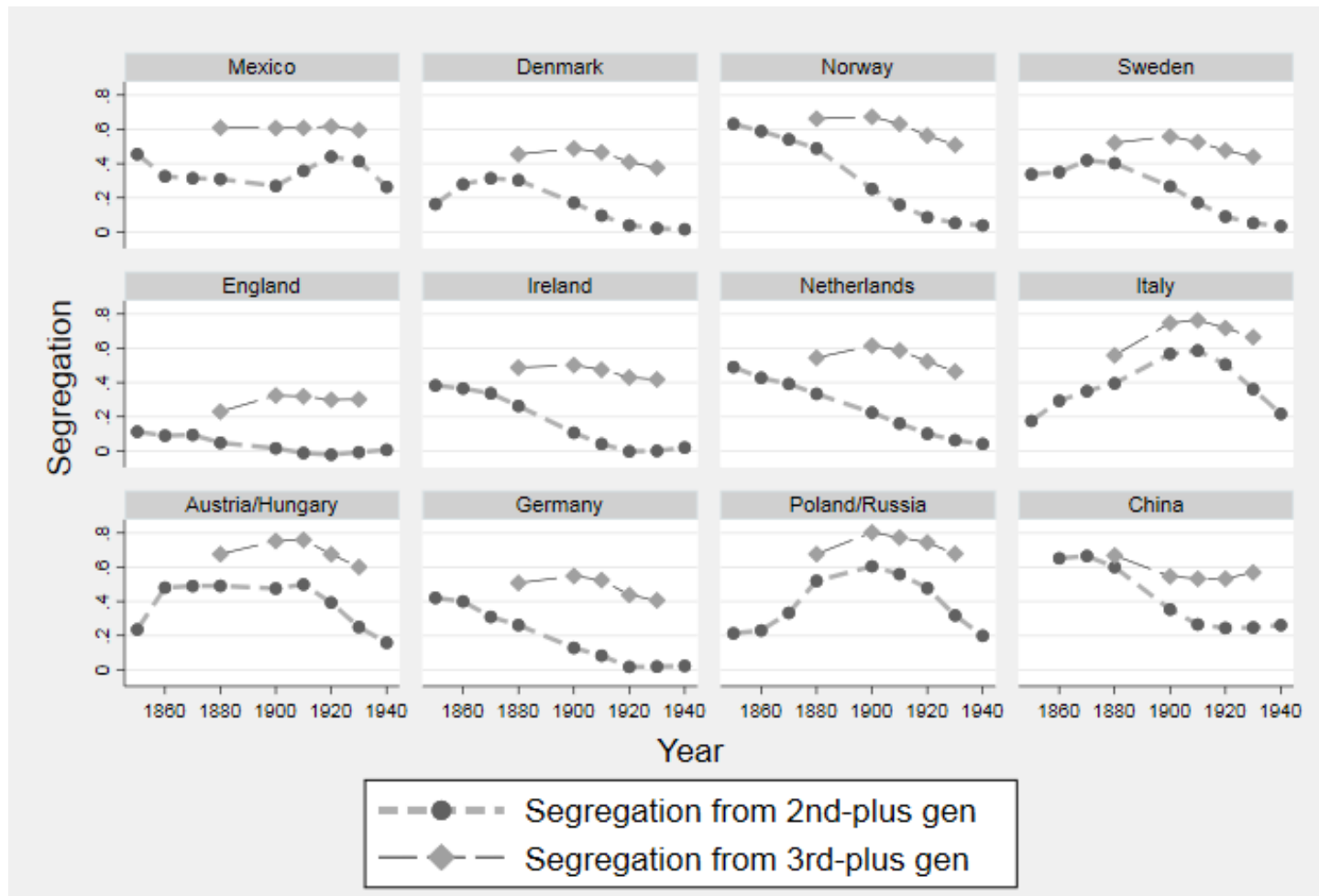
Zunz, Olivier. *The changing face of inequality: Urbanization, industrial development, and immigrants in Detroit, 1880-1920*. University of Chicago Press, 1982.

Figure 1. National Segregation Trends by Source Region, 1850 to 1940



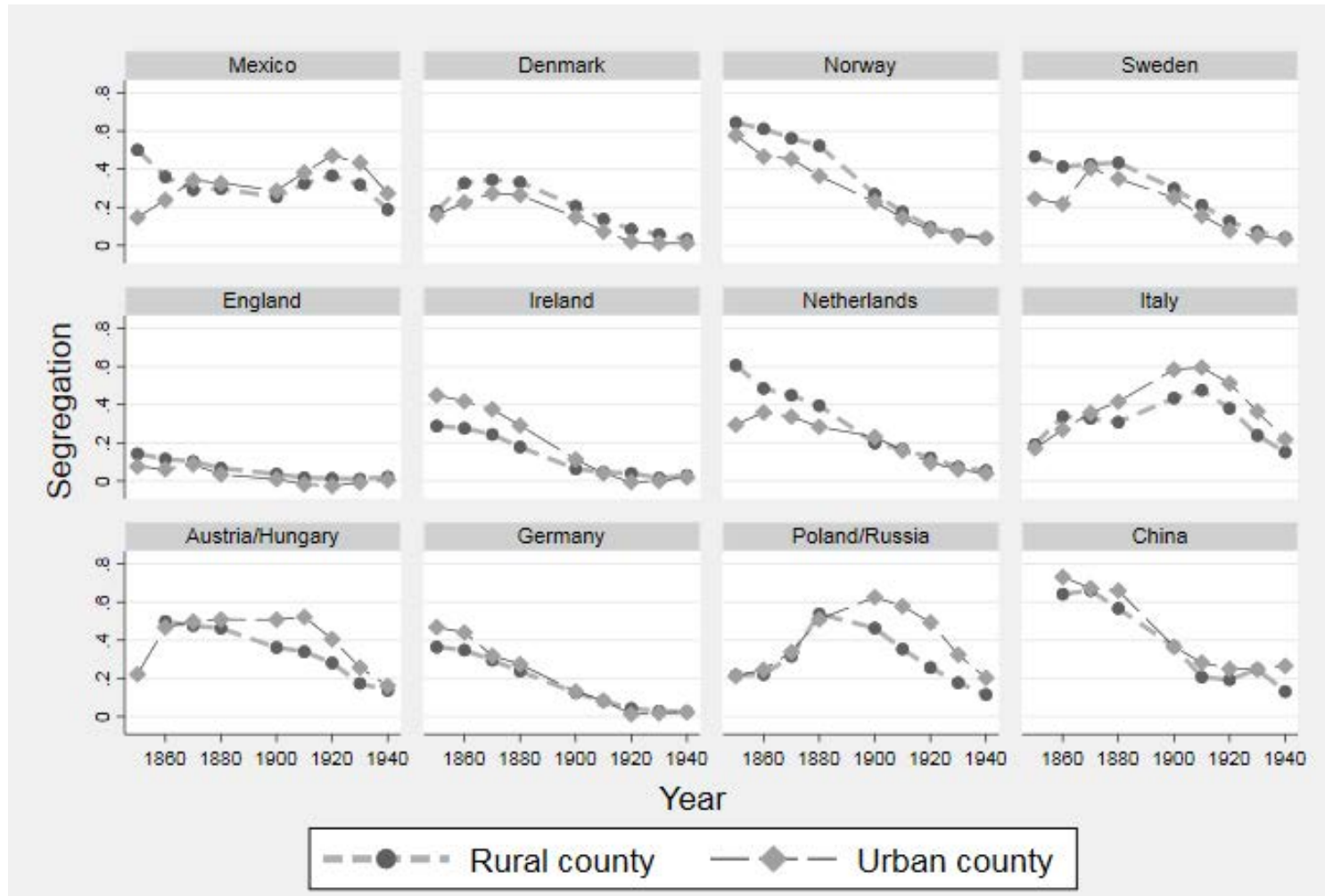
Notes: Data is from the 1850 to 1940 full-count censuses. Segregation measure calculated at county and country of birth level and then aggregated to national level after weighting by the number of households in the county/country of birth. Western Europe includes England, Scotland, Wales, Ireland, Belgium, France, Luxembourg, Netherlands, Switzerland and Germany. Northern Europe includes Denmark, Finland, Iceland, Norway, and Sweden. Southern Europe includes Albania, Greece, Italy, Malta, Portugal, and Spain. Eastern Europe includes Austria/Hungary (includes Czechoslovakia and Yugoslavia), and Russia/Poland (includes Estonia, Latvia and Lithuania).

Figure 2. Immigrants were more segregated from native-born with native-born parents



Notes: Data is from the 1850 to 1940 full-count censuses. Segregation measure calculated at county and country of birth level and then aggregated to national level after weighting by the number of households. 2nd-plus generation are US born; 3rd-plus generation are US-born to two US-born parents. Austria/Hungary includes Austria, Hungary, Czechoslovakia and Yugoslavia; Russia/Poland includes Russia, Poland, Estonia, Latvia and Lithuania.

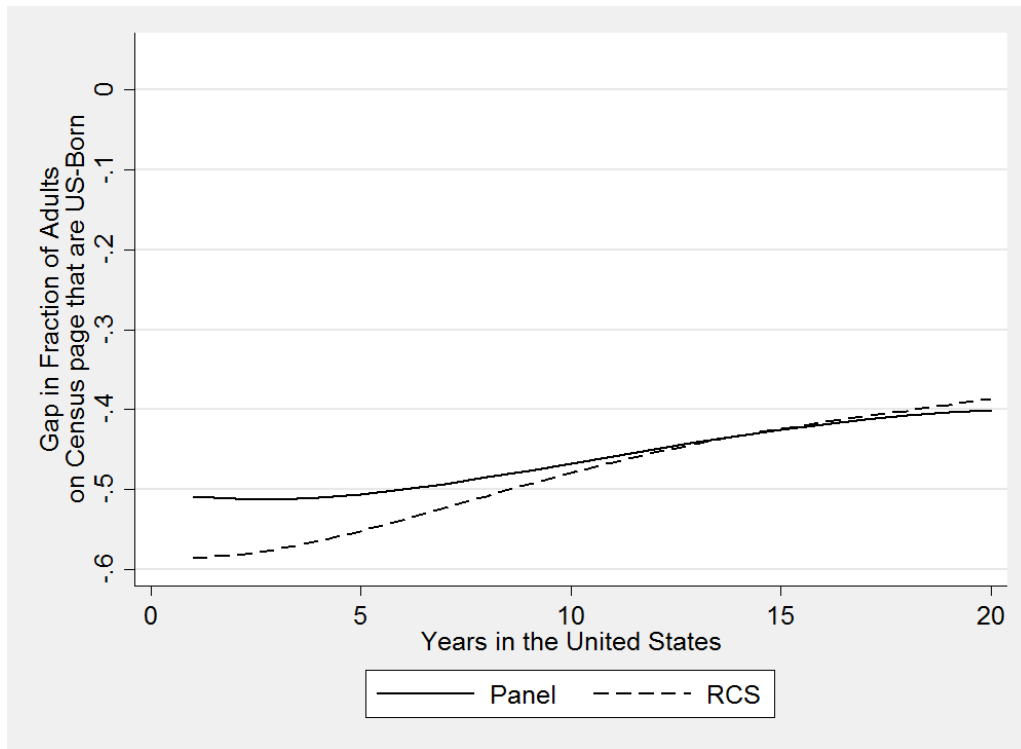
Figure 3. Rural segregation was often higher than urban segregation in the 19th century



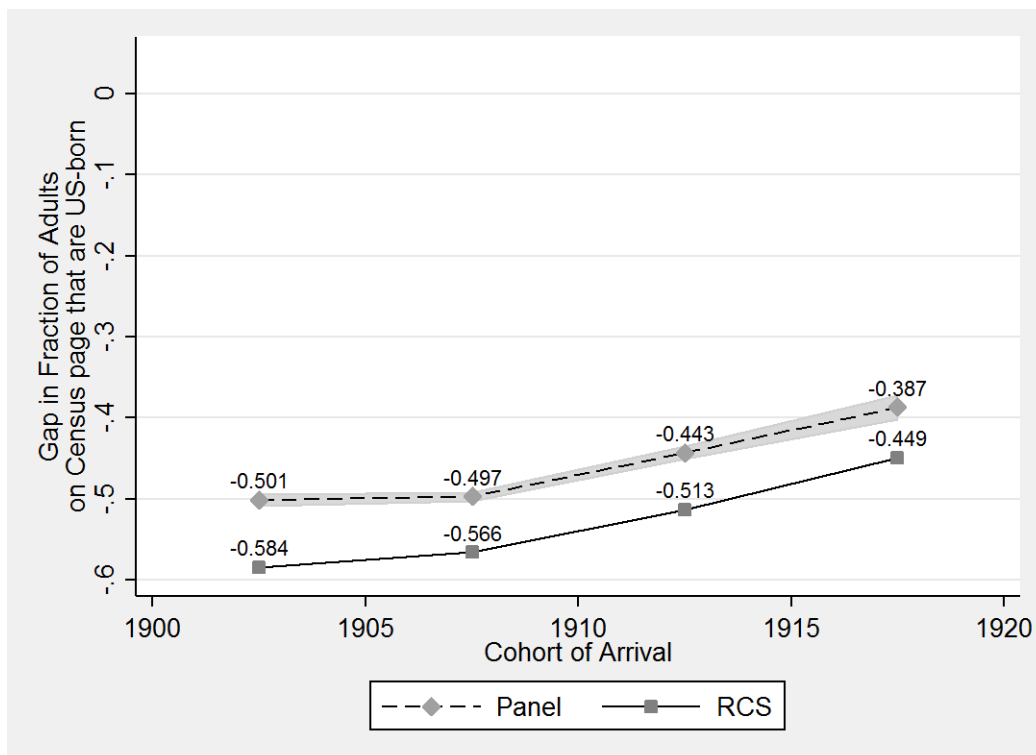
Notes: Data is from the 1850 to 1940 full-count censuses. An urban county is defined as having at least 25 percent of the population living in an urban area, or an incorporated area/town with more than 2,500 people.

Figure 4. Spatial Assimilation in the decades after arrival

Panel A: Assimilation profile for 1900-1904 cohort

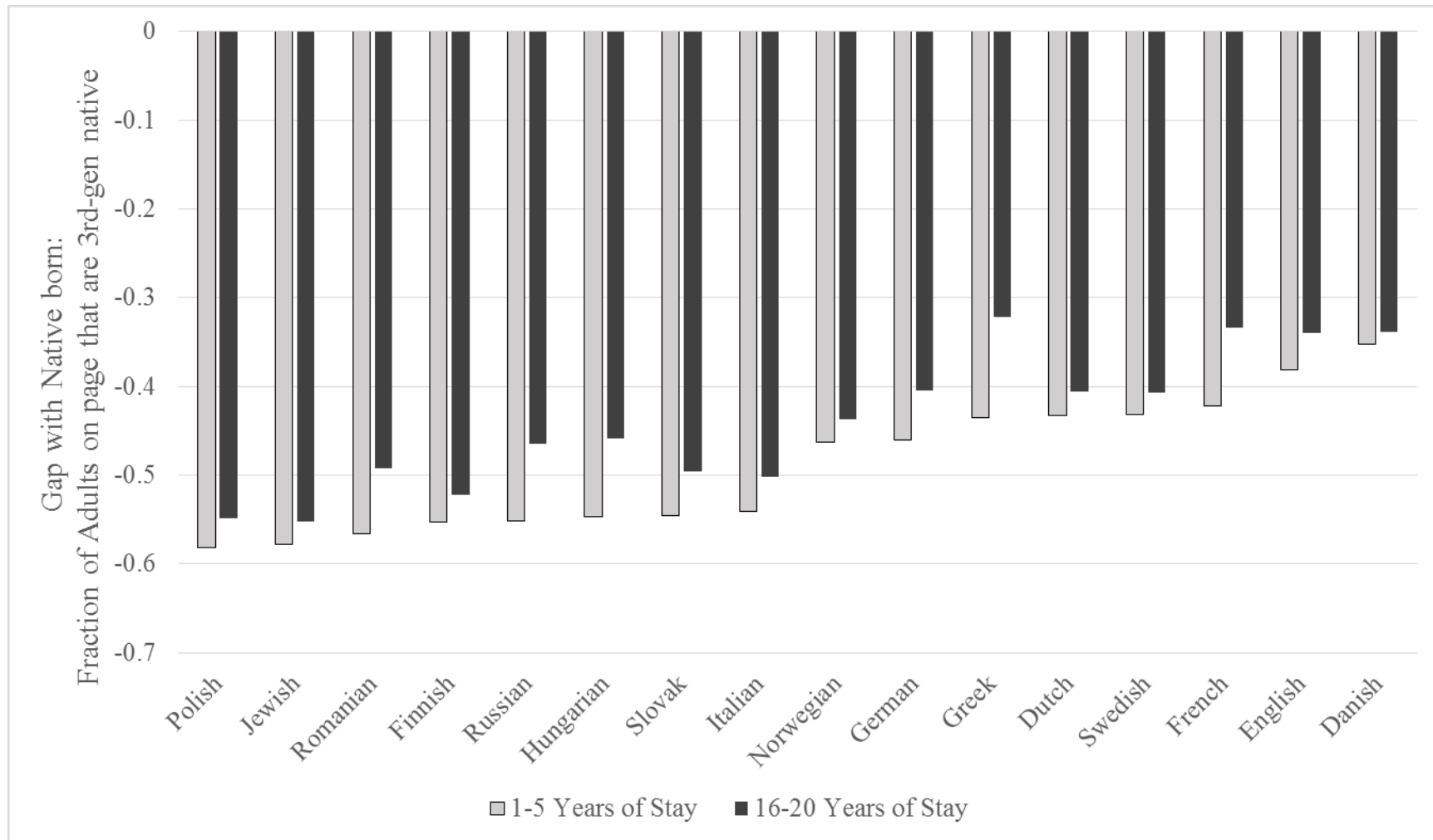


Panel B. Cohort Effects for 1900 to 1919 arrivals



Notes: Data is from linked samples between the 1910-1920 census and 1920-1930 census. Panel A plots results for the 1900 to 1904 cohort. Panel B plots predicted gap with natives at arrival, with 95 percent confidence intervals shaded. See Table A5 for underlying regression coefficients and for other measures of spatial assimilation.

Figure 5. Spatial Assimilation by Ethnicity between 1910 and 1920



Notes: Data is from the linked panel between 1910 and 1920. The figure plots the raw means in the likelihood that a neighbor is native born to two native-born parents, after correcting for age and period fixed effects with natives. The figure is split by ethnicity, which is measured by the mother tongue variable in the 1920 census.

Table 1. Top segregated ethnicities and cities with over 1,000 households

City	Ethn.	Seg	N of HH	City	Ethn.	Seg	N of HH	City	Ethn.	Seg	N of HH
1850				1860				1870			
Lowell, MA	Ireland	0.801	1,584	Lowell, MA	Ireland	0.721	2,631	Chicago, IL	Aus/Hgy	0.728	1,992
Buffalo, NY	Germany	0.714	2,806	Lawrence, MA	Ireland	0.718	1,448	San Fran, CA	China	0.726	1,449
Boston, MA	Ireland	0.692	8,769	Worcester, MA	Ireland	0.680	1,565	Chicago, IL	Sweden	0.676	1,656
Cincinnati, OH	Germany	0.638	9,016	Roxbury, MA	Ireland	0.660	2,096	Manchester, NH	Ireland	0.664	1,025
Roxbury, MA	Ireland	0.625	1,096	Boston, MA	Ireland	0.648	14,296	Worcester, MA	Ireland	0.663	2,979
Providence, RI	Ireland	0.598	1,888	Charlestown, MA	Ireland	0.637	1,478	Lawrence, MA	Ireland	0.642	2,260
Saint Louis, MO	Germany	0.564	5,782	Portland, ME	Ireland	0.630	1,163	Lowell, MA	Ireland	0.631	2,772
Chicago, IL	Germany	0.559	1,176	Kingston, NY	Ireland	0.622	1,071	Fall River, MA	Ireland	0.628	1,834
Milwaukee, WI	Germany	0.556	1,565	Providence, RI	Ireland	0.618	2,823	Milford, MA	Ireland	0.611	1,002
1880				1900				1910			
San Fran, CA	China	0.919	1,787	Passaic, NJ	Aus/Hgy	0.908	1,084	Passaic, NJ	Aus/Hgy	0.856	2,548
Chicago, IL	Aus/Hgy	0.761	4,314	Buffalo, NY	Pol/Rus	0.855	6,905	Providence, RI	Italy	0.822	3,914
Milwaukee, WI	Pol/Rus	0.729	1,085	San Fran, CA	China	0.793	2,977	Utica, NY	Italy	0.792	1,662
Chicago, IL	Pol/Rus	0.720	2,434	Providence, RI	Italy	0.787	1,407	Lorain, OH	Aus/Hgy	0.790	1,285
Cleveland, OH	Aus/Hgy	0.719	2,360	Boston, MA	Italy	0.786	4,114	Perth Amboy, NJ	Aus/Hgy	0.786	1,663
Fall River, MA	Canada	0.629	1,203	Bridgeport, CT	Aus/Hgy	0.776	1,112	Braddock, PA	Aus/Hgy	0.785	1,012
Worcester, MA	Ireland	0.616	3,482	Buffalo, NY	Italy	0.742	1,708	South Beth., PA	Aus/Hgy	0.776	1,204
Manchester, NH	Ireland	0.612	1,218	Detroit, MI	Pol/Rus	0.740	5,384	Boston, MA	Italy	0.764	8,801
New York, NY	Italy	0.585	3,562	Toledo, OH	Pol/Rus	0.740	1,776	Bridgeport, CT	Aus/Hgy	0.763	2,652
1920				1930				1940			
Niagara Fls, NY	Pol/Rus	0.803	1,514	Brawley, CA	Mexico	0.710	1,005	San Bern., CA	Mexico	0.430	1,059
Lawrence, MA	Italy	0.793	2,406	Lawrence, MA	Italy	0.673	2,933	El Paso, TX	Mexico	0.427	8,991
Lowell, MA	Greece	0.784	1,184	San Bern., CA	Mexico	0.663	1,171	San Fran, CA	China	0.423	3,599
Providence, RI	Italy	0.773	7,179	Rome, NY	Italy	0.641	1,104	Auburn, NY	Italy	0.409	1,003
Binghamton, NY	Aus/Hgy	0.733	1,126	El Paso, TX	Mexico	0.592	11,086	Lawrence, MA	Italy	0.388	2,908
Lakewood, OH	Aus/Hgy	0.732	1,162	Worcester, MA	Italy	0.578	2,063	Chicago, IL	Mexico	0.371	3,242
Worcester, MA	Italy	0.729	1,522	Chicago, IL	Mexico	0.577	3,609	Norristown, PA	Italy	0.357	1,067
Amsterdam, NY	Pol/Rus	0.716	1,613	Seattle, WA	Japan	0.556	1,549	Dallas, TX	Mexico	0.343	1,026
Lorain, OH	Aus/Hgy	0.701	2,051	Niagara Falls, NY	Italy	0.547	2,271	Worcester, MA	Italy	0.342	2,071

Notes: Data is from the 1850-1940 censuses. The table lists the city, country of birth, segregation level and number of household heads for each ethnicity.

Table 2. Top segregated ethnicities and cities with a large population

City	Ethn.	Seg	N of HH	City	Ethn.	Seg	N of HH	City	Ethn.	Seg	N of HH
1850				1860				1870			
Boston	Ireland	0.692	8,769	Boston	Ireland	0.648	14,296	Boston	Ireland	0.572	18,811
Cincinnati	Germany	0.638	9,016	Cincinnati	Germany	0.567	16,195	Chicago	Germany	0.475	13,382
New York	Ireland	0.540	37,462	New York	Germany	0.510	49,880	Chicago	Ireland	0.425	14,078
New York	Germany	0.492	16,663	Saint Louis	Germany	0.499	14,294	New York	Ireland	0.409	93,773
Philadelphia	Ireland	0.447	12,595	New York	Ireland	0.454	77,453	Saint Louis	Germany	0.397	13,285
New York	England	0.015	8,128	Baltimore	Germany	0.451	12,398	Cincinnati	Germany	0.356	10,205
				Philadelphia	Ireland	0.404	19,463	New York	Germany	0.345	34,309
				New York	England	-0.042	13,255	Philadelphia	Ireland	0.333	32,250
								Saint Louis	Ireland	0.305	11,181
1880				1900				1910			
Boston	Ireland	0.506	24,400	Chicago	Pol/Rus	0.734	27,028	New York	Pol/Rus	0.681	138,763
Chicago	Germany	0.450	28,167	New York	Pol/Rus	0.702	56,989	New York	Italy	0.677	103,350
Milwaukee	Germany	0.355	12,112	New York	Italy	0.673	44,389	Philadelphia	Italy	0.673	12,851
New York	Germany	0.309	91,306	Philadelphia	Pol/Rus	0.621	10,233	Boston	Pol/Rus	0.665	11,565
Saint Louis	Germany	0.295	23,085	Chicago	Aus/Hgy	0.591	19,577	Chicago	Italy	0.659	13,216
Buffalo	Germany	0.272	10,799	New York	Aus/Hgy	0.583	35,650	Chicago	Pol/Rus	0.620	32,381
Baltimore	Germany	0.269	15,758	Chicago	Sweden	0.271	20,074	Philadelphia	Pol/Rus	0.613	25,146
New York	Ireland	0.255	99,688	Boston	Ireland	0.258	24,856	New York	Aus/Hgy	0.546	77,526
Cincinnati	Germany	0.250	21,699	Detroit	Germany	0.252	13,573	Cleveland	Aus/Hgy	0.541	21,500
1920				1930				1940			
Boston	Italy	0.695	14,543	El Paso	Mexico	0.592	11,086	Boston	Pol/Rus	0.326	16,562
New York	Pol/Rus	0.591	231,314	Rochester	Italy	0.516	10,292	Rochester	Italy	0.315	10,724
Newark	Italy	0.582	11,508	Boston	Italy	0.499	16,598	Philadelphia	Italy	0.306	30,517
Chicago	Pol/Rus	0.575	98,333	Los Angeles	Mexico	0.477	16,287	Los Angeles	Mexico	0.295	15,707
Boston	Pol/Rus	0.574	18,189	Philadelphia	Italy	0.446	30,586	Boston	Italy	0.286	16,307
New York	Italy	0.572	152,338	Boston	Pol/Rus	0.429	21,045	Philadelphia	Pol/Rus	0.278	44,154
Chicago	Italy	0.568	22,355	Chicago	Italy	0.408	33,994	Detroit	Pol/Rus	0.266	36,660
Philadelphia	Italy	0.567	23,105	Cleveland	Italy	0.407	10,585	New York	Pol/Rus	0.260	281,424
Detroit	Pol/Rus	0.567	30,308	San Antonio	Mexico	0.403	11,315	Cleveland	Italy	0.251	10,713

Notes: Data is from the 1850-1940 censuses. The table lists the city, country of birth, segregation level and number of household heads for each ethnicity.

Table 3. Top segregated ethnicities and rural counties with over 1,000 households

County	Ethn.	Seg	N of HH	County	Ethn.	Seg	N of HH	County	Ethn.	Seg	N of HH
1850				1860				1870			
Schuylkill, PA	Ireland	0.682	2,643	Dane, WI	Norway	0.678	1,391	Winneshiek, IA	Norway	0.618	1,507
Wash., WI	Germany	0.635	2,265	Luzerne, PA	Ireland	0.626	4,282	Dane, WI	Norway	0.600	1,883
Luzerne, PA	Ireland	0.635	1,499	Ottawa, MI	Neth.	0.617	1,016	Fillmore, MN	Norway	0.569	1,752
Ulster, NY	Ireland	0.521	1,382	F. D. Lac, WI	Germany	0.613	1,370	Stearns, MN	Germany	0.564	1,560
St Clair, IL	Germany	0.460	1,644	Dodge, WI	Germany	0.609	2,761	Goodh., MN	Norway	0.562	1,136
Hartford, CT	Ireland	0.427	1,220	Clayton, IA	Germany	0.598	1,192	Ottawa, MI	Neth.	0.560	1,817
St Lawr., NY	Ireland	0.421	1,685	Auglaize, OH	Germany	0.573	1,266	Carbon, PA	Ireland	0.549	1,327
Onondaga, NY	Ireland	0.420	1,768	Sheboygan, WI	Germany	0.555	2,423	Henry, IL	Sweden	0.543	1,464
1880				1900				1910			
Otter Tail, MN	Norway	0.723	1,184	Westmoreland, PA	Aus/Hgy	0.714	2,227	Somerset, PA	Aus/Hgy	0.709	1,175
Vernon, WI	Norway	0.653	1,187	Fayette, PA	Aus/Hgy	0.675	1,654	Indiana, PA	Aus/Hgy	0.693	1,295
Windham, CT	Canada	0.647	1,455	Marion, KS	Pol/Rus	0.675	1,165	Fayette, PA	Aus/Hgy	0.590	5,433
Trempeal., WI	Norway	0.636	1,307	Hutchinson, SD	Pol/Rus	0.508	1,133	Morton, ND	Pol/Rus	0.564	1,233
Fillmore, MN	Norway	0.591	1,814	Wright, MN	Swe.	0.460	1,174	Graham, AZ	Mexico	0.554	2,095
Buffalo, WI	Germany	0.588	1,015	Graham, AZ	Mexico	0.455	1,066	Norfolk, MA	Italy	0.544	1,112
Freeborn, MN	Norway	0.559	1,004	McPherson, KS	Swe.	0.453	1,077	Marion, KS	Pol/Rus	0.436	1,060
Goodhue, MN	Sweden	0.558	1,265	Kent, RI	Canada	0.432	1,359	Grant, NM	Mexico	0.395	1,118
1920				1930				1940			
Somerset, PA	Aus/Hgy	0.683	2,196	Pinal, AZ	Mexico	0.566	1,207	Somerset, PA	Aus/Hgy	0.326	1,556
Pinal, AZ	Mexico	0.632	1,253	Somerset, PA	Aus/Hgy	0.487	1,561	Indiana, PA	Aus/Hgy	0.250	1,345
Indiana, PA	Italy	0.604	1,460	Indiana, PA	Aus/Hgy	0.362	1,352	Sullivan, NY	Pol/Rus	0.231	1,480
Greenlee, AZ	Mexico	0.596	1,853	Sullivan, NY	Pol/Rus	0.336	1,448	Oxford, ME	Canada	0.197	1,497
Indiana, PA	Aus/Hgy	0.585	2,046	Grant, NM	Mexico	0.309	1,094	Merced, CA	Portugal	0.193	1,181
Clearfield, PA	Pol/Rus	0.530	1,026	Dona Ana, NM	Mexico	0.300	2,039	Fayette, PA	Aus/Hgy	0.191	5,119
Norfolk, MA	Italy	0.527	2,323	Merced, CA	Portugal	0.298	1,121	Suffolk, NY	Italy	0.170	1,941
Clearfield, PA	Aus/Hgy	0.524	1,820	San Patricio, TX	Mexico	0.285	1,363	Fayette, PA	Pol/Rus	0.163	1,621

Notes: Data is from the 1850-1940 censuses. The table lists the county, country of birth, segregation level and number of household heads for each ethnicity. This is based on counties that have at most 25 percent of the population in an urban area (>2,500 residents).

Table 4. Spatial Assimilation using longitudinal data

	1910	1920	1930	Change over decade	N
<i>Panel A: Fraction of Page 2nd generation</i>					
Native-Born	0.858	0.867	0.872		
Foreign-born Year of Arrival					
1900-1904	0.369	0.469		0.099	50,385
1905-1909	0.338	0.446		0.108	53,007
1910-1914		0.437	0.533	0.096	100,641
1915-1919		0.479	0.545	0.066	13,158
<i>Panel B: Fraction of Page 3rd generation</i>					
Native-Born	0.678	0.679	0.677		
Foreign-born Year of Arrival					
1900-1904	0.184	0.237		0.053	50,385
1905-1909	0.172	0.228		0.056	53,007
1910-1914		0.229	0.259	0.030	100,641
1915-1919		0.266	0.285	0.018	13,158
<i>Panel C: Have a 2nd-gen Next-Door Neighbor</i>					
Native-Born	0.903	0.909	0.911		
Foreign-born Year of Arrival					
1900-1904	0.406	0.505		0.099	42,120
1905-1909	0.391	0.494		0.103	39,043
1910-1914		0.479	0.565	0.086	87,207
1915-1919		0.535	0.615	0.080	10,295
<i>Panel D: Has a 3rd-gen Next-Door Neighbor</i>					
Native-Born	0.781	0.784	0.776		
Foreign-born Year of Arrival					
1900-1904	0.238	0.304		0.066	42,120
1905-1909	0.227	0.296		0.069	39,043
1910-1914		0.297	0.332	0.035	87,207
1915-1919		0.345	0.384	0.039	10,295

Notes: Data are from the 1910-1920 and 1920-1930 linked sample for immigrants from Ward (2018). Data are also from a 1910, 1920, 1930 1 percent random sample of natives. The sample sizes for the native born in Panels A and B are 181,464 in 1910, 210,324 in 1920, and 253,841 in 1930. The sample sizes for the native born in Panels C and D are 97,980, 118,238, and 146,079.

Online Appendix, not for publication.

Table A1. Segregation from 2nd-generation by country of birth

Country	Year								
	1850	1860	1870	1880	1900	1910	1920	1930	1940
Canada	0.144	0.118	0.146	0.120	0.078	0.038	0.010	0.008	0.009
Mexico	0.455	0.325	0.314	0.309	0.269	0.357	0.441	0.414	0.264
Cuba		-0.114	0.129	0.053	0.249	0.289	0.177	0.189	0.136
Denmark	0.163	0.279	0.314	0.303	0.171	0.096	0.039	0.022	0.016
Finland				0.510	0.563	0.497	0.398	0.278	0.163
Norway	0.632	0.590	0.541	0.489	0.252	0.159	0.086	0.053	0.039
Sweden	0.337	0.350	0.419	0.402	0.267	0.171	0.090	0.052	0.035
England	0.112	0.089	0.094	0.048	0.015	-0.012	-0.022	-0.008	0.006
Scotland	0.130	0.112	0.108	0.059	0.018	-0.009	-0.025	0.000	0.006
Ireland	0.383	0.365	0.337	0.263	0.107	0.041	-0.003	0.000	0.020
Belgium	0.395	0.331	0.362	0.310	0.211	0.193	0.145	0.107	0.071
France	0.261	0.278	0.202	0.165	0.075	0.069	0.051	0.048	0.039
Netherlands	0.490	0.427	0.392	0.334	0.224	0.161	0.101	0.062	0.041
Switzerland	0.362	0.351	0.268	0.220	0.100	0.059	0.029	0.031	0.028
Greece				0.155	0.185	0.370	0.293	0.203	0.139
Italy	0.175	0.293	0.349	0.395	0.568	0.586	0.505	0.361	0.217
Portugal	0.092	0.333	0.402	0.350	0.369	0.312	0.398	0.318	0.202
Spain	0.127	0.104	0.124	0.062	0.105	0.328	0.320	0.304	0.208
Austria/Hungary	0.236	0.481	0.490	0.491	0.476	0.499	0.393	0.250	0.159
Germany	0.421	0.400	0.310	0.262	0.129	0.083	0.017	0.019	0.023
Poland/Russia	0.213	0.230	0.333	0.520	0.605	0.559	0.478	0.318	0.199
China		0.652	0.666	0.601	0.353	0.265	0.244	0.247	0.261
Japan					0.694	0.608	0.442	0.399	
Turkey					0.274	0.411	0.390	0.279	0.204

Notes: Data is from the 1850 to 1940 full-count censuses. See Figure 2 for graphical depiction for 12 selected countries.

Table A2. Segregation from 3rd-generation by country of birth

Country	1880	1900	1910	1920	1930
Canada	0.120	0.389	0.380	0.355	0.352
Mexico	0.309	0.608	0.609	0.618	0.597
Cuba	0.053	0.550	0.577	0.482	0.488
Denmark	0.303	0.488	0.467	0.409	0.376
Finland	0.510	0.774	0.748	0.693	0.652
Norway	0.489	0.674	0.632	0.563	0.509
Sweden	0.402	0.559	0.526	0.477	0.440
England	0.048	0.324	0.319	0.299	0.302
Scotland	0.059	0.330	0.324	0.297	0.315
Ireland	0.263	0.502	0.475	0.430	0.418
Belgium	0.310	0.589	0.548	0.478	0.425
France	0.165	0.403	0.398	0.381	0.364
Netherlands	0.334	0.615	0.588	0.523	0.464
Switzerland	0.220	0.451	0.416	0.361	0.338
Greece	0.155	0.498	0.587	0.518	0.462
Italy	0.395	0.748	0.764	0.718	0.665
Portugal	0.350	0.586	0.614	0.669	0.648
Spain	0.062	0.427	0.566	0.545	0.558
Austria/Hungary	0.491	0.753	0.760	0.676	0.601
Germany	0.262	0.549	0.525	0.438	0.407
Poland/Russia	0.520	0.804	0.773	0.744	0.679
China	0.601	0.546	0.532	0.533	0.570
Japan		0.746	0.707	0.595	0.573
Turkey		0.544	0.615	0.616	0.565

Notes: Data is from the 1850 to 1940 full-count censuses. See Figure 2 for graphical depiction for 12 selected countries.

Table A3. Rural and Urban country segregation

Country		Year								
		1850	1860	1870	1880	1900	1910	1920	1930	1940
Canada	Rural	0.162	0.128	0.115	0.078	0.048	0.036	0.025	0.006	0.033
Canada	Urban	0.082	0.090	0.181	0.161	0.089	0.038	0.008	0.008	0.007
Mexico	Rural	0.501	0.360	0.291	0.298	0.254	0.324	0.366	0.319	0.188
Mexico	Urban	0.147	0.239	0.342	0.328	0.288	0.382	0.472	0.434	0.274
Denmark	Rural	0.181	0.327	0.345	0.332	0.206	0.137	0.084	0.058	0.033
Denmark	Urban	0.159	0.225	0.273	0.264	0.148	0.073	0.020	0.010	0.012
Finland	Rural				0.559	0.573	0.519	0.429	0.360	0.212
Finland	Urban				0.357	0.560	0.490	0.390	0.263	0.156
Norway	Rural	0.644	0.612	0.562	0.523	0.268	0.176	0.095	0.059	0.038
Norway	Urban	0.578	0.467	0.455	0.365	0.229	0.143	0.080	0.050	0.039
Sweden	Rural	0.467	0.413	0.426	0.434	0.299	0.211	0.126	0.073	0.042
Sweden	Urban	0.246	0.217	0.405	0.350	0.250	0.156	0.079	0.048	0.033
England	Rural	0.143	0.115	0.103	0.068	0.036	0.017	0.013	0.010	0.021
England	Urban	0.077	0.062	0.087	0.034	0.010	-0.017	-0.026	-0.009	0.005
Scotland	Rural	0.158	0.135	0.131	0.088	0.059	0.050	0.030	0.023	0.017
Scotland	Urban	0.096	0.087	0.090	0.039	0.006	-0.020	-0.032	-0.001	0.006
Wales	Rural	0.383	0.344	0.318	0.253	0.098	0.055	0.027	0.033	0.008
Wales	Urban	0.155	0.216	0.276	0.208	0.085	0.014	-0.010	-0.018	-0.009
Ireland	Rural	0.289	0.277	0.244	0.178	0.063	0.046	0.040	0.016	0.030
Ireland	Urban	0.449	0.418	0.376	0.291	0.113	0.040	-0.005	0.000	0.019
France	Rural	0.293	0.296	0.198	0.157	0.080	0.104	0.058	0.044	0.047
France	Urban	0.232	0.262	0.205	0.171	0.074	0.063	0.050	0.048	0.039
Netherlands	Rural	0.607	0.485	0.450	0.396	0.199	0.167	0.121	0.074	0.055
Netherlands	Urban	0.295	0.360	0.337	0.284	0.232	0.159	0.096	0.060	0.039
Switzerland	Rural	0.371	0.342	0.264	0.210	0.115	0.080	0.047	0.039	0.033
Switzerland	Urban	0.347	0.364	0.272	0.229	0.092	0.052	0.025	0.029	0.027
Italy	Rural	0.191	0.339	0.328	0.309	0.435	0.476	0.381	0.240	0.151
Italy	Urban	0.171	0.270	0.357	0.416	0.585	0.595	0.512	0.364	0.219
Portugal	Rural		0.431	0.353	0.328	0.298	0.114	0.341	0.236	0.161
Portugal	Urban	0.115	0.234	0.419	0.361	0.378	0.362	0.406	0.321	0.204
Austria/Hungary	Rural		0.499	0.477	0.463	0.362	0.340	0.281	0.174	0.137
Austria/Hungary	Urban	0.222	0.467	0.498	0.509	0.508	0.522	0.406	0.257	0.160
Germany	Rural	0.365	0.349	0.295	0.238	0.124	0.081	0.043	0.029	0.024
Germany	Urban	0.469	0.441	0.319	0.274	0.130	0.084	0.012	0.017	0.023
Poland/Russia	Rural	0.216	0.221	0.316	0.538	0.463	0.354	0.257	0.177	0.115
Poland/Russia	Urban	0.212	0.246	0.339	0.509	0.626	0.579	0.493	0.325	0.203
China	Rural		0.642	0.661	0.567	0.366	0.207	0.193	0.248	0.131
China	Urban		0.732	0.674	0.660	0.366	0.283	0.250	0.247	0.265
Japan	Rural					0.410	0.664	0.370	0.260	
Japan	Urban					0.298	0.542	0.451	0.407	

Notes: Data is from the 1850-1940 Censuses. The table shows the highest segregation levels for cities and ethnicities that have over 1,000 households. We drop values if they have less than 4,000 households in total in an urban or rural area.

Table A4. Spatial Assimilation using Segregation Measure

	1910	1920	1930	Change over Decade	N
Raw County-Level Segregation Measure					
Foreign-born Cohort of Arrival					
1900	0.383	0.302		-0.080	50,385
1905	0.386	0.306		-0.080	53,007
1910		0.314	0.216	-0.099	100,641
1915		0.253	0.180	-0.073	13,158

Notes: Data are from the 1910-1920 and 1920-1930 linked sample for immigrants from Ward (2018). The data reports the raw means of the main segregation measure in the panel data, merging at the county level.

Table A5. Spatial Assimilation regression estimates

Data Structure:	Fraction of page 2nd gen		Fraction of page 2nd gen		Next-door HH is 2nd gen		Next-door HH is 3rd gen	
	Panel	RCS	Panel	RCS	Panel	RCS	Panel	RCS
Years in US	-0.00867 (0.00563)	-0.00431*** (0.000507)	-0.0120** (0.00503)	-0.0110*** (0.000403)	-0.0257* (0.0155)	-0.0612*** (0.00187)	-0.0171 (0.0195)	-0.0534*** (0.00168)
Years in US sq	0.00197* (0.00103)	0.00303*** (9.31e-05)	0.00236** (0.000938)	0.00307*** (7.61e-05)	0.00473* (0.00287)	0.0102*** (0.000308)	0.00363 (0.00356)	0.00883*** (0.000276)
Years in US cub	-8.81e-05 (6.86e-05)	-0.000193*** (6.51e-06)	-0.000146** (6.35e-05)	-0.000200*** (5.45e-06)	-0.000264 (0.000204)	-0.000562*** (1.99e-05)	-0.000208 (0.000235)	-0.000495*** (1.79e-05)
Years in US quad	1.19e-06 (1.53e-06)	3.86e-06*** (1.53e-07)	2.99e-06** (1.45e-06)	4.14e-06*** (1.30e-07)	5.22e-06 (4.99e-06)	1.06e-05*** (4.40e-07)	3.99e-06 (5.28e-06)	9.40e-06*** (3.98e-07)
Arrival Cohort 1900-1904	-0.114*** (0.00381)	-0.135*** (0.000533)	-0.0737*** (0.00356)	-0.0861*** (0.000479)	-0.144*** (0.00837)	-0.140*** (0.00140)	-0.113*** (0.00801)	-0.106*** (0.00133)
Arrival Cohort 1905-1909	-0.110*** (0.00273)	-0.117*** (0.000499)	-0.0728*** (0.00243)	-0.0754*** (0.000449)	-0.135*** (0.00582)	-0.119*** (0.00133)	-0.110*** (0.00506)	-0.0906*** (0.00127)
Arrival Cohort 1910-1914	-0.0560*** (0.00414)	-0.0643*** (0.000531)	-0.0394*** (0.00392)	-0.0423*** (0.000484)	-0.0831*** (0.0152)	-0.0627*** (0.00140)	-0.0678*** (0.0137)	-0.0497*** (0.00134)
Constant (Arrival Cohort 1915-1919)	-0.387*** (0.00767)	-0.449*** (0.000939)	-0.400*** (0.00672)	-0.426*** (0.000767)	-0.332*** (0.0205)	-0.308*** (0.00370)	-0.433*** (0.0241)	-0.377*** (0.00337)
Observations	434,382	5,605,690	434,382	5,605,690	391,137	2,847,670	391,137	2,847,670
R-squared	0.043	0.077	0.013	0.027	0.016	0.021	0.011	0.013

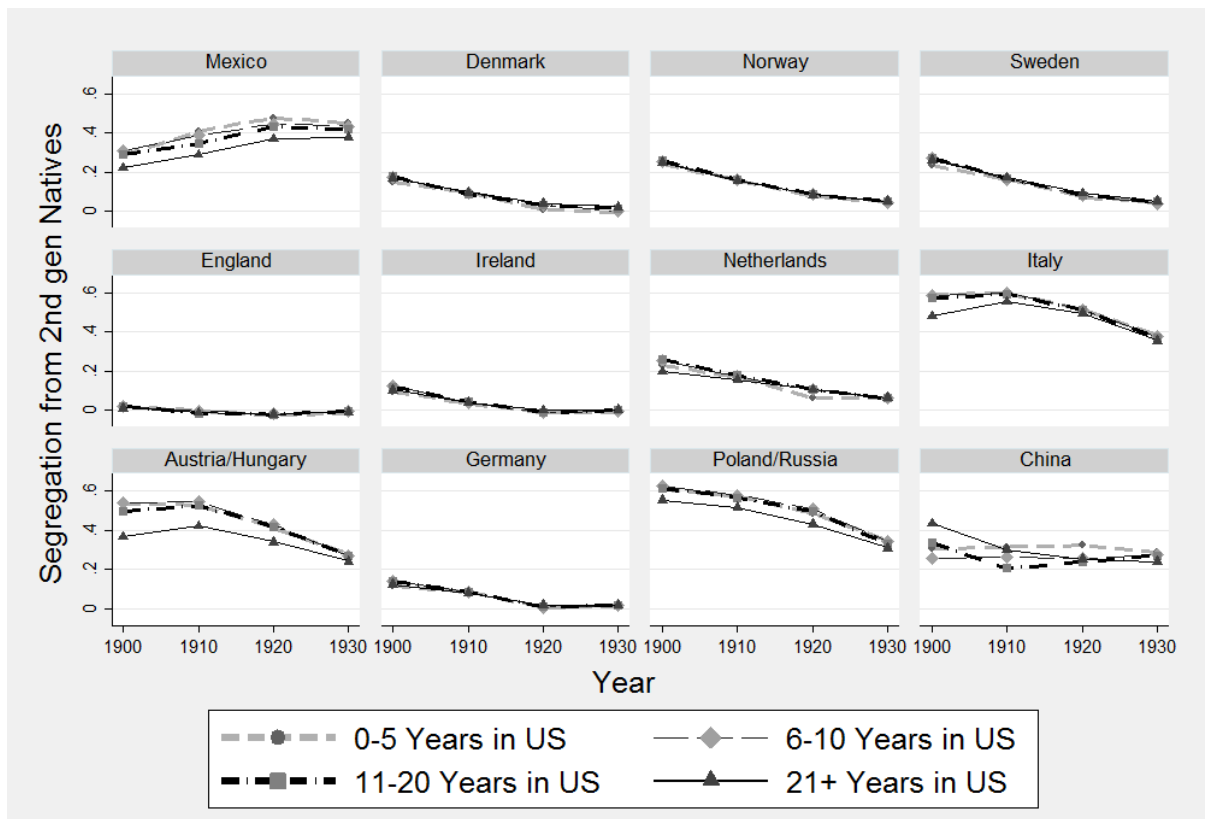
Notes: Data is from 1910-1930 linked sample (panel) and 1 percent sample from 1910-1930 census (RCS or repeated cross section). The dependent variable is the predicted gap between immigrants and natives after accounting for age and year effects.

Table A6. Fraction of native-born on census page when accounting for geography

	Overall	Within State	Within County
Years in US	-0.00867 (0.00563)	-0.00944** (0.00392)	-0.00341 (0.00404)
Years in US sq	0.00197* (0.00103)	0.00232*** (0.000739)	0.00134* (0.000744)
Years in US cub	-8.81e-05 (6.86e-05)	-0.000125** (5.26e-05)	-6.71e-05 (5.10e-05)
Years in US quad	1.19e-06 (1.53e-06)	2.28e-06* (1.26e-06)	1.12e-06 (1.18e-06)
Arrival Cohort 1900-1904	-0.114*** (0.00381)	-0.112*** (0.00396)	-0.104*** (0.00350)
Arrival Cohort 1905-1909	-0.110*** (0.00273)	-0.111*** (0.00256)	-0.107*** (0.00251)
Arrival Cohort 1910-1914	-0.0560*** (0.00414)	-0.0644*** (0.00340)	-0.0657*** (0.00333)
Constant (Arrival Cohort 1915-1919)	-0.387*** (0.00767)	-0.295*** (0.00608)	-0.257*** (0.00613)
Observations	434,382	434,382	434,382
R-squared	0.043	0.043	0.042

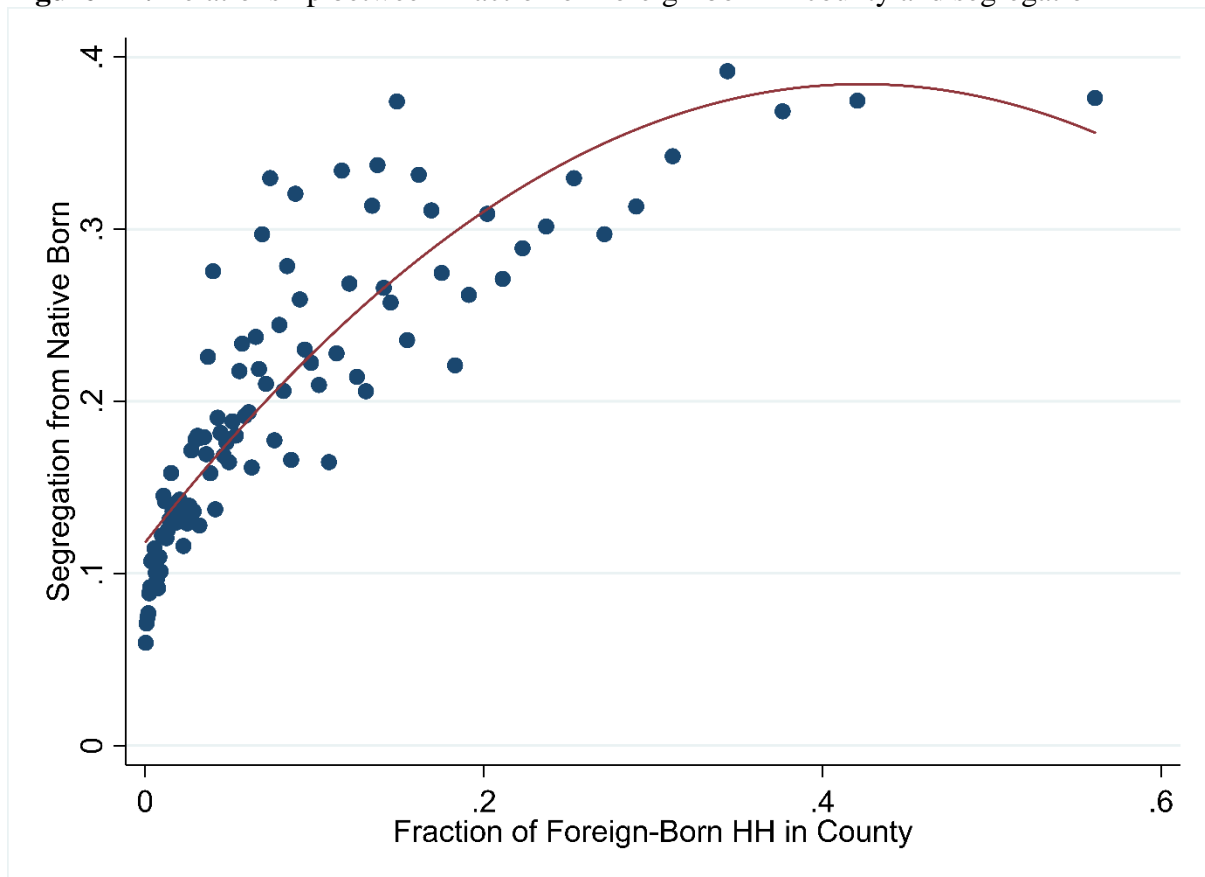
Notes: Data is from 1910-1930 linked sample (panel) and 1 percent sample from 1910-1930 census (RCS or repeated cross section). The dependent variable is the predicted gap between immigrants and natives after accounting for age and year effects in the first column, including state fixed effects in the second columns, and including county fixed effects in the third column. See Figure A3 for estimated profiles.

Figure A1. Segregation by years in the United States, by country of birth.



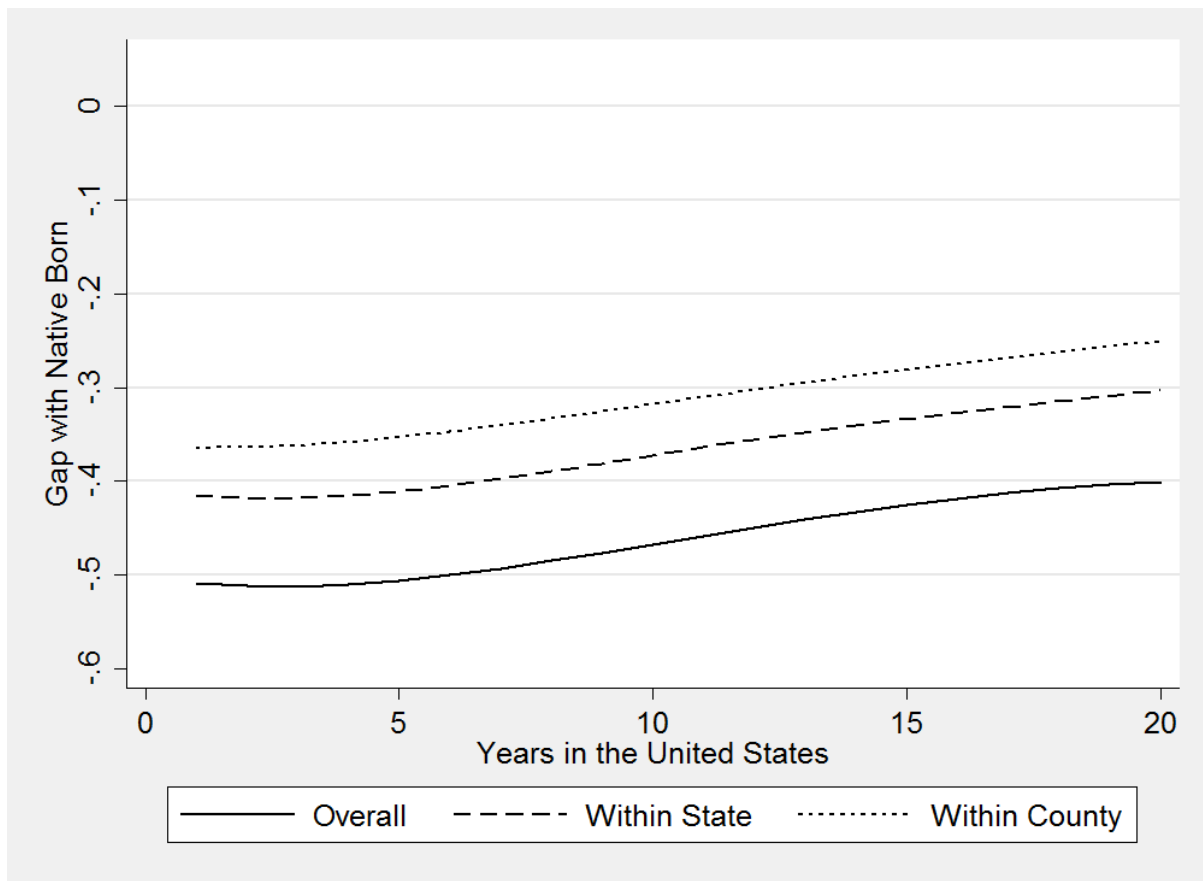
Notes: Data is from the 1850 to 1940 full-count censuses. Segregation is calculated for each group from native-born households. The pattern shows little differences across years in the United States, suggesting little spatial assimilation. Little spatial assimilation is consistent with our estimates with panel data.

Figure A2. Relationship between Fraction of Foreign-born in county and segregation



Notes: Data is from the 1850 to 1940 full-count censuses. This is a bin scatter plot that shows the relationship between fraction of foreign born in county with segregation at the county level. The underlying data is at the county-ethnicity-year level.

Figure A3. Spatial assimilation profiles when accounting for geography



Notes: Data is from the 1850 to 1940 full-count censuses. The within state profile is estimated after controlling for state of residence, and the within county profile is estimated after controlling for county of residence. See Table A6 for underlying coefficients.

Appendix A. Further details on cleaning the data

We use the full-count data between 1850 and 1940 from the University of Minnesota Population Center. At the time of writing this paper, the 1850, 1880 and 1900-1940 censuses have been cleaned; the 1900 to 1940 are cleaned on a preliminary basis.³² Therefore, we need to clean the 1860 and 1870 Censuses ourselves. The primary variables we are interested in cleaning are country of birth, county, city and household head. The process of cleaning the 1860 and 1870 datasets are described in further detail below.

- County of birth

To clean the country of birth strings, we rely heavily on the strings already cleaned by the University of Minnesota Population Center for the 1850, 1880 and 1900 to 1940 full-count data. We create files that yield the most common country of birth codes (BPL) for each country of birth string (BPLSTR).

Armed with these files, we simply merge them to the uncleaned censuses starting with the nearest year – for example, the 1860 uncleaned census to the cleaned BPLSTR codes from 1850. For BPLSTR that are unmatched, we merge them onto later cleaned census files to update the BPL codes. For this process, we merge first to the 1880 or 1850, depending on closeness in time, and then to the 1900 to 1940 Census files. This is because border changes following World War I cause the pre-World War I censuses to be more reliable for assigning BPL codes. However, boundary changes do not bias results in text since we group countries by large region (i.e. Eastern Europe is one group).

After this initial pass, we have cleaned 99 percent of the country of birth strings. Following this, we tabulated a list of strings for each census and cleaned those which appeared more than 100 times. These were more common in the earlier censuses in the mid-19th century when individuals would sometimes list a town or a state within Germany. For country of birth strings which appeared less than 100 times, we left their country (bpl code) as missing and dropped them from the dataset.

- Page indicators

We need to identify all immigrants who live next to each other on the same page. Rather than identify census page by NARA roll, reel and page, we used the codes for image id in the

³² There is some evidence that group quarters variables have some inaccuracies in the full-count data, which may bias the household measure.

uncleaned data to determine whether an individual was on the same page. The image id is a code that Ancestry.com uses that combines string information from roll and page number, so it yields the same information but in one succinct variable. There are some instances in the Censuses where the page information was clearly inaccurate as there were over 50 households listed on a page. On the extreme end, there were 20,000 households listed on the same page in the 1860 Census, a problem that could not be fixed by resorting to information about the NARA roll or page number; however, this is not problematic for our main next-door measure. Moreover, the 1880 census include both sides of the census sheet to be on the same page, yielding of an average of about 100 individuals per sheet rather than the 50 in other censuses. While this does not strongly bias results, it may influence results in our robustness check of a “page-based” measure in Appendix C. Therefore, we sort by serial number and person number to ensure that we are capturing households in order and then create “synthetic pages” the start anew after 50 people.

- Relationship to head

We keep only the head of the household for our main segregation measure, but information about household head prior to the 1880 census was not explicitly listed in the Census. However, family numbers are provided within the raw data, which appears to separate individuals by household and not by nuclear family. Therefore, we keep the first family member listed in the 1860 and 1870 censuses to proxy for the household head.

- Identifying Households and Group Quarters

We do not have institutional or group quarters identifiers for the unclean censuses in 1860 and 1870. IPUMS codes group quarters based on the number of unrelated members in the household, typically if there are more than ten individuals who are unrelated to the household head. For 1860 and 1870, lacking relationship string data, we simply keep the first listed household member and drop households if there are more than twenty individuals in a family number who have different surnames.

- County

We merge the uncleaned county strings with the ICPSR county codes, which we referenced from the IPUMS website. <https://usa.ipums.org/usa/volii/ICPSR.shtml>

- City

For city, we merge the uncleaned strings with the IPUMS city codes. https://usa.ipums.org/usa-action/variables/CITY#codes_section. There are a few times where a city in earlier census years is part of a city in later years; for example, Northern Liberties, PA was coded as a separate city in 1850, but was later a part of Philadelphia. To consistently code cities, we include smaller cities as part of the main city; this occurs for Brooklyn as part of New York City, Georgetown as part of Washington DC, and Kensington, Mayamensing, Northern Liberties, Southwark and Spring Garden as part of Philadelphia.

- Urban

Urban status is not provided in the uncleaned census files. Following Logan and Parman (2017a), we define a county as urban as those with greater than 25 percent of the population living in an urban area, as defined by the IPUMS variable URBAN. We calculate the fraction of a county in an urban area using the 1850 to 1940 IPUMS samples.

- Country groupings

One issue when presenting results by country of birth is that countries change borders over time, especially before and after World War I. We make the following groupings

1. Russia / Poland includes Russia, Poland, Estonia, Latvia and Lithuania
2. Austria / Hungary includes Austria, Hungary and Czechoslovakia

Appendix B. Measuring ethnic segregation

We follow Logan and Parman (2017a) for creating the segregation measure, but we make a few distinct changes to the formulas. The reason why we change the formula is because unlike black-white segregation which has two defined groups (black or white), ethnic segregation has multiple groups (Irish, German, Russian, etc.). Black and white are mutually exclusive sets where the union (mostly) forms the population prior to 1940; however, the union of immigrants from a certain country of birth and the native born do not form the entire population. Yet much of the following discussion closely follows Appendix 1 in Logan and Parman (2017).

The formula we use in the main results to calculate segregation measures is as follows:

$$\eta_c = \frac{E(\overline{native_c}) - native_c}{E(\overline{native_c}) - E(\underline{native_c})} \quad (1)$$

where $E(\overline{native_c})$ is the expected number of immigrants who have a native-born neighbor under random assignment, $native_c$ is the number of immigrants from country c who are observed to have a native-born neighbor, and $E(\underline{native_c})$ is the expected number of immigrants who have a native-born neighbor under complete segregation. Remember that only household heads are included in these numbers. While the expected number of native-born neighbors seems like a straightforward concept, one must adjust for the fact we observe two neighboring households for those in the center of the census manuscript, but only one neighboring household for those at the top or bottom.

Let us define the following variables:

- $n_{c,N=2}$ – number of immigrants from country c with 2 observed neighbors
- $n_{c,N=1}$ – number of immigrants from country c with 1 observed neighbor
- n_{fb} – number of immigrants from all countries
- n_{all} – number of all households in area. Note that $n_{all} - n_{fb}$ is the number of native born

The expected number of immigrants from a country of birth c with a native-born neighbor under random assignment is as follows:

$$\begin{aligned}
E(\overline{\text{native}}_c) &= n_{c,N=2} \cdot p(\text{native neighbor} | N = 2) + n_{c,N=1} \cdot p(\text{native neighbor} | N = 1) \quad (\text{B1}) \\
&= n_{c,N=2} \left(1 - \left(\frac{n_{fb} - 1}{n_{all} - 1} \right) \left(\frac{n_{fb} - 2}{n_{all} - 2} \right) \right) + n_{c,N=1} \left(1 - \left(\frac{n_{fb} - 1}{n_{all} - 1} \right) \right)
\end{aligned}$$

The logic behind the formula is under random assignment and for those with two observed neighbors, the probability of having a foreign-born neighbor on one side is $\left(\frac{n_{fb}-1}{n_{all}-1}\right)$ and the probability of having foreign born neighbors on both sides is $\left(\frac{n_{fb}-1}{n_{all}-1}\right)\left(\frac{n_{fb}-2}{n_{all}-2}\right)$. Since we are interested in the case where an immigrant has at least one native-born neighbor, the probability of this occurring for an immigrant with two neighbors is simply one minus the probability of having two foreign-born neighbors, or $\left(1 - \left(\frac{n_{fb}-1}{n_{all}-1}\right)\left(\frac{n_{fb}-2}{n_{all}-2}\right)\right)$. It is straightforward to modify this formula where instead of measuring segregation of the foreign born of country c from natives, measuring their segregation from those outside the country of birth. This would change the formula to where instead of $\frac{n_{fb}-1}{n_{all}-1}$ measuring the likelihood a next-door neighbor was foreign-born, $\frac{n_c-1}{n_{all}-1}$ would measure the likelihood a next-door neighbor was from the same country of birth.

Now we turn to calculate the expected number of native-born neighbors under complete segregation, or $E(\underline{\text{native}}_c)$. Complete segregation from natives would occur if all immigrants from an ethnicity lived together along a line, leaving the two households on the sides of the neighborhood being either native-born or from a different country of birth. Complete segregation from the native born implies that the two households on either side are from different countries of birth; for example, an Irish neighborhood could be surrounded by German neighbors on both sides. Therefore, the lower bound for expected number of native-born neighbors $E(\underline{\text{native}}_c)$ is equal to zero. Setting the lower bound equal to zero is not accurate for the special case when there are only one or no other foreign-born immigrants from another country living in the county. This event was uncommon, for example, not occurring in the 1880 Census. However, if one were to calculate the measure for smaller levels of geography, such as the enumeration district, then there may not be immigrants from other sources in the same enumeration district; if so, one should resort to the Logan and Parman (2017a) method of calculating the lower bound.

For the section of the paper where the native born are those who are 3rd generation, or those who are United States born to two United States born parents, we change the numbers that enter the formula. The number of foreign born remains as the number who are first generation; however, the number of native born are is simply the number of third generation natives in the area.

Appendix C. An alternative measure of segregation that includes non-household heads

The page-based measure

The main measure of segregation is based on whether either of the next-door neighbor household heads are native born. This measure necessarily drops non-household heads, such as spouses, parents or servants. Moreover, the method drops non-households such as mining and railroad camps, poor houses and universities. Therefore, the household measure may provide an incomplete picture of interaction with the native born for the average immigrant.

We take an alternative approach to measuring segregation that does not require dropping non-households and non-household heads in the household. The approach is based on whether the foreign born are located on the same page as the native born, rather than whether the next-door head was native born. If the foreign born are not evenly spread throughout a county, then they will not appear on the same pages as the native born. Those on the census page are in close proximity since the census was taken on a line. The alternative segregation measure we use is the same basic formula for the main measure of segregation as in Equation (1):

$$\delta_c = \frac{E(\overline{native_c}) - native_c}{E(\overline{native_c}) - E(\underline{native_c})} \quad (C1)$$

For this measure, now $E(\overline{native_c})$ is the expected fraction of native-born on the page within a county or city under random assignment. This is simply the total number of native-born in the city or county divided by the total number of pages. For this measure, we only include those aged 18 and older to reduce child-rearing bias. The variable $native_c$ is the observed fraction of native-born individuals on the page for immigrants from source country c , and $E(\underline{native_c})$ is the expected fraction of native-born on the page under complete segregation. Similar in the main section, we treat $E(\underline{native_c}) = 0$ since the foreign-born would be located either entirely on pages with other foreign born from the same country, or foreign born from a different country. Each foreign-born individual on a page has the same difference between the expected number and total number of native on the page; to aggregate the measure to the county level, we simply weight the measure by the number of foreign-born individuals on the page.

This “page-based” measure is similar in spirit to the next-door neighbor measure, but it captures segregation in a slightly different way. Besides the difference between using

individuals instead of households, the page-based measure also measures segregation on the intensive margin of how many native-born does one live near, rather than just whether the individual lives near a native-born individual or not. We compare this page-based measure with the main household-based measure in Table C1, and show that the correlation between the two measures is 0.941. See Figure C1 for the binscatter relationship between the main household-based measure and this page-based measure. The difference between the measures could reflect a difference in measuring segregation, or the fact that we are able to include non-household heads and non-households in the measure; when calculate the “page-based” measure with only households, then the correlation with the main neighbor-based measure is 0.953. Therefore, the measures are closely related but do have slightly different results.

We present the page-based segregation trends by country of birth in Figure C2, which plots both trends for segregation from the 2nd-plus generation and segregation from the 3rd-plus generation. The broad relative levels and trends from the page-based measure are roughly the same as the neighbor-based measure. First, segregation levels are higher for Southern and Eastern Europeans at the turn of the 20th century relative to Western and Northern Europeans in the mid-19th century. Second, segregation trended to decrease past 1910 for all sources, and increased for Southern and Eastern Europeans between 1880 and 1910. Third, Chinese and Mexican segregation are relatively high, though the maximum levels are slightly lower than that of Southern and Eastern Europeans. The segregation trends by rural and urban areas are shown in Figure C3, which also demonstrate that trends were similar over time across rural and urban areas (except for Ireland). Moreover, segregation levels for Northern Europeans were very high in the mid-19th century. However, note that the levels of segregation across the page-based and neighbor-based measure are different, which is discussed more in the next section.

Figure C1. Relationship between household-based measure and page-based measure

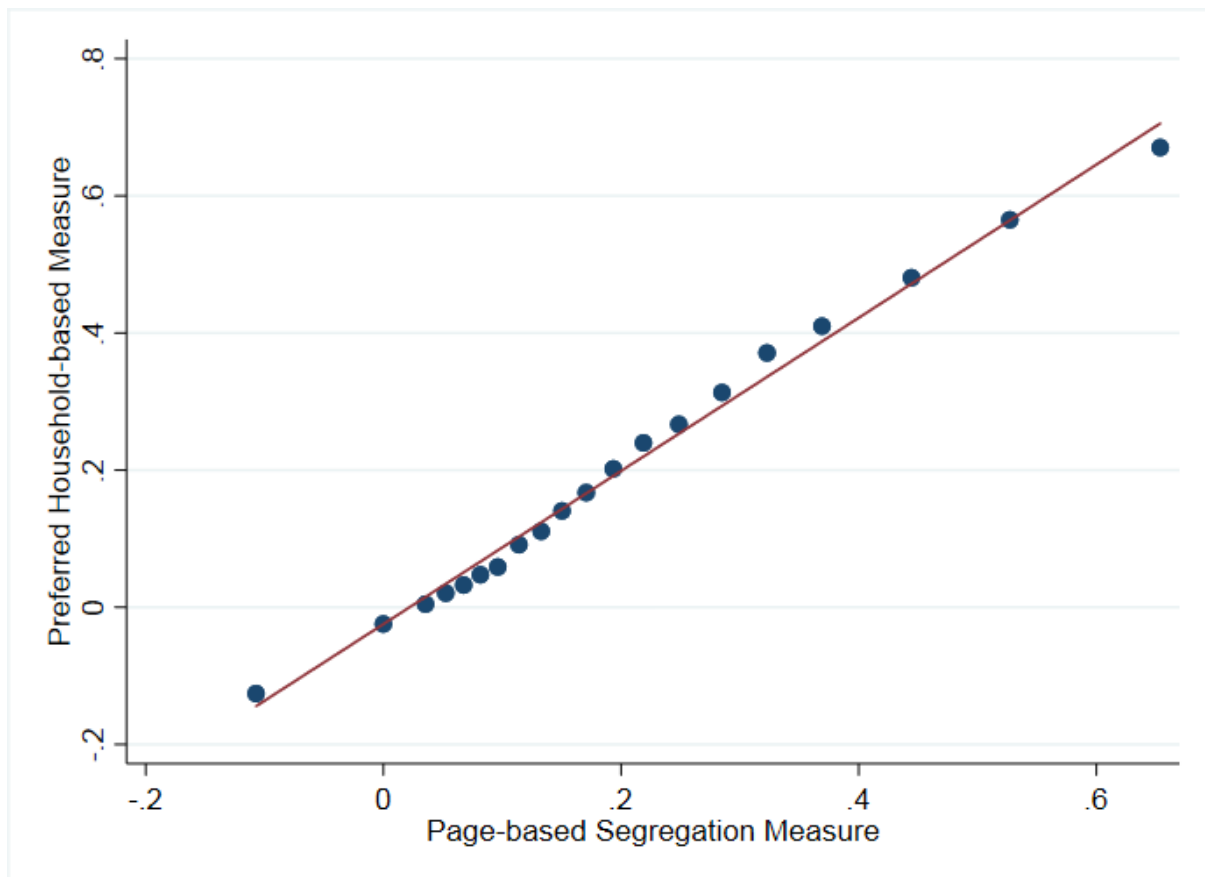
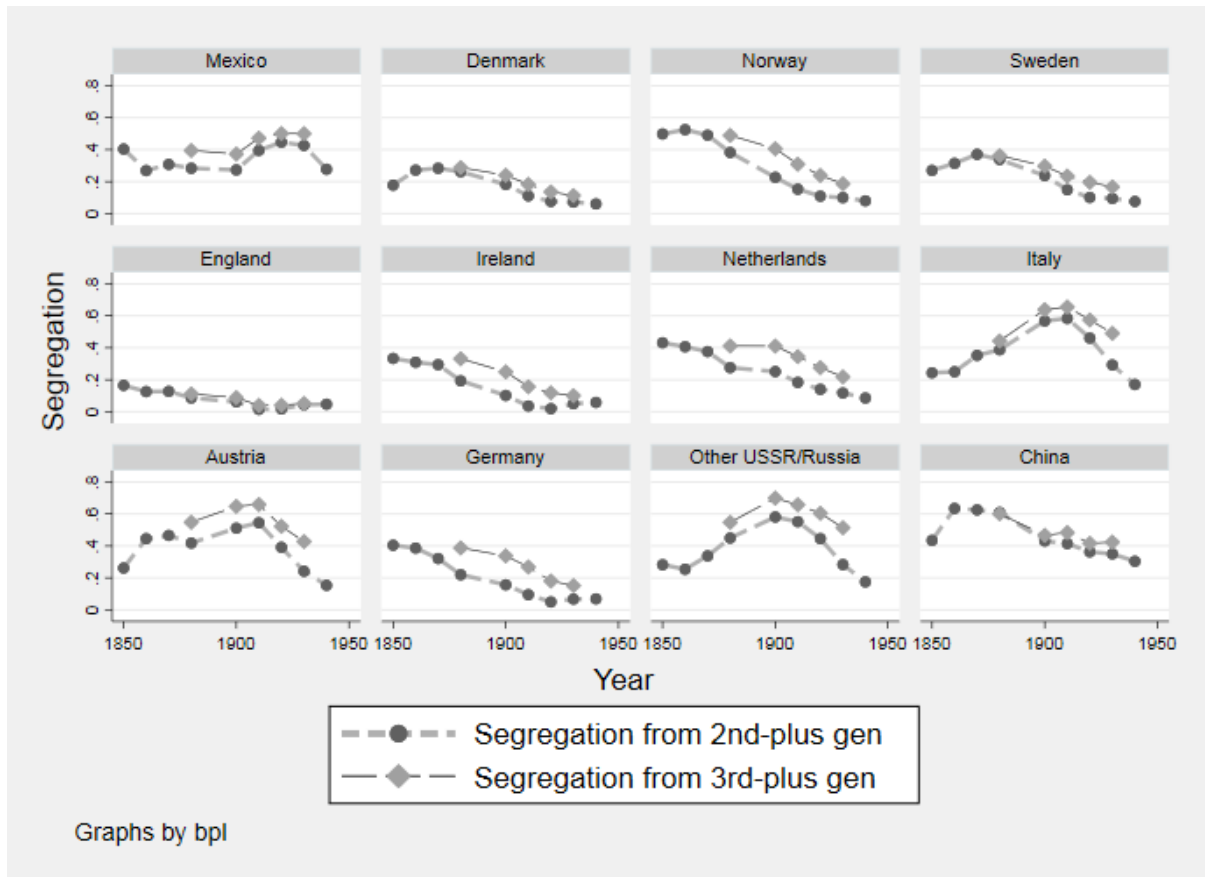
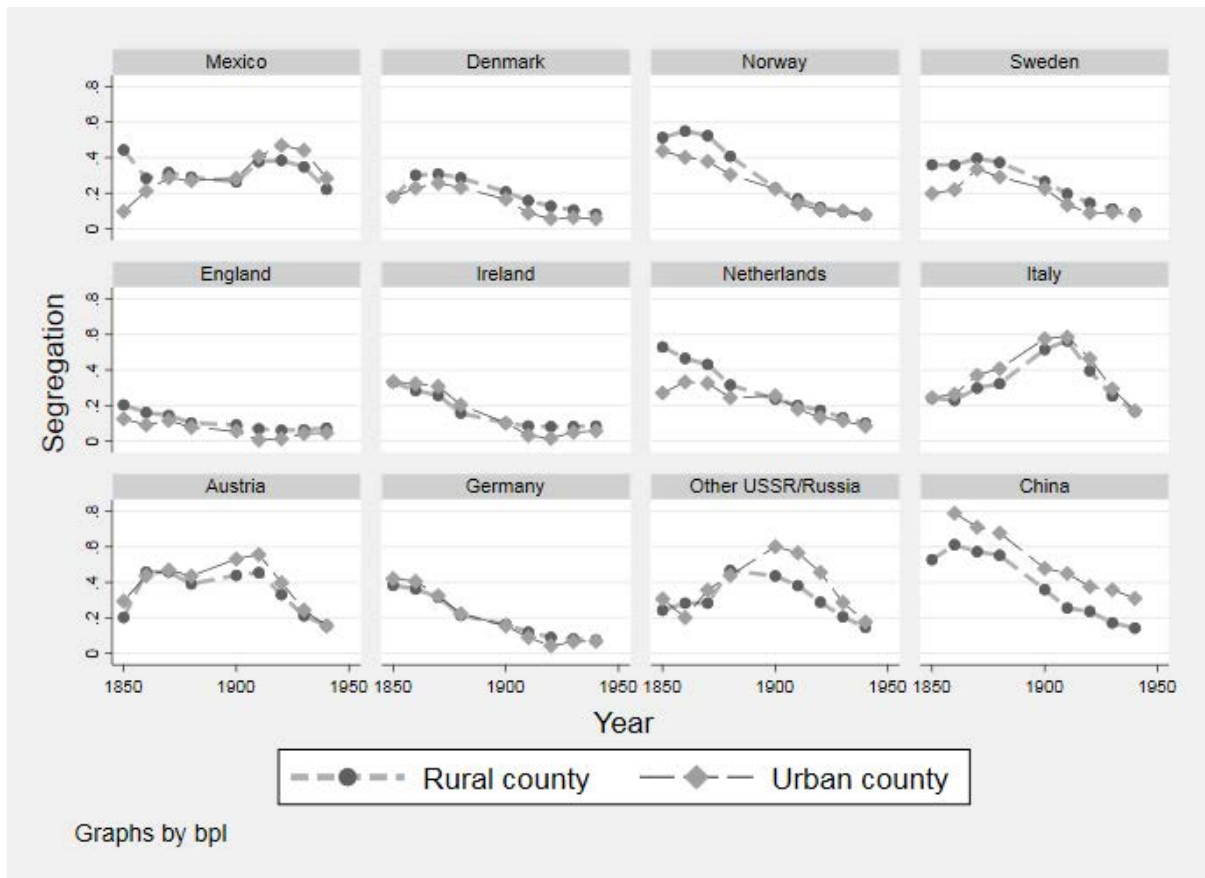


Figure C2. Trends in Segregation by Country of Birth, Page-Based Measure



Notes: Data is from the 1850 to 1940 full-count censuses. The page-based measure is discussed in Appendix C. This figure mimic Figure 2 from the main text.

Figure C3. Trends in Segregation by Country of Birth and Urban/Rural Counties, Page-Based Measure



Notes: Data is from the 1850 to 1940 full-count censuses. The page-based measure is discussed in Appendix C. This figure mimics Figure 3 from the main text.

Table C1. Correlation between preferred measure and page-based measure

	Household-based	Page-Based	Page-based, only HH
Main Household-based Measure	1		
Page-based Measure	0.9411	1	
Page-based Measure w/ only Household heads	0.953	0.9532	1

Notes: Correlation between measures when weighting for the number of households in county/year/country of birth cell.

Appendix D. Measuring Segregation from the Out-group

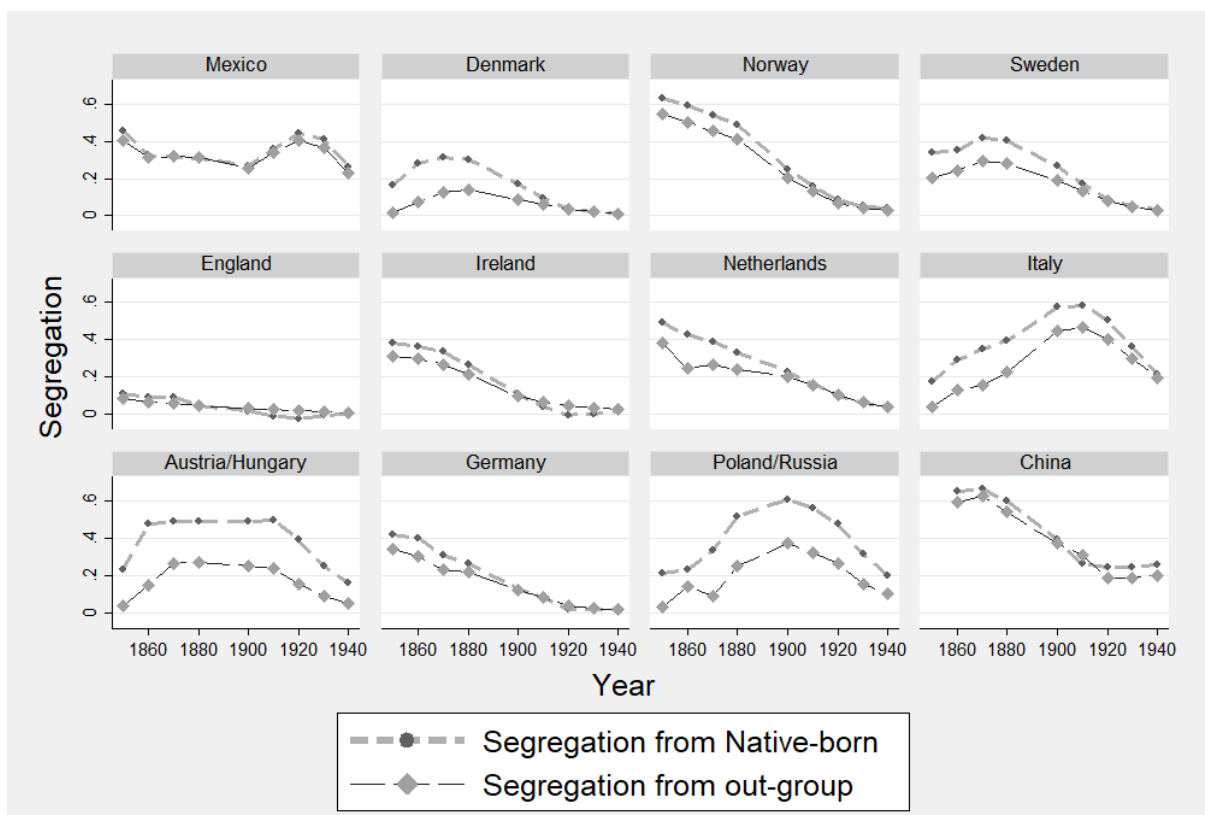
Our preferred measure of segregation is based on immigrants' (from a given country c) segregation from the *native-born*; that is, the in-group is based on country of birth, and the out-group are those born in the United States. Rather than using native-born as the out-group, one could use individuals from all other countries besides country c as the out-group. The fix for this in the formulas from Appendix B is simple: instead of counting the number of foreign-born with a *native-born* neighbor household head, we count the number of foreign-born with an *out-group* neighbor household head.

There are a few advantages for measuring segregation from other countries of birth rather than segregation from the native born; primarily, one does not measure negative levels of segregation for larger populations as we have done with our preferred measure. For example, we measure a negative level of segregation for Germans in New York City in 1940 because they were more likely to live next to a native-born household head than under random assignment. This is because Germans were more likely to next to US-born individuals rather than other non-German immigrants (e.g., from Southern or Eastern Europe). A negative level of segregation is not typical among standard segregation measures, such as the dissimilarity or isolation index. However, we prefer the main measure in text because we believe that living near the native-born is more relevant for measuring assimilation rather than segregation from the out-group; however, both measures are clearly informative for understanding immigrants' lived experience in the 19th and early 20th centuries.

In Figure D1, we present the segregation trends for our preferred measure and when measuring segregation from the out-group; this figure mirrors that of Figure 3. There are a few important differences in the trends and levels between the two measures. First, Eastern Europeans have a smaller level of segregation from the out-group than they do from the native born, and therefore were not as highly segregated from other individuals as southern Europeans. However, part of this may be because an immigrant from a given ethnicity or language may hail from different countries of birth; for example, Jewish immigrants from Russia/Poland, Germany, or Austria may live near each other and lower the measured level of segregation from other countries of birth. However, the level of segregation also falls for Southern Europeans, indicating that they also were less segregated from all others compared with segregated from the native born. Given that segregation levels are lower for Southern and Eastern Europeans when measuring segregation from the out-group, this leaves Chinese

immigrants as one of the most segregated sources, especially in the 19th century. Despite the level of segregation being lower for some sources, trends over time are largely similar.

Figure D1. Segregation from native-born versus segregation from out-group.



Notes: Data is from the 1850 to 1940 full-count censuses. The out-group measure is discussed in Appendix D. Segregation is measured from the second generation.

Table D1. Segregation from out-group (all other countries of birth)

Country	Year								
	1850	1860	1870	1880	1900	1910	1920	1930	1940
Canada	0.121	0.105	0.113	0.117	0.100	0.076	0.053	0.036	0.023
Mexico	0.405	0.311	0.320	0.317	0.257	0.337	0.403	0.363	0.231
Cuba		0.012	0.197	0.117	0.223	0.205	0.106	0.059	0.020
Denmark	0.016	0.075	0.125	0.139	0.091	0.063	0.036	0.021	0.012
Finland				0.230	0.345	0.332	0.289	0.212	0.133
Norway	0.545	0.504	0.454	0.414	0.208	0.129	0.070	0.041	0.027
Sweden	0.203	0.241	0.295	0.280	0.189	0.132	0.082	0.048	0.028
England	0.083	0.066	0.060	0.046	0.030	0.024	0.020	0.013	0.010
Scotland	0.057	0.045	0.038	0.028	0.014	0.013	0.009	0.011	0.008
Ireland	0.312	0.301	0.269	0.214	0.096	0.065	0.047	0.034	0.030
Belgium	0.162	0.177	0.215	0.212	0.103	0.093	0.071	0.056	0.035
France	0.111	0.086	0.067	0.041	0.023	0.025	0.015	0.011	0.007
Netherlands	0.384	0.247	0.268	0.243	0.203	0.156	0.107	0.063	0.040
Switzerland	0.135	0.087	0.074	0.050	0.031	0.023	0.013	0.010	0.007
Greece				0.014	0.080	0.193	0.126	0.067	0.043
Italy	0.042	0.129	0.157	0.226	0.440	0.470	0.404	0.297	0.195
Portugal	0.048	0.185	0.223	0.207	0.285	0.287	0.280	0.216	0.142
Spain	0.042	0.039	0.019	0.018	0.038	0.132	0.114	0.069	0.047
Austria/Hungary	0.037	0.151	0.264	0.275	0.244	0.239	0.157	0.089	0.054
Germany	0.345	0.305	0.229	0.217	0.125	0.084	0.035	0.023	0.018
Poland/Russia	0.030	0.142	0.090	0.253	0.377	0.321	0.267	0.157	0.101
China		0.594	0.624	0.539	0.335	0.309	0.186	0.187	0.203
Japan					0.650	0.595	0.343	0.281	
Turkey					0.116	0.200	0.086	0.062	0.036

Notes: Data is from the 1850 to 1940 full-count censuses. See Figure D1 for graphical depiction for 12 selected countries.

Table D2. Segregation from out-group (from non 1st or 2nd generation from same source), by country of birth

Country	1880	1900	1910	1920	1930
Canada	0.154	0.150	0.137	0.122	0.112
Mexico	0.515	0.495	0.504	0.507	0.468
Cuba	0.121	0.257	0.263	0.173	0.122
Denmark	0.145	0.111	0.095	0.071	0.053
Finland	0.256	0.357	0.352	0.323	0.280
Norway	0.459	0.318	0.257	0.199	0.152
Sweden	0.287	0.215	0.177	0.141	0.109
England	0.072	0.063	0.057	0.047	0.035
Scotland	0.043	0.029	0.025	0.019	0.019
Ireland	0.302	0.221	0.180	0.145	0.113
Belgium	0.231	0.151	0.126	0.090	0.073
France	0.064	0.038	0.036	0.025	0.018
Netherlands	0.283	0.313	0.278	0.226	0.169
Switzerland	0.068	0.050	0.043	0.031	0.026
Greece	0.019	0.081	0.197	0.128	0.068
Italy	0.232	0.448	0.485	0.433	0.367
Portugal	0.218	0.318	0.342	0.334	0.291
Spain	0.023	0.049	0.150	0.123	0.081
Austria/Hungary	0.284	0.276	0.270	0.189	0.131
Germany	0.305	0.298	0.259	0.168	0.139
Poland/Russia	0.258	0.395	0.344	0.315	0.226
China	0.541	0.354	0.379	0.259	0.298
Japan		0.651	0.599	0.345	0.293
Turkey		0.116	0.204	0.088	0.063

Notes: Data is from the 1850 to 1940 full-count censuses.