American Economic Journal: Applied Economics 2020, 12(4): 73–102
https://doi.org/10.1257/app.20170382

The Not-So-Hot Melting Pot: The Persistence of Outcomes for Descendants of the Age of Mass Migration†

By Zachary Ward*

How persistent are economic gaps across ethnicities? The convergence of ethnic gaps through the third generation of immigrants is difficult to measure because few datasets include grandparental birthplace. I overcome this limitation with a new three-generational dataset that links immigrant grandfathers in 1880 to their grandsons in 1940. I find that the persistence of ethnic gaps in occupational income is 2.5 times stronger than predicted by a standard grandfather-grandson elasticity. While part of the discrepancy is due to measurement error attenuating the grandfather-grandson elasticity, mechanisms related to geography also partially explain the stronger persistence of ethnic occupational differentials. (*JEL* J15, J22, J31, J51)

In the United States today, there are immigrants from over 150 different source countries, and more than 350 languages are spoken at home.[1] These differences in ethnic and cultural background are often accompanied by economic disparities. Researchers and policymakers have long been interested in how strongly these disparities across ethnicities (or source countries) in the first generation persist to the second and third generations (e.g., Borjas 1992, Glazer and Moynihan 1964). In a nation where the vast majority of the population are either immigrants or the descendants of immigrants, this question has broad relevance. If the old metaphor of the United States as a "melting pot" is true, then initial gaps across source countries should converge quickly, such that one's ancestry does not have a strong influence on his or her outcomes. In this paper, I ask whether the United States was a "melting pot" for descendants of the Age of Mass Migration (1850–1913), when tens of millions migrated from Europe to seek better opportunities in the United States.

A first-order problem in determining whether the United States was a "melting pot" is having data that can precisely measure convergence beyond the second generation. This difficulty arises because most datasets do not include grandparental

*Department of Economics, Baylor University, 1621 S 3rd Street, Hankamer School of Business, Waco, TX 76710 (email: zachary_ward@baylor.edu). Ilyana Kuziemko was coeditor for this article. Thanks to Tim Hatton, Priti Kalsi, Edward Kosack, Amber McKinney, and Santiago Pérez for their comments. Thanks to those attending presentations at the Australian National University, Baylor University, and University of California-Merced for their feedback. Thanks also to participants at the Asia-Pacific Economic and Business History Conference, 2017 EHA conference, and 2017 World Cliometrics Congress. Many thanks to Katherine Eriksson for generously sharing data on farmer incomes and to Lee Alston for helping me to gain access to the full-count census data.

†Go to https://doi.org/10.1257/app.20170382 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

[1] Data according to the 2012–2017 American Community Survey (Ruggles et al. 2018).

birthplace; instead, researchers must rely on those who self-identify their ancestry, which may be selectively reported and conflate the outcomes of the third generation with higher-order generations (Duncan and Trejo 2015, Duncan et al. 2017). I address this issue by linking immigrant grandfathers in the 1880 census to their second-generation sons in 1910 and their third-generation grandsons in 1940. With this new dataset, I can directly identify the third generation and their ancestry with grandparental birthplace. Therefore, I push the immigration literature, which has largely focused on the first two generations, to the third generation.[2]

These improved data further allow me to address an empirical puzzle in the historical immigration and intergenerational mobility literatures. First, the intergenerational literature finds that occupational mobility was high in the nineteenth and early twentieth centuries, which implies that economic gaps across families (and therefore, ethnicities) were eliminated relatively quickly (e.g., Long and Ferrie 2013, Feigenbaum 2018). In contrast, the immigration literature finds that the average level of economic status across European sources persisted strongly from the first to second generations, which suggests that the nineteenth and early twentieth century United States was not a "melting pot" for occupational outcomes (Abramitzky, Boustan, and Eriksson 2014). Since my data directly link grandfathers' outcomes to their grandsons' outcomes, I am able to measure the persistence of ethnic gaps across three generations and compare it to the transmission of status from grandfather to grandson. This comparison allows me to determine the importance of family influences across three generations relative to other influences that are correlated with one's ethnicity—similar to Borjas's research on the importance of "ethnic capital" across two generations (Borjas 1992, 1995).

After constructing the dataset from historical census files, I find that 51 percent of initial ethnic gaps in occupational income remained after three generations. This strong persistence suggests that the United States of the late nineteenth and early twentieth centuries was not, in fact, a "melting pot" for occupational outcomes. The slow convergence across three generations could be mostly due to family influences from grandfather to grandson; however, with the same dataset, I also estimate that only 21 percent of initial occupational income gaps across grandfathers persisted to their grandsons. Therefore, ethnic gaps persisted two and a half times more strongly than predicted by a multigenerational model, suggesting that there is some dragging force that slows the convergence of ethnic gaps. The magnitude of this dragging force can be captured in a regression that includes both the first-generation ethnic mean and grandfather's occupational income, which shows that the grandson's occupation is additionally correlated with the first-generation ethnic mean (estimated with an elasticity of 0.30). Given both the transmission from grandfather to grandson (21 percent) and the additional correlation with the first-generation ethnic mean (30 percent), ethnic gaps in occupational income persisted at a total rate of 51 percent.

There are many reasons why occupational gaps may have persisted across ethnic groups more strongly than predicted by a standard multigenerational model, such

[2] See Borjas (1992) and Card, DiNardo, and Estes (2000) for two-generation studies. Borjas (1994) estimates persistence across three generations but must rely on self-reported ancestry rather than grandparental birthplace.

as a direct effect of first-generation co-ethnics on descendants, or because there was some omitted factor that is correlated with both the first-generation's and grandson's occupational outcomes. The most prominent explanation in the literature is that ethnic groups have varying levels of "ethnic capital" (Borjas 1992, 1994, 1995). Borjas defines "ethnic capital" as the quality of the ethnic environment in which children are raised, which directly affects the next generation's outcomes via role model or social interaction effects with those from the same ethnicity.[3] Yet ethnic-specific causal factors are not the only possible explanation for intergenerational drag; for instance, it may be that immigrants lived in poorer neighborhoods that had negative effects on subsequent generations through general education or health mechanisms. Another possibility is that there was no causal mechanism driving stronger persistence at the ethnic level, but that the result is due to measurement error: that is, the estimated grandfather-grandson association is smaller than the true association because it is subject to attenuation bias (e.g., Clark 2014; Ferrie, Massey, and Rothbaum 2016; Solon 1992). Given measurement error, the ethnic mean would serve as a useful proxy for the father's true occupational status and thus would be positively correlated with the next generation's outcomes despite having no causal effect.

Rather than argue for an ethnic-specific causal mechanism, I instead point to measurement error and geography as key reasons for the stronger persistence of ethnic differentials across three generations. To address the issue of measurement error, I further link the fathers in 1910 to a second observation in 1920 so that I can use the average outcome to better proxy for the father's permanent economic status (Solon 1992). This method of using two father observations is a unique contribution to the historical mobility literature since obtaining more than one father observation is difficult due to the high cost of linking individuals across censuses—a cost which has fallen in recent years. I show that the one-father-observation elasticity is attenuated when compared with the two-father-observation elasticity.[4] If I assume that the father's occupational income is subject to classical measurement error, then the true father-son elasticity should be 54 percent higher than the one-observation estimate. Yet even after accounting for measurement error, ethnic differentials still converged more slowly than predicted by a father-son elasticity. Since measurement error does not fully explain the results, this suggests that there is some mechanism that causes the stronger persistence of ethnic occupational gaps.[5]

I argue that mechanisms associated with geography partially explain why ethnic differentials persisted so strongly across three generations. To demonstrate this, I show that ethnic differentials converged quickly within the grandfather's neighborhood after controlling for 1880 neighborhood fixed effects.[6] This result

[3] While Borjas's results are correlational, there is quasi-experimental evidence that highly educated ethnic peers in the neighborhood positively affect the educational outcomes of immigrants (Åslund et al. 2011).

[4] While one may expect that occupation is not subject to attenuation bias (unlike income), Mazumder and Acosta (2015) show that attenuation bias also exists for occupations in the Panel Study of Income Dynamics. Measurement error in the historical data could also come from a lower quality of data.

[5] Unfortunately, I do not show how measurement error influences the three-generation elasticity because at this point I cannot obtain another observation of the grandfather's occupation. This is because the microdata from the closest censuses to 1880 are either not cleaned (1870) or burned in a fire (1890).

[6] I do this with 1880 enumeration-district fixed effects. Enumeration districts are administrative areas where a census taker recorded all those within the enumeration district. They averaged about 1,600 individuals in the 1880 census.

suggests that there are omitted neighborhood effects that are correlated with both the first-generation ethnic mean and the grandson's outcome. Net of 1880 neighborhood fixed effects, 16 percent of initial ethnic differentials remained after three generations—a third of the estimate without fixed effects. A main reason why this may occur is that the location fixed effects capture the importance of farming or growing up in a rural area. When one separately drops farmers from the sample and does not include neighborhood fixed effects, ethnic gaps in occupational income persisted across three generations at 27.5 percent—about half the rate predicted when including farmers. Besides farming, any unobserved mechanism that is related to geography is accounted for in the neighborhood fixed effects, including the quality of schools or health environment. While ethnic spillovers may be one mechanism for the strong persistence of ethnic gaps, the results suggest that general place effects are of high importance.

The paper's focus on third-generation immigrants is shared by others in the immigration literature, a literature where generations are not defined within family but grouped by whether one or one's parents were born abroad or in the United States (e.g., Borjas 2001; Alba, Lutz, and Vesselinov 2001).[7] I improve on these immigrant convergence studies since I am the first to have grandfather-father-son linked data and do not have to rely on the self-reported ancestry question.[8] Despite having better data, the evidence supports the argument of Borjas (1992, 1994, 1995) that gaps in economic status are slow to disappear for descendants of immigrants. In fact, my study suggests that convergence may be even slower than Borjas (1994) estimates with data using self-reported ancestry, partially because I include farmers in my estimates—although my results are not directly comparable since I study a different time period.[9] At the same time, I do not argue for the importance of ethnic-specific causal mechanisms as in Borjas (1992, 1995).

The paper is also related to the fast-growing literature on multigenerational mobility (Solon 2018), some of which uses historical US censuses (e.g., Ferrie, Massey, and Rothbaum 2016; Long and Ferrie 2018; Olivetti, Paserman, and Salisbury 2018).[10] While most of this literature focuses on how grandfather and grandson outcomes are correlated, I extend it by showing that the grandson's outcomes are further associated with co-ethnic outcomes from 60 years prior. Therefore, one cannot infer convergence of ethnic differentials for the entire American population—even for white Americans—from a grandfather-grandson multigenerational model with a single intercept. This suggests that there are independent "group effects" across

[7] The main strand of the immigration literature explores mobility between two generations, but only a few studies have data that link father to son (Borjas 1992, León 2004). See Card, DiNardo, and Estes (2000) for another study on mobility across two generations, which uses group averages; also, Smith (2003, 2006).

[8] Others that do identify the third generation via grandparent's country of birth (e.g., Duncan and Trejo 2011) are limited in that they must explore outcomes when the child is still in the household. Recently, Duncan et al. (2017) have been able to estimate adult outcomes of third generation Mexican Americans with data from the NLSY97.

[9] Borjas (1994) uses the 1910, 1940, and 1980 censuses. I use the 1880, 1910, and 1940 censuses.

[10] See Black and Devereaux (2011) for a review of intergenerational studies and Solon (2018) for multigenerational studies. There are many other historical studies that estimate heterogeneity in two-generation mobility across geography, during economic shocks, or for different subgroups based on sex or race (e.g., Ager, Boustan, and Eriksson 2019; Bleakley and Ferrie 2016; Clark 2014; Chetty et al. 2017; Collins and Wanamaker 2017; Feigenbaum 2015; Feigenbaum 2018; Hilger 2016; Kosack and Ward 2019; Long and Ferrie 2013; Olivetti and Paserman 2015; Perez 2018; Tan 2018).

multiple generations such that being part of a group (i.e., an ethnicity) matters for the grandson's occupation. Also, while many have shown that an iterated AR(1) model predicts that multigenerational convergence is too fast, I contribute to the literature by showing that this pattern does not hold for ethnic averages. Instead, an iterated AR(1) accurately predicts the persistence of ethnic gaps in occupational income. This result appears to occur because the "ethnic capital" component converges at a faster-than-geometric rate (perhaps due to social or cultural assimilation), which contrasts with the family component converging at a slower-than-geometric rate. These different convergence rates for the family and "ethnic capital" components appear to cancel each other out such that overall ethnic averages converged at a geometric rate across three generations.

Finally, my results also relate to studies on the importance of location for subsequent generations— research that increasingly reveals that place matters. In a series of papers, Raj Chetty, Nathanial Hendren, and various coauthors show that intergenerational associations vary in strength across location and depend critically on the quality of the childhood neighborhood.[11] My paper controls for these neighborhood effects by including 1880 enumeration district fixed effects; I then show that ethnic gaps closed quickly for descendants of the same 1880 neighborhood. This result suggests that factors related to geography, such as exposure to local labor markets or the quality of the childhood environment, matter for outcomes across three generations. However, I am unable to identify causal effects of place in my data; it is also possible that whatever led immigrants to sort into different areas in the first generation also led to their grandchildren having similar occupations.

I. Prior Evidence on Convergence in the Age of Mass Migration

The plurality of today's American population can be traced back to the European immigrations between 1850 and 1913, an era known as the Age of Mass Migration (Hatton and Williamson 1998). Immigration to America was mostly free, which led to the highest documented immigration rate in United States history; the rate has since decreased due to restrictions enacted during and after World War I (Abramitzky and Boustan 2017).[12] Europeans dominated historical flows, which is evident in modern-day data: the top three ancestries in the 2014 American Community Survey are German, Irish, and English—the same as the top senders between 1850 and 1880.[13] Other sources from Southern and Eastern Europe had a larger role in the later stages of the Age of Mass Migration (post 1880) when steam technology reduced travel costs and made immigration possible for millions of Italians, Greeks, and Russians.

The standard view is that European immigrants assimilated relatively well, which may be true for the average immigrant compared to natives, but outcomes still varied

[11] See Chetty et al. (2014); Chetty, Hendren, and Katz (2016); Chetty and Hendren (2018a, b).

[12] There were explicit restrictions on Chinese immigration in 1882, and implicit restrictions on Japanese immigration in 1907. Other subcategories were barred from entry, such as anarchists, epileptics, the "feeble-minded," and prostitutes.

[13] This calculation combines both the first and second response to the ancestry question.

widely across sources (Borjas 1994, Minns 2000).[14] On average, immigrants remained stuck at the same point in the occupational distribution as they were at arrival; few sources converged to native outcomes (Abramitzky, Boustan, and Eriksson 2014). Ethnic group averages persisted not only from arrival to decades afterward, but also to the second generation, showing little between-group convergence (Abramitzky, Boustan, and Eriksson 2014; Darity, Dietrich, and Guilkey 2001).

While occupational gaps across ethnicities were largely preserved between the first and second generations, a small literature finds a weaker relationship between the first and third generations. The study most related to this paper is that of Borjas (1994), who estimates group-level persistence between the first generation in 1910, identified by country of birth, and the third-plus generation in 1980, identified by self-reported ancestry. Borjas finds that ethnic differentials persisted from the first generation to the third-plus generation at 0.20, or that four-fifths of initial ethnic differentials disappeared by the third-plus generation.[15] While this may imply that ethnic differentials are slow to converge, a 0.20 result may be due to transmission from the grandfather to the grandson rather than a slower convergence of ethnic group averages. My study improves on Borjas (1994) since I have linked grandfather-grandson data. I can also identify the grandson's origin with grandparent's country of birth rather than self-reported ancestry. I additionally include farmers, who Borjas excludes since there were no estimates for farmer income at the time.

II. Measuring Persistence at the Individual and Group Levels

The standard method for measuring income persistence across generations is to regress a son's income on his father's income, a regression which has been studied for many different countries and time periods (Black and Devereux 2011, Solon 1999).[16] A common specification is

$$y_{i,g} = \alpha_0 + \alpha_1 y_{i,g-t} + \varepsilon_{i,g}, \tag{1}$$

where $y_{i,g}$ is the log wages for individual i from generation g. When using logs, the coefficient β_1 measures the intergenerational elasticity coefficient (IGE), which is between 0 and 1 and commonly estimated to be around 0.5 for the United States (Corak 2006).

One does not always have data on fathers and sons, so instead of estimating intergenerational mobility using families, others estimate mobility at a group level. In the immigration literature, a common interest is how economic differentials in the first generation, grouped by country of birth, predict differentials in the second generation, grouped by parental country of birth (Borjas 1993; Card, DiNardo, and

[14] In addition to the economics literature, there is a long sociology literature on the assimilation of immigrants across generations. See, for example, Alba and Nee (2009), Glazer and Moynihan (1963), Portes and Rumbaut (2006), and Perlmann (2005). More recently, see Catron (2019).

[15] Alba, Lutz, and Vesselinov (2001) argue that intergenerational convergence is nearly complete when limiting the sample to European sources, implying that there was no within-family persistence from grandfather to grandson. See Borjas (2001) for a rebuttal.

[16] See Table 4.L2 in Dustmann and Glitz (2011) for a review of intergenerational assimilation studies for countries other than the United States.

Estes 2000). A regression that estimates persistence across immigrant generations is closely related to equation (1), but instead of using individual-level data, it uses the average income or education level by source country (to proxy for ethnicity e):[17]

$$\bar{y}_{e,g} = \theta_0 + \theta_1 \bar{y}_{e,g-t} + \varepsilon_{e,g}. \tag{2}$$

A common procedure for estimation is not to run equation (2), but instead to impute the father's earnings with the average earnings of the group based on a first-stage regression.[18] Then, using the imputed father's earnings, one can estimate the intergenerational elasticity in a second-stage regression (Solon 2018).

While useful, grouped estimators do not recover the intergenerational elasticity coefficient for the population if there is a separate correlation between the group average and an individual's outcome (Torche and Corvalan 2016). In other words, there may be a violation of the exclusion restriction when imputing father's income in the first stage. In a series of papers, Borjas (1992, 1995) addresses this point by including the group average directly in the individual-level intergenerational elasticity equation:

$$y_{i,e,g} = \beta_0 + \beta_1 y_{i,e,g-t} + \beta_2 \bar{y}_{e,g-t} + \varepsilon_{i,e,g}. \tag{3}$$

In this regression, Borjas refers to the average income level of the prior generation $\left(\bar{y}_{e,g-t}\right)$ as "ethnic capital," theorizing that ethnic capital causally affects the outcomes of the next generation through human capital spillovers on children of the same ethnicity. While the exact mechanism is unclear, this specification models the intercept for each country of birth as a function of the prior generation's mean income or education level; thus, ethnicities with lower levels of income in the first generation are predicted to have lower than average income in the second generation.

If one averages equation (3) by country to form a group-level regression, then the equation becomes

$$\bar{y}_{e,g} = \beta_0 + \left(\beta_1 + \beta_2\right) \bar{y}_{e,g-t} + \bar{\varepsilon}_{e,g}. \tag{4}$$

Note that this equation is the same as equation (2), where $\theta_1 = \beta_1 + \beta_2$. If the average income level has an effect above and beyond the father's income, then the persistence of ethnic differentials is stronger than the persistence between the father and son from a standard intergenerational model $\left(\beta_1 + \beta_2 > \alpha_1\right)$. Borjas (1995) terms this sum $\left(\beta_1 + \beta_2\right)$ the "mean convergence" of ethnic group averages. The primary interest of this paper is to not only measure the mean convergence across three generations, but also compare its magnitude with the grandfather-grandson association.

While Borjas terms the average outcome of the ethnic group as "ethnic capital," I refrain from using this term and instead use "ethnic mean." This is because the mechanism that drives the persistence of ethnic differentials may be not be due to

[17] This regression is typically weighted by the size of the second generation.

[18] This grouped estimator has been used in a variety of contexts such as grouping based on rare surname, first name, or state of birth (Clark 2014, Olivetti and Paserman 2015, Aaronson and Mazumder 2008).

ethnicity per se, but rather due to factors that are correlated with the ethnic mean. For example, ethnicities are clustered in different areas of the country, live in different labor markets, and are exposed to different quality public goods, all of which may lead to stronger persistence of ethnic differentials than predicted from a standard model, but are not ethnic-specific causal factors. While human capital spillovers within ethnicity may be a factor—as argued by Borjas (1992)—there are many other potential reasons for stronger persistence at a group level.

It is also possible that there is no causal mechanism for stronger persistence at the group level, but that an intergenerational model simply fails to capture the father's true occupational status. In this case, a positive correlation with the ethnic mean (β_2) occurs because the mean provides more information about the father's permanent occupational status. Since it is unclear whether there is a causal factor—and if there is one, if it is ethnic-specific—I avoid using the term "ethnic capital" and instead use "ethnic mean."

III. Data: Linking the 1880, 1910, and 1940 Censuses

A. *Data Creation*

The goal of this paper is to estimate the persistence of ethnic occupational gaps from first-generation immigrants to their grandchildren. The data requirements for these estimates are high since data that link grandfather to grandson are uncommon; moreover, I need enough observations for each source to provide a reliable estimate of a group average. I take advantage of the fact that US censuses are publicly released 72 years after enumeration, and that Ancestry.com and the University of Minnesota Population Center (IPUMS) have digitized much of this data (Ruggles et al. 2018). I link the full-count 1880, 1910, and 1940 US censuses to create a new sample of grandfathers linked to fathers and grandsons.

Before describing the linking process in more detail, I first discuss my approach to building the sample (see Figure 2). I start by locating all native-born males under 14 in either the 1880 or 1910 censuses, and then link them forward 30 years later to either the 1910 or 1940 census when they are old enough (30–44 years) to hold an occupation. This leaves me with two datasets of children linked to adult outcomes, who I label as either second generation (G2) from 1880–1910 or third generation (G3) from 1910–1940. The first generation (G1) are the fathers in 1880. I then use information on the relationship status in the household to attach the father's characteristics to those who are successfully linked. Not all linked children have observable fathers for various reasons such as death or separation from the household; this necessarily makes my results only representative of the population with observable fathers (Xie and Killewald 2013). At this point I have linked children from either 1880–1910 or 1910–1940 and have attached their father's characteristics. The final step to create grandfather-grandson links is to simply locate the subsample where a successfully linked G2 individual from 1880–1910 is also the father of a G3 individual linked from 1910–1940.

I use Feigenbaum's (2016) method for linking censuses, where I first hand-link a subsample of 2,000 individuals from 1880 to 1910, and another 2,000 from 1910 to

1940. To link individuals, I find the best match based on first name, last name, race, year of birth, and state of birth/country of birth. For the 1880 to 1910 link, I can additionally use information on father's and mother's country of birth to help pick the correct link; unfortunately, these variables are unavailable for linking between 1910 and 1940. Based on these hand-linked data, I estimate a probit model to predict the best match from a set of potential matches. I then use the probit coefficients to predict linking scores for potential matches between the full-count census to create my linked dataset. Please see online Appendix C for more detail on the exact process.

For inclusion in the regression sample of grandfathers(G1)-fathers(G2)-grandsons(G3), I impose additional restrictions. First, I keep G1 grandfathers who are between 30 and 55 to reduce measurement error from life cycle bias; note that the ages of the G2 and G3 generations are between 30 and 44 since I only link sons who were under 14 in childhood (Grawe 2006). Since my interest is in the outcomes of descendants of immigrants, I limit my sample to foreign-born G1 grandfathers, native-born G2 fathers, and native-born G3 sons. Therefore, in the terminology of the immigration literature, the G1 grandfather is a first-generation immigrant, the G2 father is a second-generation immigrant, and the G3 grandson is a third-generation immigrant.[19] Note that foreign-born were a sizeable 23.2 percent of the adult male population during the Age of Mass Migration.[20] I also only keep G1-G2-G3 sets in which an occupation is observed in all three generations.

After linking and imposing the various sample restrictions, I end with 66,327 G1 grandfathers in 1880 linked to 71,702 G2 fathers in 1910 and 96,726 G3 grandsons in 1940. This sample is necessarily a subset of the (unobservable) population of grandfathers with 30–44-year-old sons in 1910 and 30–44-year-old grandsons in 1940. It is a subset because I am unable to perfectly match people across censuses. For example, my method links 19.9 percent of under 14-year-olds in 1880 to 1910 and 24.7 percent from 1910 to 1940. With respect to the original 1880 population of G2 sons, I am only able to create links to three generations for 4.0 percent of them. However, this linking rate is misleadingly low because I do not know the counterfactual population of G2 sons who *also* have 0–14-year-old sons in 1910. Nevertheless, a linking rate of less than 100 percent is due to death, common names, and errors of data entry from either the initial census enumerators or clerks that digitized the data. This method of linking certainly leads to some false links; nonetheless, this method has been shown to produce reliable intergenerational elasticity estimates (Bailey et al. 2017).

To determine the representativeness of the sample, I compare the linked population with the linkable population in online Appendix C. The main biases in the linked sample are that I am more likely to have Germans, sons with farmer fathers, and those in the Midwest in my sample; as a trade-off, I am less likely to have Irish, those with unskilled fathers, and those from the Northeast. These biases are partially because Northeastern states have larger populations, and thus I am less likely to find

[19] Note that I keep G3 grandsons who have a foreign-born mother since I define generations through the paternal line. I prefer to keep these individuals since I am interested in all descendants of the first generation, including those who have sons that marry a native-born or a foreign-born wife.

[20] Based on 18+-year-old males from the 1880 census (Ruggles et al. 2018).

a unique match. I reweight the sample by source country, occupational category, and region of residence to ensure that my linked sample is representative of the population on observables, although this reweighting does not change qualitative results (see online Appendix C for the representativeness check). I use the weighted sample for the rest of the paper.

B. *Imputing Incomes by Occupations*

I estimate multigenerational mobility based on occupation rather than income or education since these variables are not available in earlier censuses. I show results based on broad occupational categories (i.e., white-collar, farmer, unskilled, and semiskilled) such as in Long and Ferrie (2013) and based on occupational income such as in Olivetti and Paserman (2015). For coding of white-collar, farmer, unskilled, and semiskilled categories, I classify the 3-digit *occ1950* codes from IPUMS into the four categories.[21] Broadly, white-collar workers are professional, technical, managerial, or sales workers; farmers are only farmers (owners and tenants), but not farm laborers; unskilled are laborers (farm and nonfarm), low-skilled service workers, or operatives; and semiskilled are craftsmen.

I prefer results using occupational categories since assigning income to each occupation across the entire 1880–1940 period is not straightforward. Nevertheless, I also present results based on two different occupational scores. First, I assign each occupation an income in 1880 and 1910 using the estimates from Sobek (1996, Table A1), which are based on various data sources and, in particular, the US Commissioner of Labor reports between 1886 and 1905.[22] I assign each occupation the income in 1940 with the more commonly used *occscore* variable since it is based on a closer time period (from 1950 earnings). In text I refer to this as the "1890–1950" score since it captures how occupational income changed between the late nineteenth and mid-twentieth centuries.[23]

I also present results when using the more common *occscore* throughout the entire 1880–1940 period; however, this fixes occupational income at 1950 levels throughout the entire period. I refer to estimates using the *occscore* variable as the "1950" score. I also check my results against alternative ways of estimating occupational income, such as allowing farmer income to vary by state or allowing all occupational incomes (not just farmer) to vary by census region. These checks are important since farmers are a large portion of the dataset in the early censuses; for example, 48 percent of grandfathers are farmers in 1880. The results hold under these alternative scores or specifications, such as when dropping farmers or estimating farmer income at the state level, although farming remains a key reason why I find strong persistence of ethnic differentials across three generations.

Since my mobility estimates are based on occupational income, they should not be directly compared to modern-day estimates based on income. While there

[21] See https://usa.ipums.org/usa-action/variables/OCC1950#codes_section for the list of occupations.

[22] For occupations not listed in Sobek (1996) but in the *occ1950* codes, I predict the score in 1890 based on its position in the 1950 score distribution. Note that when using these estimates from Sobek in the dataset, farmers are placed at the thirty-eighth percentile of occupational scores in 1880.

[23] I convert all income to 2015 dollars using the CPI adjustments from measuringworth.com.

is evidence that occupational income elasticities are roughly similar to actual income elasticities for 1915–1940 Iowa (Feigenbaum 2018), I cannot verify that this pattern also holds for my data between 1880 and 1940. The main limitation of the occupational income measures is that I miss income variation within occupation across ethnicities; for example, I do not capture if Irish farmers earned less than German farmers. Therefore, the reader should keep in mind that I estimate convergence of occupational distributions rather than convergence of income distributions.

C. *Descriptive Statistics*

The mean occupational income and the number of observations by first-generation country of birth are listed in Table 1. Almost half of my sample comes from Germany, which was the origin with the largest immigrant stock in 1880. The next two largest sources were Ireland and England. Together, these three sources make up 74 percent of grandsons in my dataset. Following these countries, the next largest sources are Canada, Norway, and Scotland. Note that Southern and Eastern European sources had smaller stocks in the 1880 census.[24] The importance of farming is seen in Table 1 since ethnicities that were more rural had lower occupational income and ethnicities that were more urban had higher occupational income.

The first evidence that ethnic differentials persisted from the first to third generation is shown in Figure 1. Figure 1 plots the first-generation ethnic averages from Table 1 (for my preferred log occupational score) against the third-generation ethnic average in 1940. That is, I estimate θ_1 from equation (2) when weighting by third-generation population. The figure shows a clear correlation between the ethnic means that are measured 60 years apart. Some association is expected since a standard multigenerational model predicts that occupational income should transmit from the grandfather to the grandson, but the association shown in Figure 1 is stronger than expected. The ethnic-mean elasticity is 0.51, which is higher than other estimates of the grandfather-grandson elasticity between 0.15 and 0.24 for nineteenth and early twentieth century data (Long and Ferrie 2018; Olivetti, Paserman, and Salisbury 2018). However, these grandfather-grandson estimates from others are not directly comparable with the ethnic-mean elasticity in Figure 1 due to differences in time period, population of interest, and measure of economic status.[25] Therefore, the aim of this paper is to compare the persistence of ethnic means with the predicted persistence from a multigenerational model when both estimates are consistently created from the same dataset.

[24] Austria/Hungary is grouped to include Austria, Hungary, Czechoslovakia, and Yugoslavia. Russia/Poland includes Russia, Poland, Latvia, Lithuania, Estonia, and any Baltic State. This coding does not drive results since they are a small part of the dataset.

[25] Long and Ferrie (2018) use occupational wealth instead of occupational income. Olivetti, Paserman, and Salisbury (2018, Table 2) estimate the grandfather's occupational income based on first name, while they have linked data between the G2 and G3 generations.

Table 1—Occupational Earnings and Number of Observations from Grandfather-Father-Grandson Linked Sample

	G1 in 1880		G2 in 1910		G3 in 1940	
Origin in G1	ln(Occ. score)	*N* of G1	ln (Occ. score)	*N* of G2	ln (Occ. score)	*N* of G3
Canada	9.29	5,781	9.38	6,214	10.10	8,287
Mexico	9.06	143	8.94	148	9.74	189
Denmark	9.18	516	9.25	555	10.06	783
Norway	8.96	1,756	9.16	1,903	9.94	2,747
Sweden	9.13	1,229	9.31	1,308	10.06	1,740
England	9.37	8,881	9.42	9,646	10.13	12,851
Scotland	9.41	1,877	9.49	2,013	10.13	2,613
Ireland	9.38	9,863	9.56	10,316	10.18	13,279
Belgium	9.11	242	9.16	262	10.05	401
France	9.37	1,333	9.45	1,430	10.13	1,929
Luxembourg	9.05	60	9.09	62	9.88	110
Netherlands	9.27	997	9.35	1,106	10.08	1,584
Switzerland	9.23	1,259	9.32	1,409	10.06	1,963
Italy	9.56	207	9.61	219	10.21	276
Portugal	9.29	96	9.47	102	10.12	129
Spain	9.55	32	9.60	35	10.25	48
Austria/Hungary	9.25	1,042	9.29	1,139	9.98	1,630
Germany	9.34	30,683	9.38	33,497	10.07	45,727
Russia/Poland	9.49	330	9.61	338	10.14	440
Overall	9.21	66,327	9.36	71,702	10.08	96,726

Notes: Data are from the 1880–1910–1940 linked sample of grandfathers to fathers to sons. The occupational score uses occupational income estimates from 1890 in Sobek (1996) for the 1880 and 1910 censuses, and then the 1950 IPUMS *occscore* estimates for the 1940 census. See Figure 1 for plots of the data. Austria/Hungary is grouped to include Austria, Hungary, Czechoslovakia, and Yugoslavia. Russia/Poland includes Russia, Poland, Latvia, Lithuania, Estonia, and any Baltic State.

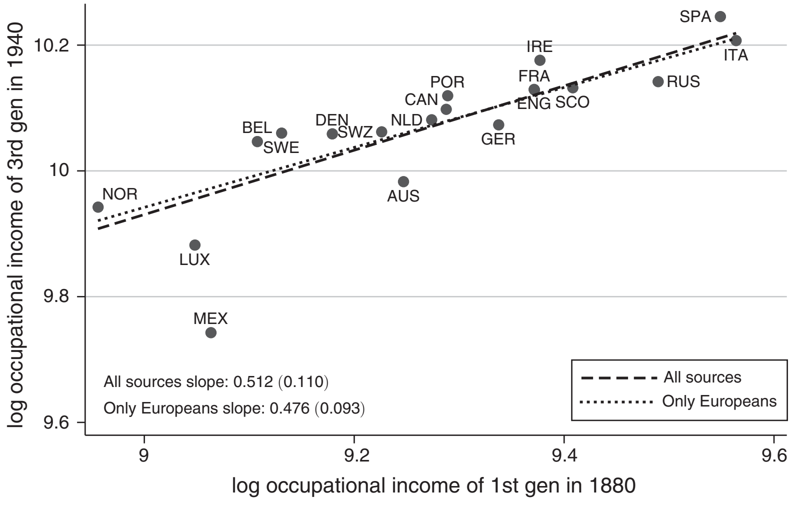

Figure 1. Persistence of Ethnic Differentials from the First Generation in 1880 to the Third Generation in 1940

Notes: Data are from the linked sample 1880–1910–1940. The underlying values are from Table 1. The slopes are weighted by the number of G3 grandchildren for each country of origin. Standard errors are in parentheses. Occupational earnings are measured in 1880 with data from Sobek (1996) and in 1940 with the *occscore* variable from IPUMS.

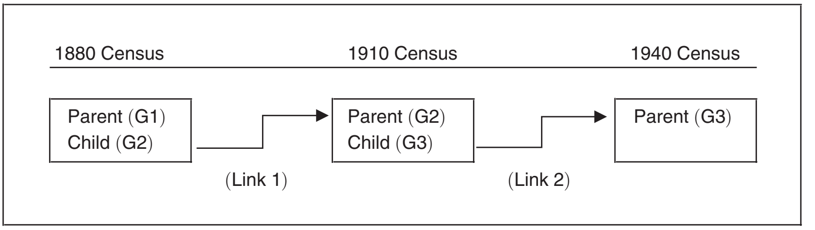

Figure 2. Steps in the Linking Process

VI. The Persistence of Occupations across Two Generations

A. *Persistence from the First to Second Generation*

In this section I demonstrate that ethnic occupational differentials persisted across two generations more strongly than predicted from a standard father-son regression. That is, I compare the convergence prediction of a standard intergenerational model from equation (1) with the convergence of ethnic occupational differentials in equation (3). I run each equation six times: one for each of the four occupational categories (farmer, white-collar, unskilled, and skilled), and one for each of the two occupational income scores (the 1890–1950 score and the 1950 score). The augmented intergenerational model includes the ethnic mean, which is measured as the fraction of the first generation in each occupational category (e.g., fraction of Irish in farming for the farming regression). For occupational income, the ethnic mean is the average logged occupational income score by ethnicity. Though not listed in the equations, I additionally include age controls to address life cycle bias: specifically, a quartic of son's age, a quartic of father's age, and a quartic of son's age interacted with father's occupation. All age controls are normalized to age 40, following Lee and Solon (2009). For all results presented in this paper, I always include these age controls.

Table 2 shows the typical intergenerational result that there is a positive association between the father's occupation and the son's occupation between 1880 and 1910. For instance, having a farmer father in 1880 is associated with a 54 percentage point increase in the likelihood of the son being a farmer. The associations for the other categories of white collar (0.30), skilled (0.17), and unskilled (0.20) workers are smaller, but still positive. After assigning each occupation an income, the intergenerational elasticity for my preferred occupational income measure is 0.50. Given the results from the occupational categories, this 0.50 elasticity reflects the strong association of being a farmer as well as weaker associations for the other occupational categories.

These father-son associations do not predict the convergence of ethnic means. For example, a simple interpretation of the 0.50 elasticity suggests that if I take an occupational income gap between *any two fathers*, only 50 percent of the gap remains for their sons. That is, ethnic means should converge at the same rate. However, this 0.50 rate stands in direct contrast to the persistence of ethnic occupational

Table 2—Persistence from the First Generation in 1880 to the Second Generation in 1910

	Farmer		White-collar	
Father outcome (β_1)	0.543	0.525	0.299	0.299
	(0.005)	(0.006)	(0.010)	(0.010)
Ethnic mean (β_2)		0.336		0.029
		(0.017)		(0.051)
Mean convergence $(\beta_1 + \beta_2)$		0.861		0.328
		(0.017)		(0.052)
	Semiskilled		Unskilled	
Father outcome (β_1)	0.165	0.164	0.202	0.184
	(0.008)	(0.008)	(0.007)	(0.007)
Ethnic mean (β_2)		0.190		0.276
		(0.049)		(0.019)
Mean convergence $(\beta_1 + \beta_2)$		0.354		0.460
		(0.051)		(0.019)
	ln(Occ. Sc.), 1890–1950		ln(Occ. Sc.), 1950	
Father outcome (β_1)	0.497	0.485	0.460	0.445
	(0.008)	(0.008)	(0.008)	(0.008)
Ethnic mean (β_2)		0.435		0.572
		(0.031)		(0.031)
Mean convergence $(\beta_1 + \beta_2)$		0.920		1.019
		(0.031)		(0.032)

Notes: Data are the 1880–1910 link from the 1880–1910–1940 linked dataset. There are 71,702 observations in each regression. The dependent variable is the second generation's outcome, which varies by either occupational category or score in each panel. For example, the top left panel uses farmer as a zero-one dependent variable and the ethnic mean is the share of co-ethnics that are farmers in the prior generation. The 1890–1950 occupational score uses occupational income estimates from 1890 in Sobek (1996) for the 1880 and 1910 censuses, and then the 1950 IPUMS *occscore* estimates for the 1940 census. The 1950 occupational score uses the *occscore* variable throughout. Standard errors are clustered by G1 grandfather. Each regression controls for life cycle bias with a quartic of son's age, quartic of father's age, and quartic of son's age interacted with father's outcome; these quartics are normalized to age 40.

status from the first to the second generations during the Age of Mass Migration (Abramitzky, Boustan, and Eriksson 2014). Therefore, it appears that a standard intergenerational model does not capture all forces that cause ethnic means to persist at a stronger rate.

The discrepancy between the associations at the father-son level and at the ethnic-level is shown in the regressions that include the ethnic mean (see second column in Table 2). These regressions show that, conditional on father's occupation, the son's occupation is additionally correlated with the ethnic mean. The estimated correlation with the ethnic mean is strong. Based on the sum of the ethnic-mean coefficient and the father-son coefficient, ethnic differentials in farming converged at 0.86, which is 59 percent higher than father-son association at 0.54. The mean convergence estimates are also larger than the father-son estimates for the semiskilled category (0.35 versus 0.17) and the unskilled category (0.46 versus 0.20). The white-collar category is different, however, since ethnic differentials in white-collar work converged at statistically the same rate as predicted from an intergenerational model.

The mean convergence of ethnic occupational income differentials was nearly twice as slow as predicted by a standard model (0.92 versus 0.50). A 0.92 estimate indicates there was almost no convergence of ethnic occupational income gaps from the first to second generations. This result may be surprising since it is commonly believed that ethnic differentials faded quickly in the early twentieth century as in the melting pot metaphor, but the result is also consistent with Abramitzky, Boustan, and Eriksson (2014) showing that mean ethnic occupational income was strongly correlated between the first and second generations.

B. *Persistence from the Second to Third Generation*

Having verified that ethnic differentials in occupations converged slowly between the first and second generations, I test whether the same pattern holds for the second and third generations, when attachment to ethnicity may have dwindled due to social assimilation in the first generation. This section exploits a unique feature of my data: direct observation of the third generation via grandparental birthplace. I re-estimate the same two-generation correlations between father and son, but now with data linking from the second generation in 1910 to the third generation in 1940. Note that I estimate how the third-generation outcome is correlated with the ethnic mean of the *second generation*; this is because I am interested in how the rate of convergence may change from the first to second to third generation.

Despite cultural assimilation across generations, third-generation occupations were still correlated with the ethnic mean of the second generation (see Table 3). This correlation holds for each occupational category and for occupational income. Therefore, a standard intergenerational model also fails to capture the convergence of ethnic occupational differentials between the second and third generations. The difference in persistence rates is vast: about 56 percent of ethnic occupational income gaps remained, compared with the prediction from the father-son elasticity that only 28 percent would remain. While the persistence of ethnic differentials between the second and third generations (0.56) was less strong than the persistence between the first and second generations (0.92), a mean convergence rate of 0.56 was still twice the magnitude as estimated by a standard father-son regression (0.28).

The strong persistence of ethnic differentials between the second and third generations is surprising since there was quick cultural assimilation in the first generation (Abramitzky, Boustan, and Eriksson 2016; Alexander and Ward 2018; Biavaschi, Giulietti, and Siddique 2017). It is true that the mean convergence elasticity is smaller between the second and third generations than between the first and second generations (0.56 versus 0.92), so an ethnic-specific causal effect may have dwindled in importance across generations. Yet since ethnic gaps were still slower to converge than predicted by a standard model, the mechanism may be related to something that is not ethnic-specific, but only correlated with the ethnic mean. For instance, it may be that ethnicities varied in their likelihood to live in different areas of the country (e.g., in rural or urban counties), and this difference both correlates with the ethnic mean and affects the son's occupation. Another possibility is that there was no causal mechanism, but rather that I am simply

Table 3—Persistence from the Second Generation in 1910 to the Third Generation in 1940

	Farmer		White-collar	
Father outcome (β_1)	0.317	0.309	0.294	0.290
	(0.005)	(0.005)	(0.008)	(0.008)
Ethnic mean (β_2)		0.173		0.916
		(0.012)		(0.067)
Mean convergence ($\beta_1 + \beta_2$)		0.482		1.206
		(0.012)		(0.067)
	Semiskilled		Unskilled	
Father outcome (β_1)	0.108	0.106	0.096	0.091
	(0.007)	(0.007)	(0.008)	(0.008)
Ethnic mean (β_2)		0.208		0.238
		(0.043)		(0.035)
Mean convergence ($\beta_1 + \beta_2$)		0.314		0.330
		(0.043)		(0.036)
	ln(Occ. Sc.), 1890–1950		ln(Occ. Sc.), 1950	
Father outcome (β_1)	0.275	0.266	0.329	0.317
	(0.005)	(0.005)	(0.007)	(0.007)
Ethnic mean (β_2)		0.295		0.420
		(0.019)		(0.025)
Mean convergence ($\beta_1 + \beta_2$)		0.562		0.737
		(0.019)		(0.025)

Notes: Data are the 1910–1940 link from the 1880–1910–1940 linked dataset. There are 96,726 observations in each regression. The dependent variable is the third generation's outcome, which varies by either occupational category or score in each panel. For example, the top left panel uses farmer as a zero-one dependent variable and the ethnic mean is the share of co-ethnics that are farmers in the prior generation. The 1890–1950 occupational score uses occupational income estimates from 1890 in Sobek (1996) for the 1880 and 1910 censuses, and then the 1950 IPUMS *occscore* estimates for the 1940 census. The 1950 occupational score uses the *occscore* variable throughout. Standard errors are clustered by G2 father. Each regression controls for life cycle bias with a quartic of son's age, quartic of father's age, and quartic of son's age interacted with father's outcome; these quartics are normalized to age 40.

measuring the father's occupation with error and that the ethnic mean captures information about the father's true occupational status. I will later directly test for the importance of geography and measurement error in driving the persistence of ethnic occupational differentials, but first I measure how ethnic gaps persisted across three generations.

V. The Persistence of Occupations across Three Generations

Given that ethnic differentials persisted strongly from the first to second generation and from the second to third generation, they likely also persisted strongly from the first to third generation. Indeed, I have already shown that this was the case in Figure 1 where the elasticity between the first-generation and third-generation ethnic average was 0.51. However, I do not know how this estimate compares with a grandfather-grandson elasticity in a multigenerational model. Now, I re-estimate the same father-son equations as in the previous section, but this time between grandfather and grandson.

Table 4—Persistence from the First Generation in 1880 to the Third Generation in 1940

	Farmer		White-collar	
Grandfather's outcome (β_1)	0.226 (0.005)	0.216 (0.005)	0.204 (0.010)	0.205 (0.010)
Ethnic mean (β_2)		0.173 (0.013)		−0.077 (0.058)
Mean convergence ($\beta_1 + \beta_2$)		0.390 (0.013)		0.128 (0.058)
	Semiskilled		Unskilled	
Grandfather's outcome (β_1)	0.031 (0.007)	0.030 (0.007)	0.045 (0.007)	0.042 (0.007)
Ethnic mean (β_2)		0.155 (0.043)		0.050 (0.017)
Mean convergence ($\beta_1 + \beta_2$)		0.185 (0.044)		0.092 (0.018)
	ln(Occ. Sc.), 1890–1950		ln(Occ. Sc.), 1950	
Grandfather's outcome (β_1)	0.214 (0.005)	0.206 (0.005)	0.284 (0.008)	0.272 (0.008)
Ethnic mean (β_2)		0.303 (0.022)		0.463 (0.032)
Mean convergence ($\beta_1 + \beta_2$)		0.508 (0.023)		0.734 (0.032)

Notes: Data are the 1910–1940 link from the 1880–1910–1940 linked dataset. There are 96,726 observations in each regression. The dependent variable is the third generation's outcome, which varies by either occupational category or score in each panel. For example, the top left panel uses farmer as a zero-one dependent variable and the ethnic mean is the share of co-ethnics that are farmers in two generations prior. The 1890–1950 occupational score uses occupational income estimates from 1890 in Sobek (1996) for the 1880 and 1910 censuses, and then the 1950 IPUMS *occscore* estimates for the 1940 census. The 1950 occupational score uses the *occscore* variable throughout. Standard errors are clustered by G2 father. Each regression controls for life cycle bias with a quartic of son's age, quartic of father's age, and quartic of son's age interacted with father's outcome; these quartics are normalized to age 40.

Table 4 shows that grandfather-grandson occupational income elasticity was 0.21—which is 40 percent of the mean convergence elasticity at 0.51.[26] Once again, farming was a key force driving the occupational income elasticity since farming was the most strongly persistent occupational category with a mean convergence of 0.39 (relative to the grandfather-grandson association of 0.23). The difference in mean convergence and the grandfather-grandson association was also large for the semiskilled and unskilled categories (0.19 versus 0.03 for semiskilled and 0.09 versus 0.05 for unskilled jobs). However, there was no statistical difference in convergence rates at the ethnic-level and grandfather-grandson level for white-collar jobs. Overall, this table shows the main result of the paper: ethnic occupational gaps were much slower to close across three generations than predicted by a multigenerational model.

[26] I estimate the group-averaged elasticities in Table A1 based on equation (2). These group-averaged elasticities are always within the error bounds of the mean convergence elasticities in the main specification. That is, the elasticity when summing the grandfather and ethnic mean coefficients.

A. *Robustness of Three-Generational Result*

It is possible that non-European sources are driving the stronger persistence of ethnic gaps. In particular, one may be concerned that ethnic occupational differentials persisted because Mexican Americans tended to hold lower-paid occupations across three generations, which may be due to unique structural barriers for upward mobility (Alba, Lutz, and Vesselinov 2001; Kosack and Ward 2019). After keeping only Europeans in the sample (dropping Canadians in addition to Mexicans), the persistence of ethnic means for the main 1890–1950 occupational score drops only from 0.51 to 0.47 (see Table A2). Therefore, the reason for the lack of ethnic occupational income convergence was not outliers outside of Europe.

I further show that the results are robust to different ways of grouping occupations or measuring occupational income in Table A3. While the magnitude of convergence rates varies by occupational income score, the main result holds where ethnic differentials converged at least twice as slowly as predicted by a standard multigenerational model. For example, if one allows for regional variation in occupational income, as in the income score created by Collins and Wanamaker (2017), then convergence of ethnic means appears to be three times slower than the grandfather-grandson estimate, and the magnitude of persistence was much higher than for my preferred occupational score. This result could be because occupational income does not capture the persistence of income due to ethnicities living in higher or lower income areas. Another concern is that one measure of farmers' income is inadequate to cover the heterogenous farmer population in 1880. However, the results continue to hold if one uses state-level estimates of farmer income from the 1880 Census of Agriculture, following Abramitzky, Boustan, and Eriksson's (2012) assumptions for imputing farmer income.[27] Finally, I also show that when one uses wage income from 1940 instead of occupational income, then ethnic differentials were also slower to close than predicted by a standard model; however, this measure is limited since I cannot measure self-employed income.

Besides using occupational or wage outcomes, I also show that ethnic differentials in human capital persisted more strongly than predicted from a grandfather-grandson association, which is measured with years of education in 1940 and literacy in 1880. However, this human capital result is not strong evidence since the literacy variable (i.e., can read and write in any language) recorded over 99 percent of grandfathers in 1880 as literate. Therefore, my main results are for occupational gaps.

B. *Iterating an AR*(1) *Model*

A common finding is that an iterated AR(1) model from two generations fails to predict convergence across three generations (e.g., Stuhler 2012, Solon 2018). Does an iterated AR(1) model for ethnic averages also fail to predict the convergence of ethnic gaps? In this section, I compare the actual three-generation associations with the predicted associations, where the predicted associations are created

[27] The state-level estimates of farmer income come after using the assumptions of Abramitzky, Boustan, and Eriksson (2012) in online Appendix Table 3.

by iterating the two-generation results. For example, the predicted association between the grandfather and grandson is the product of the grandfather-father association between 1880–1910 in Table 2 and the father-grandson association between 1910–1940 in Table 3; I then compare this product to the actual association I estimated between 1880–1940 in Table 4. See online Appendix B for how to predict the first-generation ethnic mean coefficient.

As opposed to the common result where the iterated AR(1) model overestimates the convergence of occupational income gaps, an AR(1) model accurately predicts the convergence of ethnic occupational income gaps (see Table 5). For example, the iterated AR(1) model predicts that 52 percent of ethnic occupational income gaps remain after three generations; this estimate is surprisingly close to the actual mean convergence of 51 percent. However, the results are not as accurate for different occupational categories; in fact, sometimes the iterated AR(1) model underestimates the convergence within occupational categories, in contrast to the standard result where the AR(1) model overestimates the convergence.

While an AR(1) ethnic-mean model accurately predicts the convergence of ethnic occupational income gaps, they still follow the common AR(1) result where the grandfather's importance is understated. This is seen in the middle of Table 5 where the predicted grandfather coefficient is smaller than the actual grandfather coefficient. At the same time, the predicted ethnic-mean coefficient is larger than the actual coefficient (see bottom of Table 5). Therefore, the iterated AR(1) ethnic mean model simultaneously *understates* the importance of grandfather's occupation and *overstates* the importance of the first-generation ethnic mean. These biases cancel out such that the sum of the grandfather and first-generation ethnic mean coefficients in the AR(1) model is close to the actual sum.

The differences between the predicted convergence in the AR(1) model and the actual convergence suggests that I have misspecified influences on the grandson's occupation. I estimate a full-fledged AR(2) model where I include (i) the father's occupation, (ii) the grandfather's occupation, (iii) the ethnic mean for the father's generation, and (iv) the ethnic mean for the grandfather's generation. The results are shown in Table B1 in the online Appendix.[28] For occupational income, the coefficients on the grandfather's and grandson's occupations are both positive (0.23 for father and 0.10 for grandfather), which is consistent with the magnitude and results from other three-generational studies during the nineteenth and early twentieth centuries (Long and Ferrie 2018; Olivetti, Paserman, and Salisbury 2018). However, I also find a *positive* coefficient on the second-generation ethnic mean, but a *negative* coefficient on the first-generation ethnic mean, which measures that the "ethnic mean" effect converges at a faster-than-geometric rate.

If ethnic averages followed the family result that gaps across families converged at a slower-than-geometric rate, then ethnic gaps should have also converged at a slower-than-geometric rate. Yet, I find that ethnic gaps converged at a geometric rate. These results are reconciled because the importance of the "ethnic mean" component dwindles rapidly across generations, at a faster-than-geometric rate. Note

[28] See equation (B3) from online Appendix B. I also estimate the AR(2) model after collapsing the data to the ethnic level in online Appendix Table B2.

Table 5—Predicted and Actual Association between First Generation in 1880 and Third Generation in 1940

	Farmer	White-collar	Skilled	Unskilled	ln(Occ. Sc.), 1890–1950	ln(Occ. Sc.), 1950
Predicted mean convergence	0.414	0.396	0.113	0.153	0.517	0.745
	(0.013)	(0.066)	(0.022)	(0.018)	(0.024)	(0.035)
Actual mean convergence	0.390	0.128	0.185	0.092	0.508	0.734
	(0.013)	(0.058)	(0.044)	(0.018)	(0.022)	(0.032)
Difference (actual − predicted)	−0.024	−0.268	0.072	−0.061	−0.009	−0.011
p-value	*0*	*0*	*0.106*	*0*	*0.607*	*0.663*
Predicted grandfather coefficient	0.162	0.087	0.0175	0.017	0.127	0.142
	(0.003)	(0.004)	(0.001)	(0.002)	(0.003)	(0.004)
Actual grandfather coefficient	0.216	0.205	0.030	0.042	0.206	0.272
	(0.005)	(0.010)	(0.007)	(0.007)	(0.005)	(0.008)
Difference (actual − predicted)	0.054	0.118	0.013	0.025	0.079	0.131
p-value	*0*	*0*	*0.0766*	*0*	*0*	*0*
Predicted ethnic mean coefficient	0.251	0.309	0.096	0.136	0.390	0.603
	(0.013)	(0.066)	(0.022)	(0.018)	(0.024)	(0.035)
Actual ethnic mean coefficient	0.173	−0.077	0.155	0.050	0.303	0.463
	(0.012)	(0.058)	(0.043)	(0.017)	(0.022)	(0.031)
Difference (actual − predicted)	−0.078	−0.382	0.059	−0.086	−0.087	−0.140
p-value	*0*	*0*	*0.179*	*0*	*0*	*0*

Notes: Data are from the 1880–1910–1940 linked sample. There are 96,726 observations in each regression. The predicted columns assume an AR(1) process using coefficients from Tables 2 and 3; see online Appendix B for equations to predict the coefficients. The actual correlation is taken from Table 4. The dependent variable is the third generation's outcome, which varies by occupational category or score in each row. The 1890–1950 occupational score uses occupational income estimates from 1890 in Sobek (1996) for the 1880 and 1910 censuses, and then the 1950 IPUMS *occscore* estimates for the 1940 census. The 1950 occupational score uses the *occscore* variable throughout. Standard errors are clustered by G1 father. The *p*-value tests the difference between the predicted coefficient and actual coefficient.

that if the mechanism for the stronger persistence of ethnic gaps were due to ethnic spillovers, then one would expect a positive coefficient on both the first-generation and second-generation ethnic means since both generations may influence grandchildren; however, this is not what I find in the data. The dwindling importance of the ethnic-mean effect could be due to social or economic assimilation in the first two generations where attachment to ethnicity or culture falls due to longer exposure to the US environment. This result also suggests that the multigenerational drag from factors correlated with ethnicity ultimately fade and that only the family component remains. Yet, of course, all the estimates are correlational and not causal, so it is unclear what is driving these associations; further, they may be partially driven by measurement error.

VI. Reasons for the Persistence of Ethnic Gaps

A. *Measurement Error*

Why did ethnic occupational gaps persist so strongly during the Age of Mass Migration? There are many possible causal mechanisms, such as labor market discrimination, the quality of schools, or the influence of role models; unfortunately, most cannot be explored in the data. In this section I instead examine the alternative possibility that there was no causal mechanism, but that the difference across models is due to measurement error. That is, the reason I find a statistical correlation with the ethnic mean may be that the ethnic mean provides more information about the father's true occupational status (Solon 1992, Clark 2014).

To address the possibility of measurement error driving the results, I take the strategy of Solon (1992) and use multiple reports of the father's occupation. While I would prefer to test the importance of measurement error for the grandfather-grandson regressions, unfortunately I cannot observe more than one grandfather observation because the closest censuses to 1880 do not have this information cleaned and available.[29] Instead, I use multiple observations of the father's occupation. To gain another observation of the father, I link the 1910 fathers to the 1920 census using the same machine-learning techniques from Feigenbaum (2016). I link 45 percent of the fathers, leaving me with a two-generational dataset of 42,410 G3 sons and G2 fathers (see online Appendix B for more information on the 1910–1920 linking process).

Measurement error is indeed one reason why I find a difference between the father-son association and the ethnic mean convergence rate. This result is seen in Table 6, where I use the simple average of the father's outcome across the 1910 and 1920 censuses to proxy for the father's permanent economic status. Table 6 shows that when only having one father observation from 1910, the intergenerational elasticity coefficient is 0.27. When averaging two observations of the father's log occupational income, then the elasticity increases by 21 percent from 0.27 to 0.33. This increase is consistent with the common measurement error relationship shown by Solon (1992) for income-based measures and Mazumder and Acosta (2015) for occupational-based measures. Interestingly, the ethnic mean coefficient falls by the same magnitude as the father-son coefficient increases such that they cancel each other out. This suggests that the "ethnic-mean effect" is partially picking up unobservable information about the father.

The two-observation elasticity of 0.33 is still far from the mean convergence elasticity of 0.55, suggesting that measurement error is not fully driving the results. However, two father observations still may not capture the permanent occupational status of the father. Under the assumption of classical measurement error, I can predict what the true father-son elasticity would be based on the movement from the one- to two-observation elasticity (0.27 to 0.33). Based on the classical error assumption, the projected true father-son elasticity would be 54 percent higher at 0.42.[30] Since the true elasticity of 0.42 is less than the mean convergence elasticity of 0.55, measurement error is not the sole reason why I find stronger persistence of occupational income at the ethnic level. However, it also suggests that the "effect"

[29] The 1870 census has yet to be cleaned by IPUMS. The microdata from the 1890 census were lost in a fire.

[30] With classical measurement error, $\text{plim}\,\hat{\beta}_1 = \beta_1\left(\text{var}\left(y_f^*\right)/\left(\text{var}\left(y_f^*\right) + \left(\text{var}(v)/T\right)\right)\right)$, where y_f^* is the true occupational income of the father, v is a measurement error term, and T is the number of observations of the father (Zimmerman 1992). Going from one to two observations of the father's occupation (or $T = 1$ to $T = 2$), the estimated $\hat{\beta}_1$ goes from 0.271 to 0.329. Plugging these numbers into the equation yields a β_1 of 0.419.

Table 6—Persistence from the Second Generation in 1910 to the Third Generation in 1940, Using Two Observations of the Father's Occupation

	Farmer		White-collar	
Father's outcome in 1910 (β_1)	0.284		0.286	
	(0.004)		(0.006)	
Father's mean outcome in 1910/1920 (β_1)		0.335		0.386
		(0.004)		(0.006)
Ethnic mean in 1910 (β_2)	0.175	0.119	0.976	0.835
	(0.016)	(0.016)	(0.096)	(0.095)
Mean convergence ($\beta_1 + \beta_2$)	0.459	0.454	1.262	1.221
	(0.016)	(0.016)	(0.096)	(0.095)
	Semiskilled		Unskilled	
Father's outcome in 1910 (β_1)	0.091		0.078	
	(0.006)		(0.006)	
Father's mean outcome in 1910/1920 (β_1)		0.128		0.119
		(0.006)		(0.007)
Ethnic mean in 1910 (β_2)	0.240	0.210	0.204	0.170
	(0.061)	(0.061)	(0.051)	(0.051)
Mean convergence ($\beta_1 + \beta_2$)	0.331	0.338	0.282	0.289
	(0.061)	(0.061)	(0.050)	(0.050)
	ln(Occ. Sc.), 1890–1950		ln(Occ. Sc.), 1950	
Father's outcome in 1910 (β_1)	0.271		0.320	
	(0.004)		(0.005)	
Father's mean outcome in 1910/1920 (β_1)		0.329		0.424
		(0.004)		(0.006)
Ethnic mean in 1910 (β_2)	0.307	0.236	0.441	0.314
	(0.028)	(0.027)	(0.037)	(0.037)
Mean convergence ($\beta_1 + \beta_2$)	0.578	0.565	0.761	0.738
	(0.027)	(0.027)	(0.037)	(0.036)

Notes: Data are the 1910–1920–1940 link. There are 42,410 observations in each regression. The dependent variable is the third generation's outcome, which varies by either occupational category or score in each panel. For example, the top left panel uses farmer as a zero-one dependent variable and the ethnic mean is the co-ethnics that are farmers in 1910. The 1890–1950 occupational score uses occupational income estimates from 1890 in Sobek (1996) for the 1880 and 1910 censuses, and then the 1950 IPUMS *occscore* estimates for the 1940 census. The 1950 occupational score uses the *occscore* variable throughout. The father's mean outcome averages his occupational measure in 1910 and 1920. Standard errors are clustered by G2 father. Each regression controls for life cycle bias with a quartic of son's age, quartic of father's age, and quartic of son's age interacted with father's outcome; these quartics are normalized to age 40.

of the ethnic mean is much less than expected from a regression with one father observation.

While measurement error is not fully driving differences between the mean convergence elasticity and father-son elasticity, it remains a key explanation for the different magnitudes. If I assume that attenuation bias for the grandfather-grandson relationship is of the same proportion that I find between father and son (that is, the true elasticity is 54 percent higher than the one-observation elasticity), then grandfather-grandson elasticity would increase from the observed elasticity of 0.21 to a "true" elasticity of 0.32. A 0.32 grandfather-grandson elasticity is still less than my three-generation mean convergence of 0.51. Overall, these results confirm that measurement error partially explains why I find a positive coefficient for ethnic mean, but at the same time measurement error does not appear to fully explain the positive coefficient.

B. *The Importance of Geography*

Since measurement error does not fully drive results, this suggests that there is some factor that is correlated with the ethnic mean that causes ethnic gaps to persist more strongly than predicted between grandfather and grandson. In this section I explore whether factors related to geography are driving the results. For example, one reason for the slow convergence of ethnic gaps may be that immigrants lived in different types of neighborhoods, and outcomes were largely influenced by childhood location. There is already evidence that this may be the case. For instance, Borjas (1995) shows that the correlation between the ethnic mean and the next generation's income narrows substantially when controlling for neighborhood. This suggests that the ethnic mean proxies for general neighborhood effects, such as access to high-quality schools or jobs. Moreover, recent evidence from Chetty and Hendren (2018a,b) demonstrates that childhood neighborhood matters for long-run outcomes in modern-day data; here, I can directly control for some of these unobservable neighborhood effects.

It may not be apparent why geography is important across three generations, but a key feature in the data is that grandsons tended to live in the same areas as their grandfathers. This is best seen in Figure 3, which plots where immigrants from Ireland and Germany lived in 1880 and where their descendants lived in 1940. Initially, the first generation was clustered into different parts of the country, such as the Irish in the Northeast and Germans in the Midwest and Mid-Atlantic. After 60 years, the grandchildren of Irish and German immigrants still lived in the same areas of the country. While this is a straightforward result of children living in the same location as their parents, it also shows that *spatial* assimilation, or the even spread of ethnicities across the country, did not take place (Eriksson and Ward 2018). The lack of geographic mobility across generations could imply little income convergence given the large per capita income gaps across regions between 1880 and 1940 (Mitchener and McLean 1999).

To account for geography, I include 1880 enumeration-district fixed effects in my grandfather-grandson associations (i.e., equation (3)). Note that this method also controls for the childhood enumeration district of the G2 father. Therefore, this equation estimates the mean convergence net of the neighborhood effects. A benefit of this method is that enumeration districts are local areas of about 1,600 in 1880, so I capture numerous unobservables about the local environment. However, the ethnic mean coefficient can only be identified for those from mixed-ethnic enumeration districts, which are 77 percent of my data. I will also estimate results when controlling for fixed effects at the county level, which allows me to identify the ethnic mean coefficient for nearly the entire dataset; the basic argument holds if one uses either enumeration district or county fixed effects.

Ethnic occupational differentials converged much more rapidly for those with grandfathers from the same 1880 neighborhood. This is seen in Table 7, where I show the grandson's association with the grandfather's occupational income and

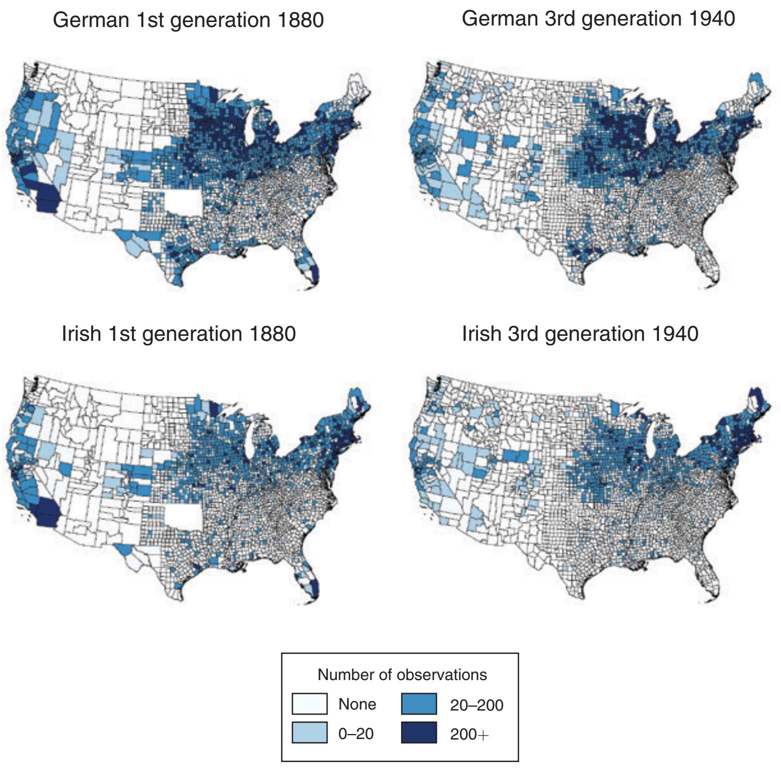

Figure 3. Persistence of Location from the First Generation in 1880 to Third Generation in 1940

Notes: Data report the (weighted) number of observations for the first generation in 1880 and the third generation in 1940.

the first-generation ethnic mean, both with and without enumeration district fixed effects.[31] The mean convergence, net of neighborhood fixed effects, was faster than the gross convergence without fixed effects. For example, with the preferred occupational income score, the estimated convergence without fixed effects is 0.51; with fixed effects, it was 0.16. Therefore, net of 1880 neighborhood factors that are correlated with both the ethnic mean and the grandson's occupation, ethnic differentials converged three times as rapidly as the gross convergence for the entire country.

There are two reasons why the convergence is more rapid when controlling for the grandfather's neighborhood. First, the grandfather coefficient falls in magnitude by half. This result suggests that the original correlation between the grandfather and grandson partially captures effects from where the grandfather lived. The second reason is that the coefficient on the first-generation ethnic mean also drops, this time by about 82 percent. This result suggests that there is little multigenerational drag for co-ethnics within the grandfather's neighborhood; rather, ethnic

[31] See Table A4 for occupation category results, which show a similar pattern as Table 7 where the ethnic mean coefficient drops after controlling for enumeration district fixed effects.

Table 7—Association between First and Third Generation after Accounting for Neighborhood

	ln(Occ. Sc.), 1890–1950		ln(Occ. Sc.), 1950	
Panel A. Use enumeration district fixed effects:				
Grandfather outcome	0.206	0.099	0.272	0.123
	(0.005)	(0.007)	(0.008)	(0.009)
Ethnic mean	0.303	0.056	0.463	0.088
	(0.022)	(0.030)	(0.032)	(0.041)
Mean convergence	0.508	0.155	0.734	0.210
	(0.023)	(0.031)	(0.032)	(0.042)
1880 neighborhood fixed effects	N	Y	N	Y
Panel B. Use county fixed effects:				
Grandfather outcome	0.205	0.140	0.270	0.179
	(0.005)	(0.005)	(0.008)	(0.008)
Ethnic mean	0.298	0.091	0.457	0.144
	(0.022)	(0.022)	(0.032)	(0.032)
Mean convergence	0.504	0.231	0.728	0.323
	(0.022)	(0.023)	(0.032)	(0.032)
1880 county fixed effects	N	Y	N	Y
Panel C. Drop farmers and use enumeration district fixed effects:				
Grandfather outcome	0.063	0.039	0.075	0.045
	(0.009)	(0.010)	(0.011)	(0.013)
Ethnic mean	0.212	0.075	0.384	0.138
	(0.039)	(0.043)	(0.059)	(0.064)
Mean convergence	0.275	0.114	0.459	0.182
	(0.040)	(0.044)	(0.060)	(0.064)
1880 neighborhood fixed effects	N	Y	N	Y

Notes: Data are from the 1880–1940 portion of the 1880–1910–1940 linked sample. There are 96,726 observations in each regression. The dependent variable is the third generation (G3) outcomes, which vary across panels. The 1890–1950 occupational score uses occupational income estimates from 1890 in Sobek (1996) for the 1880 and 1910 censuses, and then the 1950 IPUMS *occscore* estimates for the 1940 census. The 1950 occupational score uses the *occscore* variable throughout. I control for life cycle effects with a quartic of grandson's age, quartic of grandfather's age, and quartic of grandson's age interacted with grandfather's outcome; these quartics are normalized to age 40. Standard errors are clustered by G1 grandfather. The neighborhood fixed effect controls for the 1880 enumeration district.

differentials matter less if one descends from the same area of the country. For my preferred occupational score, the coefficient on the first-generation mean is marginally statistically significant at the 10 percent level.[32]

The basic pattern that ethnic occupational differentials converged more rapidly for descendants from the same areas of the country also holds if I use 1880 county fixed effects instead of enumeration district fixed effects. County fixed effects

[32] One potential issue with the results in Table 7 is that the ethnic mean is measured at the national level rather than at the local level. To the extent that the ethnic mean effect operates through mechanisms related to interactions or local channels, the national ethnic mean may matter little for those in a given enumeration district. Fortunately, since I have the complete-count 1880 census data, I can instead measure the ethnic mean at the enumeration district level. The results from this specification are similar to when using the national-level ethnic mean, in that the convergence of ethnic differentials falls in magnitude after including enumeration-district fixed effects (see online Appendix Table A5).

may be preferable because 99 percent of my grandfather-grandson links had a mixed-ethnic county in 1880. Controlling for county fixed effects also causes the mean convergence coefficient to fall, but not as steeply as when controlling for enumeration district fixed effects. For my preferred occupational score, controlling for county fixed effects causes a drop in the mean convergence elasticity from 0.51 to 0.23; the enumeration district control led to a further drop to 0.16 (see panel B in Table 7). These results are consistent with the argument that mechanisms related to geography are being captured by the ethnic-mean coefficient.

Location fixed effects may simply be capturing the effect of growing up in a rural community. Given that one was raised in a farming area, he may be more likely to be a farmer, no matter his ethnicity. In general, farming is key for understanding historical occupational income elasticities; for instance, Olivetti and Paserman (2015) show that two-generational elasticities between 1850 and 1940 fall by about half after dropping farmers. Indeed, I have already shown that farming was the most persistent occupational category, at least when compared with the white-collar, unskilled, or semiskilled categories.

Farming is a key reason for why ethnic gaps in occupational income closed slowly for descendants of immigrants in the Age of Mass Migration. Going back to the specification without enumeration district fixed effects, the persistence of ethnic differentials drops from 0.51 with farmers to 0.28 without farmers (see panel C in Table 7). Part of this drop is because the grandfather-grandson elasticity drops from 0.21 to 0.06. However, even though farmers were dropped, there is still an extra association between the grandson's occupational income and the first-generation ethnic mean (0.21), indicating that ethnic differentials persisted more strongly than predicted from a standard multigenerational model. Therefore, multigenerational models still fail to capture the convergence of ethnic differentials for non-farming occupations.

If the enumeration district fixed effects were solely capturing mechanisms related to farming, then adding these fixed effects to the non-farmer specification should not strongly influence estimates of persistence. Yet, when one includes enumeration district fixed effects in the non-farmer sample, then the persistence of ethnic differentials drops from 0.28 to 0.11. The ethnic-mean coefficient (0.08) with location fixed effects is marginally statistically significant at the 10 percent level. This result suggests that there are unobserved neighborhood factors that are both correlated with the non-farming ethnic mean in 1880 and the non-farming grandson's occupation in 1940. Therefore, the geographical controls appear to capture other aspects of the environment that shape outcomes across three generations that are unrelated to farming.

The results in this section suggest that factors related to geography and farming are key for understanding the persistence of occupational differentials across different sources during the Age of Mass Migration. However, I do not know the precise causal mechanisms. Geographic reasons could be related to anything associated with different environments across the country, such as having a different quality schooling or health environment. It is also possible that the neighborhood fixed effects capture sorting rather than a causal effect of place; it may be that whatever led immigrants to live in different areas of the country also led their grandchildren to have more similar outcomes.

VII. Conclusions

In this paper I measure the convergence of ethnic occupational differentials across three generations, and test whether they converge at a slower rate than predicted by a standard multigenerational model. With a new dataset of first-generation immigrant grandfathers in 1880 linked to third-generation native grandsons in 1940, I show that the grandson's occupation is correlated with the occupations of co-ethnics in the first generation 60 years prior, even after accounting for the grandfather's occupation. Since this extra correlation exists, an ethnicity with high occupational income in the first generation is also more likely to have high occupational income in the second and third generations. The results show that a standard multigenerational model fails to predict the persistence of ethnic group averages across three generations.

The primary goal of this paper is to measure the convergence of ethnic occupational differentials across three generations. Unfortunately, I do not fully uncover the causal mechanisms behind slow rates of convergence. I provide evidence that measurement error partially explains why ethnic differentials are slower to converge than predicted by a grandfather-grandson association, because the ethnic mean captures information about the grandfather's occupational status. I further argue that factors related to geography partially explain why there is slower convergence across generations. Farming was a particularly persistent occupation. It may be that children entered farming both because their father was a farmer and because they were raised in a farming community—not necessarily because of an ethnic-specific causal mechanism. Overall, the results suggest that location matters, whether because of the childhood environment or because of exposure to different types of labor markets (Chetty, Hendren, and Katz 2016; Chetty and Hendren 2018a, 2018b), however, I am unable to show causal evidence for this result.

The results in this paper come from a dataset of mostly European groups between 1880 and 1940, and therefore do not apply to today's immigrants from Central America, South Asia, and East Asia. Yet there are many reasons to expect the persistence of ethnic gaps to be stronger for more recent times than during the Age of Mass Migration. For example, immigrant segregation has been increasing since the mid-twentieth century, and segregation reinforces group averages over time (Borjas 1995; Cutler, Glaeser, and Vigdor 2008). Moreover, a long line of sociology work suggests that assimilation for recent migrant groups is different from historical European assimilation due to more salient ethnic differences for recent immigrants (Portes and Zhou 1993; Perlmann 2005). On the other hand, Asian immigrants are particularly upwardly mobile between the first and second generations, which may cause the persistence of group averages to be less than in a standard intergenerational model (Hilger 2016).

Unfortunately, our understanding of how recent ethnic differentials evolve past the second generation is limited due to measurement issues related to the ancestry question in the census. Today only a few datasets include the ideal variable of grandparent's country of birth, and these often contain too few observations to accurately estimate ethnic outcomes for the variety of immigrant origins. It may be that the increasing availability of linked census data will lead to a more complete understanding of the convergence (or lack thereof) of ethnic differentials for entrants in the second half of the twentieth century. This issue is increasingly important as more third-generation Americans from Latin America and Asia will enter the labor market in the next few decades.

REFERENCES

Aaronson, Daniel, and Bhashkar Mazumder. 2008. "Intergenerational Economic Mobility in the United States, 1940 to 2000." *Journal of Human Resources* 43 (1): 139–72.

Abramitzky, Ran, and Leah Boustan. 2017. "Immigration in American Economic History." *Journal of Economic Literature* 55 (4): 1311–45.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2012. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review* 102 (5): 1832–56.

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

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2016. "Cultural Assimilation during the Age of Mass Migration." NBER Working Paper 22381.

Ager, Philipp, Leah Boustan, and Katherine Eriksson. 2019. "The Intergenerational Effects of a Large Wealth Shock: White Southerners after the Civil War." NBER Working Paper 25700.

Alba, Richard, Amy Lutz, and Elena Vesselinov. 2001. "How Enduring Were the Inequalities among European Immigrant Groups in the United States?" *Demography* 38 (3): 349–56.

Alba, Richard, and Victor Nee. 2009. *Remaking the American Mainstream: Assimilation and Contemporary Immigration*. Cambridge, MA: Harvard University Press.

Alexander, Rohan, and Zachary Ward. 2018. "Age at Arrival and Assimilation during the Age of Mass Migration." *Journal of Economic History* 78 (3): 904–37.

Åslund, Olof, Per-Anders Edin, Peter Fredriksson, and Hans Grönqvist. 2011. "Peers, Neighborhoods, and Immigrant Student Achievement: Evidence from a Placement Policy." *American Economic Journal: Applied Economics* 3 (2): 67–95.

Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey. 2017. "How Well Do Automated Methods Perform in Historical Samples? Lessons from U.S. Historical Data." NBER Working Paper 24019.

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

Black, Sandra E., and Paul J. Devereux. 2011. "Recent Developments in Intergenerational Mobility." In *Handbook of Labor Economics*, Vol. 4, edited by Orley Ashenfelter and David Card, 1487–1541. Amsterdam: North-Holland.

Bleakley, Hoyt, and Joseph Ferrie. 2016. "Shocking Behavior: Random Wealth in Antebellum Georgia and Human Capital across Generations." *Quarterly Journal of Economics* 131 (3): 1455–95.

Borjas, George J. 1992. "Ethnic Capital and Intergenerational Mobility." *Quarterly Journal of Economics* 107 (1): 123–50.

Borjas, George J. 1993. "The Intergenerational Mobility of Immigrants." *Journal of Labor Economics* 11 (1): 113–35.

Borjas, George J. 1994. "Long-Run Convergence of Ethnic Skill Differentials: The Children and Grandchildren of the Great Immigration." *Industrial and Labor Relations Review* 47 (4): 553–73.

Borjas, George J. 1995. "Ethnicity, Neighborhoods, and Human-Capital Externalities." *American Economic Review* 85 (3): 365–90.

Borjas, George J. 2001. "Long-Run Convergence of Ethnic Skill Differentials, Revisited." *Demography* 38 (3): 357–61.

Braun, Sebastian Till, and Jan Stuhler. 2018. "The Transmission of Inequality across Multiple Generations: Testing Recent Theories with Evidence from Germany." *Economic Journal* 128 (609): 576–611.

Card, David, John DiNardo, and Eugena Estes. 2000. "The More Things Change: Immigrants and the Children of Immigrants in the 1940s, the 1970s, and the 1990s." In *Issues in the Economics of Immigration*, edited by George J. Borjas, 227–70. Chicago: University of Chicago Press.

Catron, Peter. 2019. "The Citizenship Advantage: Immigrant Socioeconomic Attainment in the Age of Mass Migration." *American Journal of Sociology* 124 (4): 999–1042.

Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. 2017. "The Fading American Dream: Trends in Absolute Income Mobility since 1940." *Science* 356 (6336): 398–406.

Chetty, Raj, and Nathaniel Hendren. 2018a. "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects." *Quarterly Journal of Economics* 133 (3): 1107–62.

Chetty, Raj, and Nathaniel Hendren. 2018b. "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates." *Quarterly Journal of Economics* 133 (3): 1163–1228.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review* 106 (4): 855–902.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." *Quarterly Journal of Economics* 129 (4): 1553–1623.

Clark, Gregory. 2014. *The Son Also Rises: Surnames and the History of Social Mobility*. Princeton, NJ: Princeton University Press.

Collins, William J., and Marianne H. Wanamaker. 2017. "Up from Slavery? African American Intergenerational Economic Mobility since 1880." NBER Working Paper 23395.

Corak, Miles. 2006. "Do Poor Children Become Poor Adults? Lessons from a Cross-Country Comparison of Generational Earnings Mobility." In *Dynamics of Inequality and Poverty*, Vol. 13, edited by John Creedy and Guyonne Kalb, 143–88. Bradford, UK: Emerald Group Publishing Limited.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. 2008. "Is the Melting Pot Still Hot? Explaining the Resurgence of Immigrant Segregation." *Review of Economics and Statistics* 90 (3): 478–97.

Darity, William, Jason Dietrich, and David K. Guilkey. 2001. "Persistent Advantage or Disadvantage?: Evidence in Support of the Intergenerational Drag Hypothesis." *American Journal of Economics and Sociology* 60 (2): 435–70.

Duncan, Brian, Jeffrey Grogger, Ana Sofia Leon, and Stephen J. Trejo. 2017. "New Evidence of Generational Progress for Mexican Americans." NBER Working Paper 24067.

Duncan, Brian, and Stephen J. Trejo. 2011. "Intermarriage and the Intergenerational Transmission of Ethnic Identity and Human Capital for Mexican Americans." *Journal of Labor Economics* 29 (2): 195–227.

Duncan, Brian, and Stephen J. Trejo. 2015. "Assessing the Socioeconomic Mobility and Integration of U.S. Immigrants and Their Descendants." *ANNALS of the American Academy of Political and Social Science* 657 (1): 108–35.

Duncan, Brian, and Stephen J. Trejo. 2017. "The Complexity of Immigrant Generations: Implications for Assessing the Socioeconomic Integration of Hispanics and Asians." *ILR Review* 70 (5): 1146–75.

Dustmann, Christian, and Albrecht Glitz. 2011. "Chapter 4—Migration and Education." In *Handbook of the Economics of Education*, Vol. 4, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessman, 327–439. Amsterdam: North-Holland.

Eriksson, Katherine, and Zachary A. Ward. 2018. "The Ethnic Segregation of Immigrants in the United States from 1850 to 1940." NBER Working Paper 24764.

Feigenbaum, James J. 2015. "Intergenerational Mobility during the Great Depression." https://scholar.harvard.edu/files/jfeigenbaum/files/feigenbaum_jmp.pdf.

Feigenbaum, James J. 2016. "A Machine Learning Approach to Census Record Linking." https://scholar.harvard.edu/files/jfeigenbaum/files/feigenbaum-censuslink.pdf.

Feigenbaum, James J. 2018. "Multiple Measures of Historical Intergenerational Mobility: Iowa 1915 to 1940." *Economic Journal* 128 (612): F446–81.

Ferrie, Joseph, Catherine Massey, and Jonathan Rothbaum. 2016. "Do Grandparents and Great-Grandparents Matter? Multigenerational Mobility in the US, 1910–2013." NBER Working Paper 22635.

Glazer, Nathan, and Daniel P. Moynihan. 1963. *Beyond the Melting Pot: The Negroes, Puerto Ricans, Jews, Italians and Irish of New York City*. Cambridge, MA: MIT Press.

Grawe, Nathan D. 2006. "Lifecycle Bias in Estimates of Intergenerational Earnings Persistence." *Labour Economics* 13 (5): 551–70.

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

Hilger, Nathaniel. 2016. "Upward Mobility and Discrimination: The Case of Asian Americans." NBER Working Paper 22748.

Kosack, Edward and Zachary Ward. 2019. "El sueño americano? The Generational Progress of Mexican Americans prior to World War II." https://drive.google.com/file/d/1Ur40lnbPy9eBxFMifNhEN-JVQz9u8P7z/view.

Lee, Chul-In, and Gary Solon. 2009. "Trends in Intergenerational Income Mobility." *Review of Economics and Statistics* 91 (4): 766–72.

León, Alexis. 2004. "Does 'Ethnic Capital' Matter? Identifying Peer Effects in the Intergenerational Transmission of Ethnic Differentials." https://pdfs.semanticscholar.org/bb37/46b65f8c368ea3fed99f62709d14def69f1c.pdf.

Lindahl, Mikael, Mårten Palme, Sofia Sandgren Massih, and Anna Sjögren. 2015. "Long-Term Intergenerational Persistence of Human Capital an Empirical Analysis of Four Generations." *Journal of Human Resources* 50 (1): 1–33.

Long, Jason, and Joseph Ferrie. 2013. "Intergenerational Occupational Mobility in Great Britain and the United States since 1850." *American Economic Review* 103 (4): 1109–37.

Long, Jason, and Joseph Ferrie. 2018. "Grandfathers Matter(ed): Occupational Mobility across Three Generations in the US and Britain, 1850–1911." *Economic Journal* 128 (612): F422–45.

Mazumder, Bhashkar, and Miguel Acosta. 2015. "Using Occupation to Measure Intergenerational Mobility." *ANNALS of the American Academy of Political and Social Science* 657 (1): 174–93.

Mitchener, Kris James, and Ian W. McLean. 1999. "U.S. Regional Growth and Convergence, 1880–1980." *Journal of Economic History* 59 (4): 1016–42.

Minns, Chris. 2000. "Income, Cohort Effects, and Occupational Mobility: A New Look at Immigration to the United States at the Turn of the 20th Century." *Explorations in Economic History* 37 (4): 326–50.

Olivetti, Claudia, and M. Daniele Paserman. 2015. "In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850–1940." *American Economic Review* 105 (8): 2695–2724.

Olivetti, Claudia, M. Daniele Paserman, and Laura Salisbury. 2018. "Three-Generation Mobility in the United States, 1850–1940: The Role of Maternal and Paternal Grandparents." *Explorations in Economic History* 70: 73–90.

Özek, Umut, and David N. Figlio. 2016. "Cross-Generational Differences in Educational Outcomes in the Second Great Wave of Immigration." NBER Working Paper 22262.

Perlmann, Joel. 2005. *Italians Then, Mexicans Now: Immigrant Origins and the Second-Generation Progress, 1890–2000*. New York: Russell Sage Foundation.

Portes, Alejandro, and Rubén G. Rumbaut. 2006. *Immigrant America: A Portrait*. 3rd ed. Berkeley: University of California Press.

Portes, Alejandro, and Min Zhou. 1993. "The New Second Generation: Segmented Assimilation and its Variants." *ANNALS of the American Academy of Political and Social Science* 530: 74–96.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2018. "IPUMS USA: Version 8.0 [dataset]." IPUMS, Minneapolis, MN. https://doi.org/10.18128/D010.V8.0 (accessed June 7, 2018).

Smith, James P. 2003. "Assimilation across the Latino Generations." *American Economic Review* 93 (2): 315–19.

Smith, James P. 2006. "Immigrants and the Labor Market." *Journal of Labor Economics* 24 (2): 203–33.

Sobek, Matthew. 1996. "Work, Status, and Income: Men in the American Occupational Structure since the Late Nineteenth Century." *Social Science History* 20 (2): 169–207.

Solon, Gary. 1992. "Intergenerational Income Mobility in the United States." *American Economic Review* 82 (3): 393–408.

Solon, Gary. 1999. "Chapter 29—Intergenerational Mobility in the Labor Market." In *Handbook of Labor Economics*, Vol. 3A, edited by Orley C. Ashenfelter and David Card, 1761–1800. Amsterdam: North-Holland.

Solon, Gary. 2018. "What Do We Know so Far about Multigenerational Mobility?" *Economic Journal* 128 (612): F340–52.

Stuhler, Jan. 2012. "Mobility across Multiple Generations: The Iterated Regression Fallacy." IZA Discussion Paper 7072.

Ward, Zachary. 2017. "Birds of Passage: Return Migration, Self-Selection and Immigration Quotas." *Explorations in Economic History* 64: 37–52.

Ward, Zachary. 2020. "Replication Data for: The Not-So-Hot Melting Pot: The Persistence of Outcomes for Descendants of the Age of Mass Migration." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E120064V1.

Xie, Yu, and Alexandra Killewald. 2013. "Intergenerational Occupational Mobility in Great Britain and the United States since 1850: Comment." *American Economic Review* 103 (5): 2003–20.

Zimmerman, David J. 1992. "Regression Toward Mediocrity in Economic Stature." *American Economic Review* 82 (3): 409–29.