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

Zachary Ward[†]

Australian National University

June 2017

Abstract: How many generations does it take for skill gaps across immigrant sources to converge? Using data that links immigrant grandfathers in 1880 to grandsons in 1940, we show that ethnic group averages converge at a slower rate than a standard multigenerational model predicts. Third-generation outcomes are correlated with the average skill level of the first generation, above and beyond the effect of the grandfather. Ethnic skill gaps converge more quickly for descendants of grandfathers from the same 1880 neighborhood, suggesting that the clustering of the first generation into enclaves partially drives persistence of ethnic skill gaps for multiple generations.

JEL Classification: J61, J62, N21, N22

Keywords: intergenerational mobility, ethnic capital, assimilation

[†]email: Zach.A.Ward@gmail.com. Research School of Economics, HW Arndt Building 25A, College of Business and Economics, The Australian National University, Canberra, ACT 2600, Australia.

I.   Introduction

The intergenerational mobility literature measures the transmission of income from father to son, or when the data permits, from grandfather to grandson (Solon, 1999; Black and Devereux, 2011; Solon, 2015). A standard finding from the United States is that about half of an income gap between two fathers disappear by the next generation; however, this prediction is too optimistic for the convergence of group averages.[1] For example, the black-white wage gap has remained wide for over 150 years and is nowhere near full convergence.[2] Group averages based on ethnicity or country of birth are also preserved between first and second-generation immigrants at a higher rate than predicted by an intergenerational elasticity coefficient (Borjas, 1992). This is not a recent phenomenon: skill gaps across European sources during the Age of Mass Migration (1850-1914) closed only slightly between the first and second generations (Abramitzky et al., 2014), in direct contrast with estimates of even higher rates of intergenerational mobility during this earlier period (Long and Ferrie, 2013; Feigenbaum, 2017).

Borjas (1992, 1994, 1995) argues that slower convergence of ethnic skill differentials may be due to varying levels of "ethnic capital" (or average skill level) in the first generation, where the initial skill level acts as a dragging force on convergence for future generations.[3] The provocative implication of this theory is that ethnic skill gaps take hundreds of years to close, much longer than predicted by a standard multigenerational model between grandfather and grandson. Yet there is no evidence to show how an ethnic capital model performs across multiple generations with a dataset that links grandfather to grandson.[4] It may be that ethnic gaps close quickly for the third generation, as in the melting pot metaphor.

In this paper, we test whether ethnic skill differentials converge at a slower rate than predicted by a standard multigenerational model with a large and unique linked dataset of immigrant grandparents in 1880, second-generation fathers in 1910 and third-generation

---

[1] Corak (2006) prefers an intergenerational elasticity coefficient of 0.47 in the United States. Mazumder and Acosta (2015) use recent data from the PSID to estimate it between 0.50 and 0.60.

[2] See Collins and Wanamaker (2017) for a discussion on the black-white wage gap and intergenerational mobility. See also Darity et al. (2001) for a discussion for both black-white mobility and immigrant mobility in a historical setting. There are many other studies on convergence of group averages; most related to our study are Borjas (1994) and Card et al. (2000) who use country of birth or parent's country of birth for immigrants.

[3] The mechanism behind this correlation is unclear. It may relate to various theories, such as human capital spill overs within ethnicity, clustering of immigrants into ethnic enclaves, or discrimination.

[4] Borjas (1994) shows that the average skill level of the first generation in 1910 correlates with the third generation in 1980, but does not have grandfather-grandson data to decompose the effect into ethnic capital or direct influence from the grandfather.

grandchildren in 1940. Therefore, we measure convergence of ethnic skill differentials with a dataset of mostly white Europeans who arrived during the Age of Mass Migration – groups who are believed to have assimilated quickly relative to recent arrivals (Perlmann, 2005). Not only do we observe three generations, but we also observe the exact geographical location of the grandfather, allowing us to gauge the importance of neighborhood effects driving the persistence of ethnic group averages over time (Borjas, 1995).

After constructing the datasets from historical census files, we first analyze the link between father and son to verify that ethnic capital was important for second generation outcomes during the Age of Mass Migration.[5] We show that ethnic groups converge more slowly at the group level than predicted by a standard intergenerational model. This is due to a strong correlation between the son's occupation and ethnic capital in the previous generation – above and beyond the effect of father's occupation. We further show that group averages converge slowly between the *second* generation in 1910 and third generation in 1940.

Given these correlations from the first to second and second to third generations, we can predict the expected convergence across three generations assuming an AR(1) process.[6] Since we have grandfather-grandson linked sets, we can further compare how an AR(1) model performs relative to the actual data. We show that the actual convergence of ethnic group averages is about 30 to 35 percent slower than predicted from iterating an AR(1) model, suggesting that group averages decay at a slower than geometric rate between the first and third generations. This is mostly due to an AR(1) model under predicting the importance of grandfather's occupation rather than under predicting the importance of ethnic capital across three generations.

We then fit the data to an AR(2) model to predict the convergence of skill gaps beyond the third generation. The AR(2) model estimates that 20 percent of a skill gap between two sources in the first generation still exists by the fourth generation; after 7 generations, 5 percent remains; and after 10 generations, only 1 percent remains. Depending on the size of the initial skill gap, it may take over a hundred years for ethnic group averages to converge.

---

[5] León (2005) demonstrates the importance of ethnic capital during the Age of Mass Migration based on the effect of average literacy rate for adults on children still in the household. We differ by exploring the effects on adult outcomes using the linked sample.

[6] An AR(1) process assumes that a generation's outcomes only depends on the immediately preceding generation and no generation before that.

Having established that ethnic group averages converge slowly, we then test for the importance of neighborhoods in explaining this persistence. This line of work follows Borjas (1995) who shows that the effect of ethnic capital on the second generation declines once controlling for neighborhood fixed effects. This result implies that one reason why group averages persist is general neighborhood factors; since the first generation clusters into different neighborhoods, their outcomes are determined by factors separate from ethnicity such as the quality of schools, access to jobs, prevalence of crime or any other neighborhood characteristic. Since we observe the full-count censuses, we can compare grandsons who had grandfathers from the same enumeration district – geographical units that averaged 2,400 individuals in 1880. The results when controlling for enumeration district fixed effects show that ethnic skill gaps converge more quickly for descendants from the same neighborhood. A natural implication of this result is that the slow convergence of ethnic group averages over multiple generations would be quicker if ethnicities were more spatially integrated in the first generation. Alternatively, ethnic groups averages do not converge because all neighborhoods are not of the same quality.

Our paper is related to the fast-growing literature on intergenerational mobility using newly available historical sources.[7] While the earlier literature regresses a son's income on a father's income for the whole population with a single intercept, improved data sources allow researchers to explore heterogeneity in intergenerational links during economic shocks or for different subgroups based on sex or race.[8] A separate but related literature pushes beyond the classic father-son regression to three or more generations.[9] Our research combines both strands since we examine how multi-generational mobility varies across different subgroups (i.e., ethnicities). We contribute to this literature on many fronts. First, our analysis shows that third-generation outcomes are correlated with the average skill level in the first generation, above and beyond the direct effect of grandfather's occupation; this correlation suggests that

---

[7] There are several studies on intergenerational mobility using high-quality data from Scandinavian countries. See Black and Devereaux (2011) for a review of intergenerational studies and Solon (2015) for multigenerational studies.

[8] For research on intergenerational links during economic shocks, see Ager, Boustan and Eriksson (2016) for evidence from loss of slave wealth after the Civil War, Bleakley and Ferrie (2016) for the intergenerational effects of a lottery win, and Feigenbaum (2015) for the Great Depression. For how intergenerational links vary across decades, see Long and Ferrie (2013) and Chetty et al. (2017). For historical studies on intergenerational mobility for different races, see Collins and Wanamaker (2017) for black-white differences, and Hilger (2016) for Asian immigrants. Finally, to see how intergenerational mobility varies across males and females, see Olivetti and Paserman (2015). Not all of these papers use linked father-son data.

[9] See, for example, Solon (2015); Long and Ferrie (2015); Ferrie, Rothbaum and Massey (2016); Lindahl et al. (2015).

multigenerational models should contain ethnic-group specific intercepts. One cannot infer convergence for the entire American population – even for white Americans – from a grandfather-grandson intergenerational elasticity coefficient with a single intercept, as is often done. Moreover, while many others have shown that an iterated AR(1) model under predicts the transmission of status between grandfather and grandson, we additionally show that an iterated AR(1) *ethnic capital* model under predicts persistence of ethnic group averages for three generations (Braun and Stuhler, 2016; Lindahl et al., 2015; Stuhler, 2012).

Our 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 et al., 2001).[10] The third generation are defined as US-born to US-born parents, and whose origin is commonly identified by self-reported ancestry in the Census. It is well known that responses to the ancestry question may be selectively reported and therefore biases our understanding of skill convergence across three immigrant generations (Duncan and Trejo, 2011; Duncan and Trejo, 2015). We improve on these immigrant convergence studies since we are the first to have grandfather-father-son linked data and do not have to rely on the self-reported ancestry question.[11] Despite having better data, the evidence supports the argument of Borjas (1992, 1994, 1995) that ethnic skill differentials are slow to disappear. In fact, our study suggests that convergence may be even slower than Borjas (1994) estimates with data using self-reported ancestry, although our results are not directly comparable since we study different time periods.[12]

Finally, our results also relate to studies on the importance of neighborhood effects for subsequent generations – research that increasingly reveals that place matters. In a series of papers, Raj Chetty, Nathanial Hendren and various co-authors show that intergenerational outcomes vary widely across location and are strongly dependent on the childhood

---

[10] The main strand of the immigration literature explores mobility between two generations, but only a few studies have data that links father to son (Borjas, 1992; León, 2005). See Card et al. (2000) for another study on mobility across two generations, which uses group averages; also, Smith (2003) and Smith (2006). See Fulford et al. (2015) for a separate but related argument that immigrant ancestry influences county-level economic outcomes decades later.

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

[12] Borjas (1994) uses the 1910, 1940 and 1980 Censuses. We use the 1880, 1910 and 1940 Censuses.

neighborhood.[13] Our paper shows that ethnic skill gaps disappear quickly for those from the same *grandfather*'s neighborhood, suggesting that these local neighborhood effects matter for generations beyond the second. While it is unclear what is driving these neighborhood effects, our results imply that ethnic skill differentials persist because first-generation immigrants located in different quality neighborhoods.

II.   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 1914, 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 first enacted after World War I (Abramitzky and Boustan, 2016).[14] 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 exact same as the top senders between 1850 and 1880.[15] 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 with natives, but outcomes still varied widely across sources (Abramitzky et al., 2014; Borjas, 1994; Minns, 2000).[16] On average, immigrants remained stuck at the same point in the occupational distribution as they were at arrival; few sources converged to native outcomes, perhaps because United States-specific human capital such as English fluency had a low return (Ward, 2016). Ethnic group averages persisted not

---

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

[14] 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. Note that comparisons of immigration rates over time do not account for undocumented entries. The migrant stock as a proportion of the population may be a better statistic; this suggests that immigration today is nearing the heights of the 1850-1913 period.

[15] The top ancestries are as follows: German (16%), Scots Irish/Irish (11%), English (8%), African American (8%), the United States (7%), Mexican (7%) and Italian (6%). Note that ancestry is self-reported and therefore some descendants may not correctly self-identify either by accident or on purpose (Duncan and Trejo, 2016); further, respondents may list up to two ancestries.

[16] 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).

only from arrival to decades afterward, but also they persisted to the second generation, leaving little between-group convergence (Abramitzky et al, 2014; Darity et al, 2001).[17]

While skill gaps 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 ours is that of Borjas (1994), who estimates group-level persistence between the first generation in 1910, identified by country of birth, and the third generation in 1980, identified by self-reported ancestry. Borjas finds that ethnic group averages persist from the first generation to the third generation at 0.20, or that four-fifths of ethnic skill differentials disappear by the third generation.[18] While this may imply that first-generation ethnic capital affects the third generation, a 0.20 result may just be the correlation between grandparent and grandson rather than a slower convergence of group averages. For example, Borjas also shows that the correlation with first-generation ethnic capital disappears when controlling for father's occupation, making it unclear whether ethnic group means converge at a slower rate than a standard multigenerational model predicts. Our study improves on Borjas (1994) by having a larger and more comprehensive dataset: we can identify grandson's origin with grandparent's country of birth rather than self-reported ancestry; moreover, we also have linked grandfather-father-son outcomes, so we can account for grandfather's skill level in addition to father's skill level.

III.   Measuring Persistence at the Individual and Group Level

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).[19] A common specification is:

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

---

[17] Abramitzky et al. (2014) do not directly estimate the convergence of ethnic group averages across generations, or Equation (2) from the next section. However, if one uses the values in Figure 6 from their paper, then the estimated slope is approximately 0.80, which is much higher than Borjas (1994) or Card et al. (2000) estimate at around 0.40-0.62. One difference between the results of Abramitzky et al. (2014) and Borjas (1994, 2001) is that ABE use average skill level of *permanent* migrants, while Borjas uses the average skill level of the *temporary and permanent* migrants. These are different because return migrants were negatively selected during the Age of Mass Migration (Ward, 2017). This could be rectified if one only kept first-generation immigrants who stayed more than ten years, since return rates drop drastically after ten years of stay (Greenwood and Ward, 2015). However, the 1880 census does not contain year of arrival.

[18] Alba et al. (2001) argue that intergenerational convergence nearly complete when limiting the sample to European sources. See Borjas (2001) for a rebuttal.

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

where $y_{i,g}$ is the skill level, whether in terms of log wages, log occupation score or rank, 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 (Mazumder and Acosta, 2015).[20]

One does not always have data on fathers and sons, so instead of estimating persistence using families, others estimate mobility at a group level. In the immigration literature, a common interest is how skill differentials in the first generation, grouped by country of birth, predict skill differentials in the second generation, grouped by parental country of birth (Borjas, 1993; Card et al, 2000). A regression that estimates persistence across immigrant generations is closely related to equation (1), but instead of using individual-level data, it uses mean skill level by source country c:[21]

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

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 a first-stage regression.[22] Then, using imputed father's earnings by group membership, one can estimate the intergenerational elasticity in a second-stage regression (Solon, 2015).

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. For example, second-generation children may be influenced by other factors related to their father's country of birth, whether due to human capital spill overs or ethnic discrimination. In a series of paper, Borjas (1992, 1995) addresses this point by including the group average directly in the individual-level intergenerational elasticity equation:

$$y_{i,c,g} = \beta_0 + \beta_1 y_{i,c,g-1} + \beta_2 \bar{y}_{c,g-1} + \varepsilon_{i,g} \tag{3}$$

---

[20] Others have used a rank-rank specification instead of a log-log specification (e.g. Chetty et al., 2014), which estimates the persistence of location within the income distribution.

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

[22] 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).

In this regression, Borjas refers to the average skill level of the prior generation ($\bar{y}_{c,g-1}$) as "ethnic capital", theorizing that human capital spill overs affects the outcomes of future generations. While the exact mechanism is unclear, the specification models the intercept for each country of birth as a function of the prior generation's human capital; thus, countries with lower levels of human capital in the first generation should have lower than average skills in the second generation. This approach is similar in spirit to others who have estimated intergenerational models for African Americans and whites, demonstrating the need for race-specific intercepts (e.g. Collins and Wanamaker, 2017; Hertz, 2005).

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

$$\bar{y}_{c,g} = \beta_0 + (\beta_1 + \beta_2)\bar{y}_{c,g-1} + \varepsilon_{c,g} \qquad (4)$$

Note that this equation is the same as equation (2), where $\theta_1 = \beta_1 + \beta_2$. If ethnic capital has an effect above and beyond the father's skill level, then the persistence of group averages may be stronger than persistence between the father and son from a standard intergenerational model ($\beta_1 + \beta_2 > \alpha_1$). Borjas (1995) terms this sum ($\beta_1 + \beta_2$) the "mean convergence" of ethnic group averages.

It is straightforward to iterate this model to predict outcomes for the third generation. Suppose the relationship between first and second generation of immigrants is as follows:

$$y_{i,c,g-1} = \delta_0 + \delta_1 y_{i,c,g-2} + \delta_2 \bar{y}_{c,g-2} + \varepsilon_{i,c,g-1} \qquad (5)$$

and let Equation (3) be the relationship between the second and third generations.[23] After plugging Equations (4) and (5) into Equation (3) and grouping terms, the relationship between the grandson's skill level, the grandfather's skill level and ethnic capital in the grandfather's generation is

$$y_{i,c,g} = c + \beta_1\delta_1 y_{i,c,g-2} + (\beta_1\delta_2 + \beta_2\delta_1 + \beta_2\delta_2)\bar{y}_{c,g-2} + \upsilon_{i,c,g-2} \qquad (6)$$

where $\beta_1\delta_1$ is the correlation with grandfather's skill and $\beta_1\delta_2 + \beta_2\delta_1 + \beta_2\delta_2$ is the correlation with ethnic capital in the grandfather's generation.[24] Given the multiple avenues through which

---

[23] Equations (3) and (5) show that the effect of father's occupation and ethnic capital may differ across generations ($\beta_2 \neq \delta_2$), which may occur if attachment to ethnicity fades or if the second generation are less residentially segregated.

[24] After grouping terms, the constant $c = \beta_0 + \beta_1\delta_0 + \beta_2\delta_0$. The effect of grandfather's skill on grandchild's skill is a familiar term from the multigenerational literature since it is simply the product of the correlation between the grandfather and father (δ_1), and then between father and son (β_1). The effect of ethnic capital in the

grandfather's ethnic capital influences the grandson's occupation, the effect of ethnic capital may dominate with more generations.[25] Note that the mean convergence between the first and third generations is the sum of the grandfather and ethnic capital effect ($\beta_1\delta_1 + \beta_1\delta_2 + \beta_2\delta_1 + \beta_2\delta_2$), which is also the product of mean convergence between the first and second generations ($\beta_1 + \beta_2$) and second and third generations ($\delta_1 + \delta_2$) – a reflection of the AR(1) set up.

This modelling assumes an AR(1) process where the son's occupation is only influenced by the prior generation. However, there is growing evidence that a son's skill level is also correlated with the skill level of the grandfather, above and beyond the effect of the father (Mare, 2011; Solon, 2015). This could be due to a variety of theoretical reasons, such as investment from the grandparent into the grandchild or a latent factor that is inherited across generations (Clark, 2014; Solon, 2014). One could make a similar argument for first-generation ethnic capital influencing the grandchild's generation, such as the grandparent's generation providing job connections, financial resources, or serving as role models for the grandchild. Therefore, one could extend the ethnic capital model to an AR(2) process:

$$y_{i,c,g} = \gamma_0 + \gamma_1 y_{i,c,g-1} + \gamma_2 y_{i,c,g-2} + \gamma_3 \bar{y}_{c,g-1} + \gamma_4 \bar{y}_{c,g-2} + \varepsilon_{icg} \tag{7}$$

Of particular interest is if the grandfather's occupation or first-generation ethnic capital predicts the grandson's occupation, after controlling for the skill of the father and father's generation. The expected group average for the third-generation Americans from source c, abstracting from the constant, would be $(\gamma_1 + \gamma_3)\bar{y}_{c,g-1} + (\gamma_2 + \gamma_4)\bar{y}_{c,g-2}$. Depending on the coefficients, intergenerational persistence in group averages could converge at a faster or slower rate than that predicted by the AR(1) model.

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

The goal of this paper is to estimate the persistence of skill level from the first generation of immigrants to their grandchildren. The data requirements for these estimates are

---

grandfather's generation comes from a variety of avenues: first, the product of the ethnic capital effect across generations ($\beta_2\delta_2$), the indirect effect of ethnic capital in the grandfather's generation and the father's skill level ($\beta_1\delta_2$), and the indirect effect of ethnic capital in the father's generation and the grandfather's skill level ($\beta_2\delta_1$).
[25] A common assumption is that the effect of father's skill level on son's skill level is stationary across generations or that $\beta_1 = \delta_1$. We could further assume that the effect of ethnic capital is also constant across generations, or that $\beta_2 = \delta_2$. Therefore equation (6) may simplify such that the correlation with grandfather's occupation is β_1^2, and the correlation with grandfather's ethnic capital is $3\beta_2^2$. Therefore, depending the relative importance of ethnic capital and father's occupation on the next generation, the influence of ethnic capital may dominate. Indeed, Borjas (1992) documents with log wages in the NLSY that $\beta_1 = \beta_2$; if this model is correct, then the influence of ethnic capital should be three times the size of grandfather's occupation.

high since data that links grandfather to grandson is uncommon; moreover, we need enough observations for each source to provide a reliable estimate of a group average. We take advantage of the fact that United States 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., 2015). We link the full-count 1880, 1910 and 1940 US censuses to create a new sample of grandfathers linked to fathers and grandsons.[26]

Before describing the linking process in more detail, we first discuss our approach to building the sample (see Figure 2). We start by locating all 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.[27] This leaves us with two datasets of children linked to adult outcomes, who we identify as either second generation (G2) from 1880-1910 or third generation (G3) from 1910-1940. We then use 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 our results only representative of the population with observable fathers (Xie and Killewald, 2013). At this point we 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.

To be included in the regression sample of grandfathers(G1)-fathers(G2)-grandsons(G3), we impose additional restrictions. First, we keep G1 grandfathers who are between 30 and 55 to reduce measurement error from life-cycle bias (Grawe, 2006).[28] We also limit our 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

---

[26] The 1910 and 1940 are preliminary versions of the full-count samples. We clean the 1910 full-count census by matching on strings for occupation and country of birth using already cleaned strings from IPUMS. See Appendix C for more detail.

[27] Our starting sample includes males less than fourteen since those older than fourteen and still in the household might not be representative of the entire population.

[28] Remember that the G2 and G3 generations are 30-44 years old because they had to be under 14 in the census 30 years prior.

third-generation immigrant.[29] We also only keep G1-G3 sets where an occupation is observed in all three generations.

Our method of linking to build this dataset mostly follows others who use automated linking techniques (e.g. Abramitzky, 2012; 2014; Massey, 2017); that is, we link individuals with a computer algorithm that finds the best match based on first name, last name, race, year of birth and state of birth/country of birth.[30] For the 1880 to 1910 link, we additionally match on father and mother's country/state of birth; unfortunately, these variables are unavailable for linking between 1910 and 1940.[31] Please see Appendix A for more detail on the exact process.

After linking and imposing the various sample restrictions, we end with 56,011 G1 grandfathers in 1880 linked to 60,928 G2 fathers in 1910 and 83,548 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 we are unable to perfectly match people across censuses. For example, our linking strategy links 12.2 percent of under 14-year olds in 1880 to 1910 and 35.5 percent from 1910 to 1940.[32] 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; we will explore how linking error effects our results in a later section, but all of our results are robust to a more precisely linked dataset (Bailey, Henderson and Massey, 2016).

To determine the representativeness of the sample, we compare the G1 grandfathers to other similarly aged fathers in 1880, and the G2 fathers to other similarly aged fathers in 1910 in Appendix B. The main biases in the linked sample is that we are more likely to have Germans, farmers and those in the Midwest in our sample; as a trade-off, we are less likely to

---

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

[30] Note that we link only US-born individuals. Therefore, concerns about false links due to individuals changing their first name between censuses, as first-generation immigrants often do, do not apply in this case (Biavaschi, Guilietti and Siddique, forthcoming). Also, we will explore robustness to false links by restricting the links to be more exact; however, note that false links (and thus measurement error) would lead to lower rates of intergenerational persistence. Therefore, if we had a perfectly matched dataset, then my conclusions would likely be stronger; see Section "Bias in measurement from linking and occupation".

[31] Mother and father's birth place is available for sample-line observations in the 1940 Census, but we cannot link on this variable because it is unavailable for the entire census.

[32] The lower linking rate that appears in 1880 is likely due to poorer quality images for transcribing and the fact that we also match on mother and father's birth place. Linking rates are typically lower the further one goes back in time; for example, Ferrie (1996) links the 1850-1860 censuses at an overall rate of 9 percent. This may also be related to higher death rates in between censuses.

have Irish, those with unskilled fathers, and those from the Northeast. These biases are partially because Northeastern states have larger populations, and thus we are less likely to find a unique match. We reweight the sample by source country, skill level and region of residence to ensure that our linked sample is representative of the population on observables, although this reweighting does not change qualitative results (see Appendix B). We use the weighted sample for the rest of the paper.

Imputing Incomes by Occupation

We estimate multigenerational persistence based on occupation rather than income or education since these variables are not available in earlier censuses. We show results based on broad occupational categories (i.e. white-collar, farmer, unskilled and semi-skilled) and based on occupational scores, where we impute the income for each occupation. For coding of white-collar, farmer, unskilled and semi-skilled categories, we classify the 3-digit *occ1950* codes from IPUMS into categories based on the first digit.[33] Broadly, white-collar workers are professional, technical, managerial or sales workers; farmers are only farmers (owners and tenants), but not farm laborers; unskilled are farm laborers, laborers, low-skilled service workers, or operatives; and semi-skilled are craftsmen.

We prefer results using occupational categories since assigning income to each occupation across the entire 1880-1940 period is not straightforward. Nevertheless, we also present results based on two different occupational scores commonly used in the literature. First, we assign each occupation in 1880 and 1910 an income using the 1901 Cost of Living Survey; this is the score used by Borjas (1994). In 1940 we assign the more commonly used *occscore* variable since it is based on a closer time period (1950 earnings). In text we refer to this as the 1901 score since we also present results when using *occscore* throughout the entire 1880-1940 period, as done by Abramitzky et al. (2014) and Olivetti and Paserman (2015). These scores are clearly limiting because they come from different time periods and have no variation within occupation; the reader should keep in mind that we estimate convergence of occupational distributions rather than convergence of income distributions.

A further complexity when imputing income arises from farmers: while the 1950 score does give an estimate of farmer's income, the 1901 Cost of Living Survey does not. We use the farmer income estimates by state created by Ager, Boustan and Eriksson (2016) from the

---

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

revenue and cost data in the 1880 Census of Agriculture.[34] Note that other immigrant convergence studies do not estimate farmer income and thus drop farmers (Alba et al., 2001; Borjas, 1994).[35] While farmer income is certainly estimated with some error, the alternative of dropping farmers would lose the most important occupation: farmers are 47 percent of the grandfathers in 1880, 37 percent of fathers in 1910, and 14 percent of grandsons in 1940.[36] We elect to keep this valuable information; moreover, it may be why we estimate higher persistence rates than Borjas (1994) since farming is highly correlated across generations. Nevertheless, our results will show stronger persistence at the group level than the individual level for both occupational categories and occupational score.

Descriptive Statistics

The average skill level (summarized by the 1950 log occupational score) and the number of observations by country of birth are listed in Table 1. Over half of our 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 81 percent of grandsons in our dataset. Following these countries, the next largest sources are Canada, Norway and Scotland. Note that ethnic spillovers for Canadian immigrants may not be clear because they have descended from England, France or Ireland; yet our results are robust to including or excluding them from the dataset. Note that Italy, Russia and Hungary had smaller stocks in the 1880 Census.[37]

Table 1 provides the first evidence that group averages persisted from the first to the third generation. Sources that had lower than average scores in the first generation, such as

---

[34] The farmer income estimates use data from the 1880 Census of Agriculture on revenue from output and costs from labor, taxes, fertilizer and maintenance. Assumptions for calculating earnings (subtracting costs from revenue) are given in Online Appendix Table 3 in Abramitzky, Boustan and Eriksson (2012). Thus, it proxies farmer income for each state. Also, we use the farm laborer estimates by state as given by Ager, Boustan and Eriksson (2016) and originally found in the Young report (1871). When imputing incomes, we first convert all incomes (farmer and non-farmer) to 2015 dollars since they are measured in different years. This is done using the inflation estimates from www.measuringworth.com Note that results do not change if we assign scores based on the 1880 Census of Agriculture or the 1910 Census of Agriculture.

[35] Moreover, we are unable to separate farm tenants from farm owners in the 1880 census. Collins and Wanamaker (2017) link their sample of black and white famers to the 1880 Census of Agriculture to record tenancy status, but we do not pursue this strategy give our large dataset and high cost of linking. Instead, we rely on the farmer income estimates by state.

[36] The sample's fraction of farmers is larger than in the generational population in 1880, 1910 and 1940. This is because we are more likely to link sons of farmers than non-farmers, partially because farmers have more sons. We reweight our samples to match the population distribution as described in text.

[37] We group Russia and Poland together because Poland was not an independent country until after World War I. We further group most countries in the Austro-Hungarian Empire together. See Appendix E for further details. This coding does not drive results since they are a smaller part of our dataset.

Norway and Sweden, also had lower than average scores in the third generation; similarly, sources with higher than average scores continued to hold occupations that paid more. Of course, persistence of group averages could be because high-skilled fathers tend to also have high-skilled sons and not due to any extra effect from ethnic capital slowing convergence. This underscores the importance of estimating intergenerational mobility with linked grandfather-grandson data.

Immigrants' outcomes are necessarily influenced by where they located in the United States. We plot where immigrants from the two most important sources, Ireland and Germany, lived in 1880 and where their descendants lived in 1940. This is done with the full-count 1880 census and the sample of the 1940 censuses where we observe grandfather's country of birth. There are a few important points from this map, which plots the relative fraction of Irish or Germans by county.[38] First, immigrants were not evenly distributed throughout the country but located in specific areas – for example, the well-known enclaves of Irish in the Northeast and Germans in the Midwest and the Mid-Atlantic. Note that a large portion of the West had immigrants; however, most settled in the more populated areas of the Northeast and Midwest. Importantly, few immigrants located in Southern states, except for Germans in parts of Texas.[39]

The most important point from Figure 1 is the strong persistence of location from the first generation in 1880 to the third generation in 1940. The same areas that had a larger fraction of first-generation Germans or Irish in 1880 also had a larger fraction of third-generation Germans or Irish in 1940. While this is a straightforward result of children living in the same location as their parents – and it is also an important verification that the linked data is of good quality – the location data also imply that *spatial* assimilation, or the even spread of ethnicities across the country, did not take place. 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). In a later section, we will explore the importance of first-generation location for explaining persistence of skill gaps to the third generation.

Given that immigrants are spread out in different places within source, we take a different approach for measuring ethnic capital than the studies of Borjas (1992, 1994, 1995).

---

[38] The denominator of this fraction is the number from the county in the sample, not the entire county's population.
[39] Germans would often travel directly to New Orleans and either take the train or boat to Texas or Midwestern states, because it was much cheaper than traveling via New York (Cohn, 2009). Others entered Texas via Galveston after the failed German revolutions in 1848.

Borjas (1992) argues that the theoretical mechanism that drives persistence of group averages is spill overs within ethnicity during human capital accumulation. However, Borjas can only measure ethnic capital at the national level; since we have much better data, we can measure ethnic capital more precisely at the county or neighborhood level during childhood, where and when these spillovers are more likely to occur. Both the 1880 and 1910 full-count censuses allow us to measure human capital for enumeration districts, which contain on average 2,400 individuals. Therefore, we use average skill level of co-ethnics in the enumeration district as our measure of ethnic capital.[40] Note that ethnic capital is measured for both temporary and permanent migrants; unfortunately, we cannot separate those who return home from those who remain in the 1880 Census.

V.   The Persistence of Skill across Two Generations

In this section, we demonstrate how skill persists across two generations, verifying that ethnic group averages converge more slowly than the link between father and son. That is, we compare how a standard intergenerational model from Equation (1) predicts the convergence of ethnic skill differences relative to the model with ethnic capital in Equation (5). Remember that sons are between ages 30 and 44 and we restrict fathers to ages 30 to 55, limiting bias from variation in earnings across the lifecycle (Haider and Solon, 2006; Grawe 2006; Nybom and Stuhler, 2016). Yet given measurement issues from life-cycle bias, we present results including age controls in the regression: specifically, a quartic of son's age, father's age, and son's age interacted with father's occupation, where age is normalized to 40 – this follows the specification of Lee and Solon (2009). For all results presented in this paper, we always include these age controls.

Table 2 shows that in a standard intergenerational model there is a tight correlation between the father and son's occupational categories. For instance, having a farmer father in 1880 correlates with a 0.50 percentage point increase in the likelihood of the son being a farmer. Furthermore, a son is 0.31 percentage points more likely to be white-collar worker if his father was also a white-collar worker. The correlations for the other categories of unskilled (0.20) and skilled (0.16) worker are smaller, but still positive. If we assume that these results predicted

---

[40] Ethnic capital may capture other neighborhood factors, which we will later control for with enumeration district fixed effects. In our view, it does not matter if the correlation between ethnic capital and the next generation is from neighborhood factors or not, as long as the correlation exists. In other words, we are not interested in a causal interpretation of ethnic capital, but rather are interested in measuring the convergence of skill gaps across generations, just as the intergenerational literature is interested in the correlation between father and son and not the causal effect of father on son.

convergence for the whole population, then ethnic skill gaps in farming would converge the slowest, but first-generation differences in skilled and unskilled work would disappear quickly.

The bottom row of Table 1 uses log-log specification for the 1901 and 1950 occupational score, and provides a specification in line with the modern-day studies on intergenerational mobility. There is a 0.41 intergenerational elasticity when using the 1950 occupational score, while the 1901 occupational score estimates an elasticity of 0.29. Both estimates are lower than income persistence estimates from the late 20th century (~0.50), and are consistent with higher mobility in the past than today (Feigenbaum, 2017; Long and Ferrie, 2013). A higher coefficient for the 1950 score is also consistent with the literature; for example, Feigenbaum (2017) shows that the 1950 occupational score overestimates the degree of persistence for his sample of Iowa farmers and sons, likely because the 1950 score places farmers at the 10th percentile of earnings and farming is highly persistent over time. The 1901 score, on the other hand, places farmers near median earnings, and probably better reflects mobility during the earlier time period. Nevertheless, the reader should keep in mind that these occupational scores are limited for measuring income prior to 1940.

These relatively high levels of mobility between father and son do not explain the strong persistence of group averages documented elsewhere in the literature (Abramitzky et al., 2014). A simple interpretation of the 0.29 coefficient from the 1901 score suggests that if we take a skill gap between *any two fathers*, only 29 percent of the gap remains for their sons. In other words, the prediction is that group averages converge at the same rate – yet this was not the case in the early 20th century, suggesting that a standard intergenerational model does not fully capture convergence of ethnic skill gaps.

In the second column, we add measures of ethnic capital to explain why convergence of group averages was slower than the intergenerational elasticity suggests.[41] The coefficient for ethnic capital is always positive and significant for each of the occupational categories, indicating that sons held similar occupations as co-ethnics in the father's generation. The sum of the father and ethnic capital coefficients estimates how group averages converge across the first and second generation. The mean convergence for farmers is 0.62, about 25 percent higher than the prediction from a standard model between father and son (0.50). In other words, a standard intergenerational model predicts that group averages converge more quickly than they

---

[41] Ethnic capital is measured as the proportion of the 1880's ethnicity in the enumeration district who have the same occupational category. When using occupational scores, it is either the average log occupational score, or average rank.

actually do.[42] For white-collar work, the mean convergence rate is 0.58, compared with 0.31 from the father-son regression.

The rate of mean convergence when using occupational scores is 50 to 66 percent higher than the simple father-son model. With the 1901 score, the father-son regression predicts that only 29 percent of initial skill gaps remain. The rate of mean convergence estimates that 48 percent of ethnic skill gaps should remain. Therefore, this regression reconciles the estimates of high mobility in the population, as documented by Feigenbaum (2017) and Long and Ferrie (2013), with the slow convergence of group averages between the first and second generations, as documented by Abramitzky et al. (2014).

Note that all estimated coefficients in Table 2 are clearly correlational and not causal; both father's occupation and ethnic capital may be capturing general geographic effects of being raised in, for example, a Midwestern farming community. Yet the fact that second generation Americans generally lived in the same locations as their parents reinforces that gaps across ethnic groups may converge slowly due to geographical factors.

Persistence from the Second to Third Generations

We have just verified that ethnic averages converge slower than the standard model predicts between the first and second generation; this is because ethnic capital is highly correlated with the next generation's outcomes. Now we test whether the ethnic capital effect mattered between the second and third generations, when attachment to ethnicity may have dwindled. We re-estimate the same correlations as in Table 2, but now with data from the second generation in 1910 to the third generation in 1940 (that is, we estimate equation (3) while also controlling for life-cycle effects). Note that here we measure ethnic capital in 1910 for individuals from the *second* generation because we are interested in how group averages converge from first to second to third generation Americans.

Table 3 again establishes that the father's occupation is correlated with the son's occupation when using either occupational categories or occupational scores: farmer fathers are more likely to have farmer sons, and white-collar fathers are more likely to have white-collar sons. The log-log correlations for occupational scores are slightly weaker for the 1910

---

[42] The slower estimate of mean convergence is because the sum of the father and ethnic coefficient is higher than the father coefficient in the first column; however, when including ethnic capital in the model, the effect of father's occupation drops in magnitude. This suggests that part of the correlation between the father and son's occupation is because others in the area where the son was raised held similar occupations.

to 1940 dataset; now, the correlation is 0.298 when using the 1950-based score and 0.304 when using the 1901-based score. Once again, the data is consistent with the pattern that intergenerational mobility was higher in the early 20th century. However, the specification may predict that ethnic skill differentials converge too quickly if there is a missing effect of ethnic capital.

Ethnic group averages converged between the *second* and *third* generations more slowly than predicted by the population model. Similar to the relationship between the first and second generations, the model estimates that mean convergence in occupational score between the second and third generations is about 50 to 60 percent slower than predicted from a standard intergenerational model. Therefore, factors that are correlated with ethnicity, whether discrimination, locational choice or human capital spill overs, still influence occupations of third-generation Americans in 1940. This result is surprising because it is commonly believed that attachment to ethnicity decreases after the second generation as in the melting pot metaphor (Alba, 1985). It may be that slower convergence of ethnic group averages occurs even though ethnicity is not salient for third-generation Americans; this could be due to residential location, which we will show later.

VI.   The Persistence of Outcomes across Three Generations

We have established that group averages converged more slowly than predicted from a standard intergenerational model for immigrants from the Age of Mass Migration. This is because ethnic capital is correlated across generations, above and beyond the effect of father's occupation. The results imply that there should be an ethnic capital effect between the first and third generation, and a slower rate of convergence of ethnic averages than predicted by a grandfather-grandson multigenerational model.

However, before estimating the model for the grandparents-grandsons, one could predict the coefficients under the assumption that an AR(1) process exists across the two generations. For example with farming, we document that the correlation between fathers and sons after controlling for ethnic capital is 0.294 between 1880 and 1910, and 0.186 between 1910 and 1940. An iterated AR(1) model predicts that correlation between grandfather and grandson is the product of these two estimates, or 0.055. Similarly, predicting the farming coefficient from ethnic capital using the values in Tables 2 and 3 and the formula from Equation (6) yields 0.168. Therefore, we expect that the convergence of ethnic averages, or the sum of the predicted grandfather and ethnic capital coefficients, is 0.223. We can easily test whether

this is true given we have the data linking grandfather and grandson. We display the predicted coefficients from iterating the AR(1) model in Table 4 for all occupational categories and scores.

Table 4 demonstrates that an iterated AR(1) model underestimates the convergence of skill gaps across three generations. For example with farming, the predicted 0.055 grandfather-grandson coefficient is much less than the 0.120 actually estimated from the data. Underestimating the grandfather's influence with an iterated AR(1) model is a common result in the literature (Solon, 2015), though it is unclear exactly why this occurs. For example, it may be due to a causal effect due to investment from the grandfather, or a spurious correlation from measurement error. Yet the iterated AR(1) model does not underestimate the importance of grandfather's ethnic capital, but actually overestimates it; the predicted coefficient of 0.168 is higher than the true one at 0.141. The amount which an AR(1) over or under predicts the ethnic capital coefficient varies by outcome, yet the AR(1) model always underestimates the grandfather coefficient.

Due to this underestimating of grandfather's influence, the AR(1) model also under states the mean convergence across three generations. Given the relationships found in Tables 2 and 3, ethnic skill gaps should have persisted at a 0.236 rate between the first and third generations (1901 score). Yet we find that the true mean convergence is 0.310, about 30 percent slower than predicted. This suggests that ethnic group averages closed at a slower than geometric rate, at least for our dataset between 1880 and 1940.

The slow convergence in skill gaps primarily come from white-collar and farming occupations rather than unskilled or skilled work. Convergence of group averages in skilled and unskilled work occurred relatively rapidly, where 91-92 percent of initial differences in these categories closed by the third generation. It may be that persistence is stronger at the upper ends of the occupational distribution, such as for white-collar work, like persistence today for higher income levels. Farming may also be a unique occupation that is highly persistent over time given the land requirements and inheritance. Moreover, given that most descendants remained in the same locations as their grandparents, sources with many first-generation farmers, such as Norway and Sweden, remained in farming for generations.

In Appendix C, we show that the result that ethnic group averages close slower than predicted by a standard model still holds when dropping non-European sources such as Canada

or Mexico. Therefore, ethnic skill gaps converge slowly when limiting the sample to white Europeans.

A Second-Order Autoregressive Model of Ethnic Capital

How many generations will it take for the full convergence of ethnic group averages? The previous section shows that an AR(1) model predicts that group averages converge too quickly; an AR(2) model as shown in Equation (7) may more accurately predict the eventual closing of skill gaps. Therefore, in this section we regress a G3 grandson's occupation on the G2 father's and G1 grandfather's occupation, and on G2 and G1 ethnic capital.

The results of an AR(2) ethnic capital model are shown in Table 5. As expected, the values on the grandfather's occupation and grandfather's ethnic capital are positive, which demonstrates that ethnic skill gaps close at a slower than geometric rate. For the 1901 occupational score, the correlation with the group average one generation prior is 0.439 and two generation's prior is 0.112. The AR(2) coefficients sum to less than one, indicating that ethnic group averages will eventually converge rather than remain stable (sum to 1) or diverge (sum to more than 1).

We use these estimated values to predict how many generations it takes for ethnic skill gaps to converge. We show the results in Figure 3 for the 1901 log occupational score. The AR(2) model predicts that 20 percent of the occupational score gap in the first generation still exists after 4 generations. For example with Germans and Norwegians, who started with an occupational skill gap of 30 percent, the model predicts that 4 generations later the German Americans and Norwegian Americans have a 6 percent occupational score gap. Note that there are approximately 5 to 6 generations separating us today from fathers in 1880; thus, the model predicts that small differences still exist between Norwegian and German descendants today. After 10 generations, or the benchmark used by Clark (2014), then 99 percent of any initial difference between two ethnic groups disappears.

Of course, extrapolating an AR(2) model into the future is problematic. We do not model how assortative mating might slow or quicken the pace of convergence (Meng and Gregory, 2005); nor do we know how shocks to the economy affect intergenerational mobility in the future.[43] For example, between 1880 and 1940 the economy shifted away from

---

[43] We define ethnicity through the paternal line; this misses other grandparents' country of birth. See Logan and Shin (2012) for a study on historical marital rates across generations; also see Wildsmith, Gutmann and Gratton (2003).

agriculture to manufacturing; given we document a strong persistence in farming despite this structural shift, it could be that skill gaps would have existed for longer in the counterfactual scenario without technological change. Moreover, an AR(3) model of ethnic skill gaps may demonstrate rapid convergence in the fourth generation as attachment to ethnicity falls even further. Therefore, one should be careful when extrapolating these results beyond the third generation and to different time periods. Nevertheless, our data shows that, at least for Europeans, gaps by source country remain important for at least 3 generations.

Bias in measurement from linking and occupation

Our main result is that convergence of ethnic group averages across 3 generations is slower than a standard multigenerational model predicts. A large driver of slower convergence is the positive effect of ethnic capital; importantly, we consistently find that ethnic capital is more important for the next generation than the grandfather or father. An important caveat to this result is that it is also consistent with measurement error from the linking process. Given that we link on a few pieces of information, it is likely that some of our data is matched to the wrong individual (Bailey, Henderson, and Massey, 2016; Massey, 2017). Errors in the linking process would lower both coefficients of grandfather's skill and ethnic capital; yet measurement error for grandfather's occupation is likely more severe than for ethnic capital since we are more likely to pick the wrong grandfather rather than the wrong grandfather's country.[44]

In Table 6, we test for the importance of this measurement error by limiting our sample to those who are linked exactly on available characteristics, and are thus more likely to be true matches (Massey, 2017).[45] Restricting our sample to exact matches for both the 1880-1910 link and 1910-1940 link reduces the number of observations from 82,000 to about 17,000 grandsons, demonstrating that many of our matches are based on a fuzzy match either between 1880-1910 or 1910-1940. Using the exactly linked sample and the 1901 occupational score, we find the expected result that the grandfather effect nearly doubles from 0.062 to 0.117; yet at the same time, the ethnic capital effect remains stable. Therefore, the sum of the grandfather and ethnic capital effect is larger in the exactly linked dataset. Importantly, this suggests that

---

[44] The linking algorithm picks the best option that matches on first name, last name and state of birth; if the algorithm chooses the wrong match, the (false) link still has a similar last name and the same state of birth as the true link. Given that we identify ethnicity through the paternal line, surnames are a reasonable proxy for ancestry, suggesting fewer mismatches for grandparent's country of birth. The resulting measurement error could drive our results that grandfather's country matters more than the grandfather.

[45] That is, they have the same first name and last name strings, year of birth and state of birth; this drops those with slight deviations in names or year of birth.

our main results predict that ethnic skill gaps converge more quickly than they actually do. In other words, our main argument that ethnic skill differentials converge slower than predicted from a standard multigenerational model would likely be strengthened with a perfectly linked dataset.

Another measurement issue is that we only have one observation of grandfather's occupation rather than the full occupational history. It is well known that estimating intergenerational mobility with only a single observation of income or occupation downward biases estimates due to transitory shocks (Mazumder and Acosta, 2015; Solon, 1992). However, this life-cycle bias is smallest at mid-career for father and son, which is one reason why we limit our sample to G1 grandfathers between 30 and 55, G2 fathers between 30 and 44, and G3 sons between 30 and 44. If the father's and grandfather's occupation does not fully capture their skill sets, then ethnic capital may serve as a good proxy for the father's and grandfather's skills. This suggests that we would find a positive effect for ethnic capital even when none exists.[46] Unfortunately, we are unable to gauge the extent of this bias given that we only have one observation of the father or grandfather's occupation.

However, we are not concerned that measuring parent's occupation with error drives the positive coefficient on ethnic capital. We will show in the next section that once controlling for 1880 neighborhood fixed effects, the coefficient on ethnic capital drops while the grandparent coefficient remains stable. This suggests that a positive coefficient on ethnic capital is not due to proxying for grandparent's skills; rather, it suggests that ethnic capital is capturing general neighborhood characteristics in the first generation. Nevertheless, our main result that ethnic skill gaps close at a slower rate than a standard multigenerational model predicts still holds under this measurement error issue.

VII.   The importance of first-generation neighborhood

We have demonstrated that ethnic group averages close at a slow rate from the first to third generation. However, these correlations across generations do not measure any causal process. It is unclear what exactly is driving the relationship between ethnic capital in first generation and occupation in the third generation; while Borjas (1992) argues that it reflects spill overs during the human capital accumulation process, it could also reflect discrimination or other effects that persist throughout time. One potential reason for persistence over time may

---

[46] See Appendix A of León (2005) for a further discussion of this issue.

be immigrants clustering into different types of neighborhoods, and outcomes persist strongly for those from the same neighborhood. Borjas (1995) shows that the correlation between ethnic capital and the next generation's occupation narrows substantially when controlling for neighborhood, suggesting that ethnic capital proxies for general neighborhood effects, such as access to high-quality schools or jobs.

In this section, we test for the importance of ethnic capital across three generations when also controlling for the neighborhood in the first generation.[47] We do this by including 1880 enumeration district fixed effects for the adult G1 grandfather; note that this is the same childhood enumeration district of the G2 father. About 10 percent of the 1880 grandfathers do not have another grandfather from the same enumeration district, so these individuals do not identify our coefficient; however, we are still left over with plenty of variation to estimate a neighborhood fixed effects specification.

In Table 7, we show the correlation between the first and third generations before and after controlling for 1880 enumeration district fixed effects. For each outcome the coefficient on ethnic capital falls substantially in the fixed effects specification, ranging from falling 96 percent for farming to falling 10 percent for unskilled work. The stronger fall for farming is unsurprising because the ethnic capital variable proxies with being raised in a rural farming community. Given one was raised in a farming community, having more or less co-ethnic farmers in the neighborhood did not influence the grandchild's occupation – however, there was still a positive correlation between the grandfather being a farmer and the grandson being a farmer. Therefore, even within neighborhood, farming ethnic skill differentials did not converge between the first and third generation – however, they converged at the same rate predicted by a multigenerational regression with neighborhood fixed effects.

While ethnic capital did not matter for farming after netting out neighborhood effects, it did matter for white-collar work, unskilled work, and therefore occupational scores. Therefore, for these outcomes there was slower mean convergence of ethnic group averages than predicted by the relationship between grandfather and son. However, the rate of mean convergence is estimated to be much quicker with neighborhood fixed effects. When using the 1901 occupational score, the sum is 0.117, suggesting that 88 percent of any skill gap between two ethnicities in the same 1880 neighborhood disappeared by the third generation. This

---

[47] See Appendix D for the importance of neighborhood fixed effects in two-generation regressions. We do not present these results in text due to space constraints.

24

compares with a sum of 0.31 without fixed effects, indicating that 69 percent of occupational score gaps disappeared across three generations for the entire country.

These results suggest that the main channel for the ethnic capital effect is not due to human spillovers from one generation to the next. Rather, there are local neighborhood factors that cause occupational gaps to persist across multiple generations. This may be because there were different quality schooling or health environments in different 1880 neighborhoods, and that these environments had long-lasting influences on human capital accumulation and occupational choice in the future. If neighborhood factors were highly influential, then persistence of ethnic group averages across generations may be expected since immigrant clustering was relatively strong in the 1880s (Logan and Zhang, 2012). Yet there was still a positive effect of ethnic capital within the neighborhood, suggesting that ethnic skill differentials still persisted after controlling for the neighborhood; however, the persistence was much less strong. Ultimately, it is unclear what causal mechanisms drive the general neighborhood effects or the ethnic capital effect within neighborhood, but place appears to be important not only for today but also in the past (Chetty et al., 2014).

VIII. Conclusions

In this paper, we measure the convergence of ethnic skill means 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, we show that the average skill level in the first generation drags the convergence of skill gaps across three generations. Therefore, a high-skilled immigrant group is more likely to have a high-skilled second- and third-generation group; similarly, a low-skilled immigrant group is more likely to have a low-skilled second and third generation group. The results imply that an intergenerational elasticity model with a single intercept does not predict economic convergence for the entire population. While others have noted this when estimating the convergence of the black-white income gap, we show even within the group of white Americans, ethnic-specific factors slow the convergence of skill differentials.

We show that convergence models that only depend on the prior generation (i.e. AR(1) models) understate the transmission of ethnic group averages across three generations by about 30 percent. Fitting the data to an AR(2) model predicts that after 4 generations, 20 percent of initial differences in group averages remain; after 10 generations, only 1 percent of initial

differences remain. These results are consistent with an AR(1) coefficient of 0.67 for group averages, much higher than population AR(1) coefficients.

The primary goal of this paper is to measure the convergence of skill gaps across three generations, but we do not uncover the causal mechanisms behind slow rates of convergence. Yet we do provide evidence that neighborhood effects explain a large portion of the slow convergence across generations. Ethnic skill gaps converged rapidly if the grandparents lived in the same neighborhood. The results suggest that the childhood neighborhood environment is an important determinant of outcomes for the third generation, consistent with much research from modern-day data between two generations (Chetty, Hendren and Katz, 2016; Chetty and Hendren, 2016a; Chetty and Hendren, 2016b). A natural corollary to these results is that ethnic skill gaps would close more quickly if the first generation were more spatially integrated or if quality differences across neighborhoods were flattened.

The results in this paper come from a dataset of mostly European groups between 1880 and 1940, and therefore do not fully apply to today's immigrants from Central America, South Asia and East Asia. Yet there are many reasons to expect that persistence of group averages to be stronger for more recent times than for this dataset. For example, immigrant segregation has been increasing since the mid-20th century, and segregation reinforces group averages over time (Borjas, 1995; Cutler et al., 2008). Income inequality has also been increasing, which widens most income gaps. 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 (i.e., segmented assimilation theory) (Portes and Zhou, 1993; Perlmann, 2005).

Yet we have scarce knowledge on how recent immigrants' group averages evolve past the second generation 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 a group's average for the variety of immigrant origins.[48] 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 skill differentials for post Hart-Celler entrants. This issue is increasingly important as more third-

---

[48] For exceptions, see Duncan and Trejo (2011); Duncan and Trejo (2015); and Özek and Figlio (2016).

generation Americans from Latin America and Asia will enter the labor market in the next few decades.

Bibliography

Aaronson, Daniel, and Bhashkar Mazumder. "Intergenerational economic mobility in the United States, 1940 to 2000." *Journal of Human Resources* 43.1 (2008): 139-172.

Abramitzky, Ran, and Leah Boustan. *Immigration in American Economic History*. No. w21882. National Bureau of Economic Research, 2016.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. "Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration." *The American Economic Review* 102.5 (2012): 1832-1856.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." *Journal of Political Economy* 122.3 (2014): 467-506.

Ager, Philipp, Leah Boustan, and Katherine Eriksson. "Inter-generational transmission of wealth shocks: Evidence from the US Civil War." Manuscript (2016).

Alba, Richard D. *Italian Americans: Into the twilight of ethnicity*. Prentice Hall, 1985.

Alba, Richard, Amy Lutz, and Elena Vesselinov. "How enduring were the inequalities among European immigrant groups in the United States?" *Demography* 38.3 (2001): 349-356.

Alba, Richard, and Victor Nee. *Remaking the American mainstream: Assimilation and contemporary immigration*. Harvard University Press, 2009.

Bailey, Martha, Morgan Henderson and Catherine Massey. "How do Automated Methods Linking Perform? Evidence from the LIFE-M Project". (2016)

Biavaschi, Constanza, Corrado Guilietti and Zahra Siddique. "The Economic Payoff of Name Americanization." *Journal of Labor Economics*. (Forthcoming)

Black, Sandra E., and Paul J. Devereux. "Recent Developments in Intergenerational Mobility." *Handbook of Labor Economics* 4 (2011): 1487-1541.

Bleakley, Hoyt, and Joseph Ferrie. "Shocking behavior: Random wealth in antebellum Georgia and human capital across generations." *The Quarterly Journal of Economics* 131.3 (2016): 1455-1495.

Borjas, George J. "Ethnic Capital and Intergenerational Mobility." *The Quarterly Journal of Economics* 107.1 (1992): 123-150.

Borjas, George J. "The Intergenerational Mobility of Immigrants." *Journal of Labor Economics* 11.1 (1993): 113-35.

Borjas, George J. "Long-run convergence of ethnic skill differentials: The children and grandchildren of the great immigration." *Industrial & Labor Relations Review* 47.4 (1994): 553-573.

Borjas, George J. "Ethnicity, Neighborhoods, and Human-Capital Externalities." *American Economic Review* 85.3 (1995): 365-90.

Borjas, George J. "Long-run convergence of ethnic skill differentials, revisited." *Demography* 38.3 (2001): 357-361.

Braun, Sebastian Till, and Jan Stuhler. "The Transmission of Inequality Across Multiple Generations: Testing Recent Theories with Evidence from Germany." *The Economic Journal* (2016).

Card, David, John DiNardo, and Eugena Estes. "The More Things Change: Immigrants and the Children of Immigrants in the 1940s, the 1970s, and the 1990s." *Issues in the Economics of Immigration*. University of Chicago Press, 2000. 227-270.

Chetty, Raj, Nathaniel Hendren, Patrick Kline and Emmanuel Saiz. "Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States." *The Quarterly Journal of Economics* 129.4 (2014): 1553-1623.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. "The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment." *The American Economic Review* 106.4 (2016): 855-902.

Chetty, Raj, and Nathaniel Hendren. *The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects*. No. w23001. National Bureau of Economic Research, 2016a.

Chetty, Raj, and Nathaniel Hendren. *The impacts of neighborhoods on intergenerational mobility ii: County-level estimates*. No. w23002. National Bureau of Economic Research, 2016b.

Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. "The fading American dream: Trends in absolute income mobility since 1940." *Science* 356.6336 (2017): 398-406.

Clark, Gregory. *The son also rises: surnames and the history of social mobility*. Princeton University Press, 2014.

Collins, William J., and Marianne H. Wanamaker. *Up from Slavery? African American Intergenerational Economic Mobility Since 1880*. No. w23395. National Bureau of Economic Research, 2017.

Cohn, Raymond L. *Mass immigration under sail: European immigration to the Antebellum United States*. Cambridge University Press, 2009.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. "Is the melting pot still hot? Explaining the resurgence of immigrant segregation." *The Review of Economics and Statistics* 90.3 (2008): 478-497.

Darity, William, Jason Dietrich, and David K. Guilkey. "Persistent advantage or disadvantage?: Evidence in support of the intergenerational drag hypothesis." *American Journal of Economics and Sociology* 60.2 (2001): 435-470.

Duncan, Brian, and Stephen J. Trejo. "Intermarriage and the intergenerational transmission of ethnic identity and human capital for Mexican Americans." *Journal of Labor Economics* 29.2 (2011): 195.

Duncan, Brian, and Stephen J. Trejo. "Assessing the socioeconomic mobility and integration of US immigrants and their descendants." *The ANNALS of the American Academy of Political and Social Science* 657.1 (2015): 108-135.

Duncan, Brian, and Stephen J. Trejo. *The complexity of immigrant generations: Implications for assessing the socioeconomic integration of Hispanics and Asians*. No. w21982. National Bureau of Economic Research, 2016.

Duncan, Brian, Jeffrey Grogger, Ana Sofia Leon, and Stephen J. Trejo. "The Generational Progress of Mexican Americans." (2017).

Dustmann, Christian, and Albrecht Glitz. "Migration and education." *Handbook of the Economics of Education* 4 (2011): 327-439.

Feigenbaum, James J. "Multiple Measures of Historical Intergenerational Mobility: Iowa 1915 to 1940" *Economic Journal* (2017).

Feigenbaum, James J. "Intergenerational Mobility during the Great Depression." *Unpublished working paper* (2015).

Ferrie, Joseph P. "A new sample of males linked from the public use microdata sample of the 1850 US federal census of population to the 1860 US federal census manuscript schedules." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 29.4 (1996): 141-156.

Ferrie, Joseph P. *Yankeys now: Immigrants in the antebellum US 1840-1860*. Oxford University Press on Demand, 1999.

Ferrie, Joseph, Catherine Massey, and Jonathan Rothbaum. *Do Grandparents and Great-Grandparents Matter? Multigenerational Mobility in the US, 1910-2013*. No. w22635. National Bureau of Economic Research, 2016.

Fulford, Scott L., Ivan Petkov, and Fabio Schiantarelli. "Does it matter where you came from? Ancestry composition and economic performance of US counties, 1850-2010." Working paper, 2015.

Glazer, Nathan, and Daniel P. Moynihan. "Beyond the melting pot: The Negroes, Puerto Ricans, Jews, Italians and Irish of New York City." (1963).

Grawe, Nathan D. "Lifecycle bias in estimates of intergenerational earnings persistence." *Labour economics* 13.5 (2006): 551-570.

Greenwood, Michael J., and Zachary Ward. "Immigration quotas, World War I, and emigrant flows from the United States in the early 20th century." *Explorations in Economic History* 55 (2015): 76-96.

Haider, Steven, and Gary Solon. "Life-cycle variation in the association between current and lifetime earnings." *The American Economic Review* 96.4 (2006): 1308-1320.

Hatton, Timothy J., and Jeffrey G. Williamson. *The age of mass migration: Causes and economic impact*. Oxford University Press on Demand, 1998.

Hertz, T. 2005. "Rags, Riches and Race: The Intergenerational Economic Mobility of Black and White Families in the United States," in *Unequal Chances: Family Background and Economic Success*, S. Bowles, H. Gintis, and M. Osborne, eds., pp. 165-191. New York, NY: Russell Sage Foundation and Princeton University Press.

Hilger, Nathaniel. *Upward Mobility and Discrimination: The Case of Asian Americans*. No. w22748. National Bureau of Economic Research, 2016.

Lee, Chul-In, and Gary Solon. "Trends in intergenerational income mobility." *The Review of Economics and Statistics* 91.4 (2009): 766-772.

León, Alexis. "Does' Ethnic Capital'Matter? Identifying Peer Effects in the Intergenerational Transmission of Ethnic Differentials." (2005).

Lindahl, Mikael, Mårten Palme, Sofia Sandgren Massih, and Anna Sjögren. "Long-Term Intergenerational Persistence of Human Capital An Empirical Analysis of Four Generations." *Journal of Human Resources* 50.1 (2015): 1-33.

Logan, John R., and Hyoung-jin Shin. "Assimilation by the third generation? Marital choices of white ethnics at the dawn of the twentieth century." *Social science research* 41.5 (2012): 1116-1125.

Logan, John R., and Weiwei Zhang. "White ethnic residential segregation in historical perspective: US cities in 1880." *Social science research* 41.5 (2012): 1292-1306.

Long, Jason, and Joseph Ferrie. "Intergenerational occupational mobility in Great Britain and the United States since 1850." *The American Economic Review* 103.4 (2013): 1109-1137.

Long, Jason, and Joseph Ferrie. "Grandfathers Matter(ed): Occupational Mobility Acorss Three Generations in the U.S. and Britain, 1850-1910" *mimeo,* 2015.

Massey, Catherine G. "Playing with matches: An assessment of accuracy in linked historical data." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* (2017): 1-15.

Mazumder, Bhashkar, and Miguel Acosta. "Using occupation to measure intergenerational mobility." *The ANNALS of the American Academy of Political and Social Science* 657.1 (2015): 174-193.

Mitchener, Kris James, and Ian W. McLean. "US regional growth and convergence, 1880–1980." *The Journal of Economic History* 59.04 (1999): 1016-1042.

Minns, Chris. "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 (2000): 326-350.

Nybom, Martin, and Jan Stuhler. "Heterogeneous income profiles and lifecycle bias in intergenerational mobility estimation." *Journal of Human Resources* 51.1 (2016): 239-268.

Olivetti, Claudia, and M. Daniele Paserman. "In the name of the son (and the daughter): Intergenerational mobility in the United States, 1850–1940." *The American Economic Review* 105.8 (2015): 2695-2724.

Özek, Umut, and David N. Figlio. *Cross-Generational Differences in Educational Outcomes in the Second Great Wave of Immigration*. No. w22262. National Bureau of Economic Research, 2016.

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

Portes, Alejandro, and Min Zhou. "The new second generation: Segmented assimilation and its variants." *The annals of the American academy of political and social science* 530.1 (1993): 74-96.

Portes, Alejandro, and Rubén G. Rumbaut. *Immigrant America: a portrait*. Univ of California Press, 2006.

Steven Ruggles, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. *Integrated Public Use Microdata Series: Version 6.0* [dataset]. Minneapolis: University of Minnesota, 2015. http://doi.org/10.18128/D010.V6.0.

Smith, James P. "Assimilation across the Latino generations." *The American Economic Review* 93.2 (2003): 315-319.

Smith, James P. "Immigrants and the labor market." *Journal of Labor Economics* 24.2 (2006): 203-233.

Solon, Gary. "Intergenerational mobility in the labor market." *Handbook of labor economics* 3 (1999): 1761-1800.

Solon, Gary. "Theoretical models of inequality transmission across multiple generations." *Research in Social Stratification and Mobility* 35 (2014): 13-18.

Solon, Gary. *What do we know so far about multigenerational mobility?*. No. w21053. National Bureau of Economic Research, 2015.

Stuhler, Jan. "Mobility Across Multiple Generations: The Iterated Regression Fallacy". IZA Discussion Paper No. 7072, 2012.

Ward, Zachary. *The Role of English Fluency in immigrant Assimilation: Evidence from United States History*. No. 049. Centre for Economic History, Research School of Economics, Australian National University, 2016.

Ward, Zachary. "Birds of passage: Return migration, self-selection and immigration quotas." *Explorations in Economic History* 64 (2017): 37-52.

Wildsmith, Elizabeth, Myron P. Gutmann, and Brian Gratton. "Assimilation and intermarriage for US immigrant groups, 1880–1990." *The History of the Family* 8.4 (2003): 563-584.

Xie, Yu, and Alexandra Killewald. "Intergenerational occupational mobility in Great Britain and the United States since 1850: Comment." *The American economic review* 103.5 (2013): 2003-2020.

Young, Edward. *Special Report on Immigration.* Philadelphia: US Department of Treasury, 1871.

Figure 1. Persistence of location from First Generation in 1880 to Third Generation in 1940

German 1st Generation 1880

German 3rd Generation 1940

Fraction of County pop (Male 30-44)
No Data
0.000 - 0.025
0.025 - 0.050
0.050 - 0.100
0.100 - 0.250
0.250 - 1.000

Irish 1st Generation 1880

Irish 3rd Generation 1940
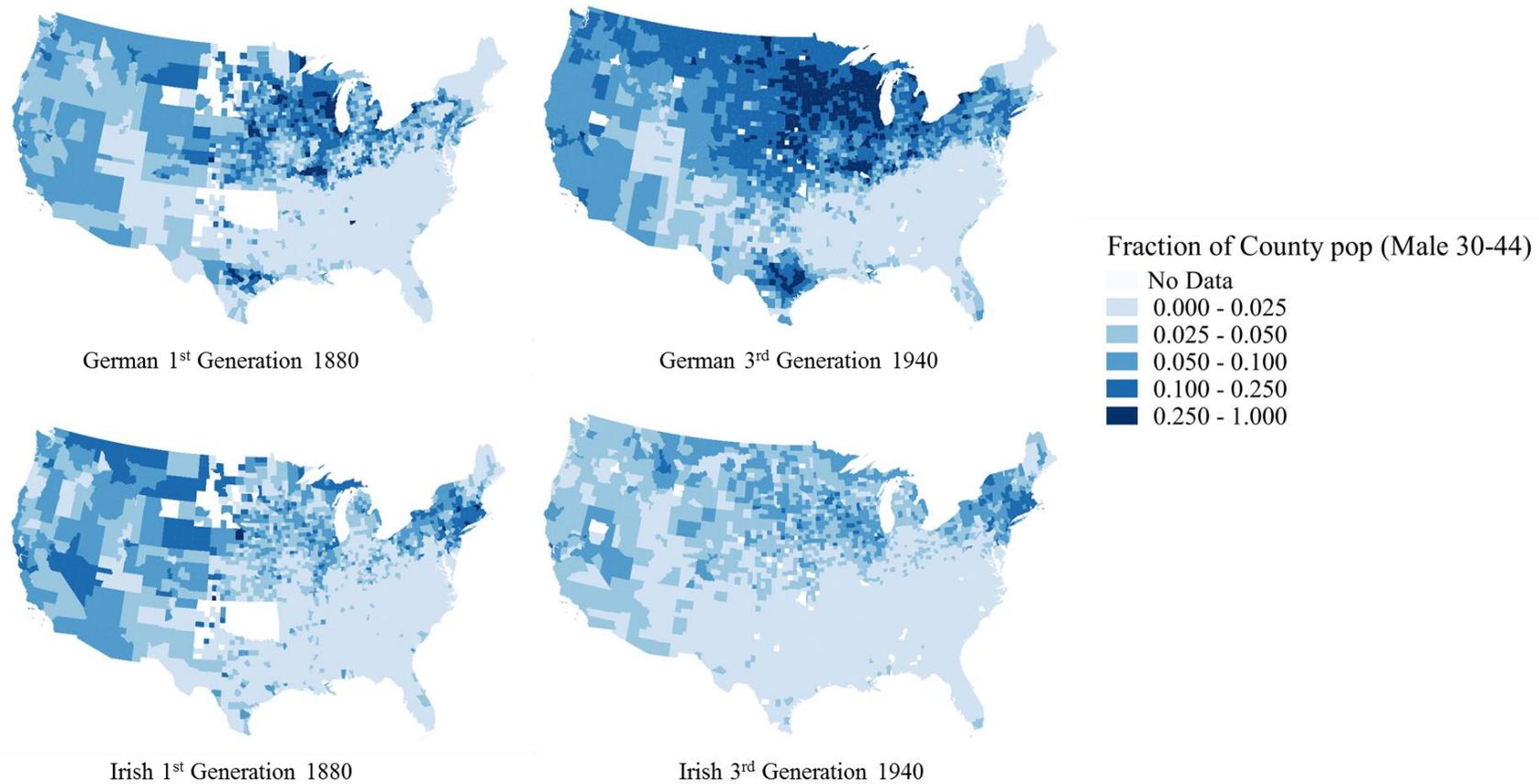
Notes: Data is from the 1880 full-count census and 1940 sample successfully linked to the 1910 census. Note the denominator is not the county's population, but the count from the sample. The sample is males aged 30 to 44 years old who hold an occupation, with the first and second-plus generation in 1880 and the third and fourth-plus generation in 1940.

Figure 2. Steps in the Linking Process

1880 Census	1910 Census	1940 Census

parent (G1) / child (G2) → parent (G2) / child (G3) → parent (G3)

(link 1)   (link 2)
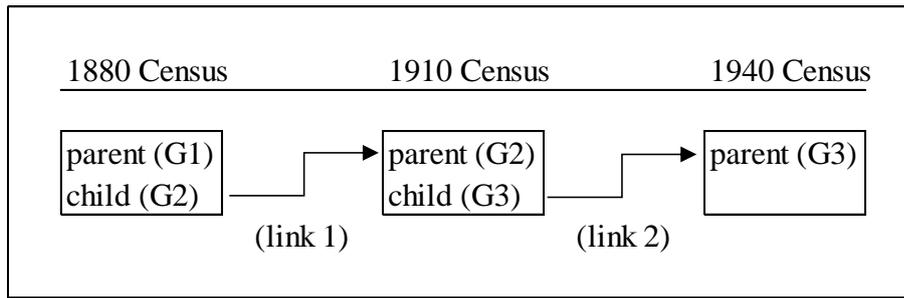
Figure 3. Predicted Convergence of Ethnic Group Averages from AR(2) Model
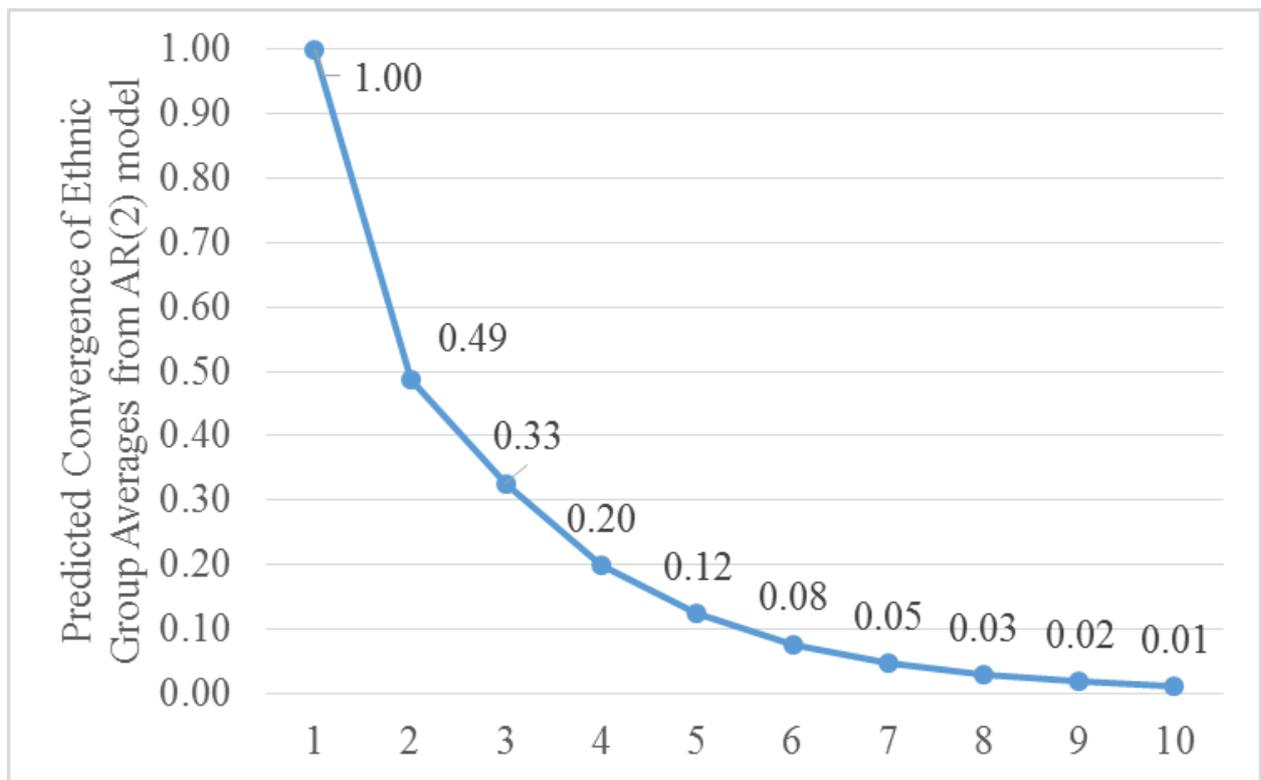
Notes: Results from the AR(2) model using the 1901 log occupational score. We assume that skill gaps converge between the first and second generation from the 1880-1910 AR(1) model.

Table 1. Skill level and number of observations from Grandfather-Father-Grandson linked sample

Origin in G1	Generation 1 in 1880		Generation 2 in 1910		Generation 3 in 1940	
	Log (Occ. Score)	N of G1	Log (Occ. Score)	N of G2	Log (Occ. Score)	N of G3
Canada	9.90	3,250	9.99	3,440	10.11	4,598
Mexico	9.70	160	9.67	169	9.73	219
Denmark	9.83	417	9.98	457	10.13	627
Norway	9.70	1,560	9.84	1,702	9.95	2,403
Sweden	9.80	1,127	9.96	1,206	10.08	1,582
England	9.97	6,223	10.03	6,798	10.13	9,163
Scotland	10.01	1,175	10.08	1,250	10.14	1,643
Ireland	9.96	9,412	10.12	9,831	10.18	13,199
Belgium	9.77	164	9.74	186	10.02	280
France	9.96	691	9.99	745	10.15	1,071
Netherlands	9.86	602	10.00	664	10.13	932
Switzerland	9.89	869	9.96	953	10.09	1,352
Italy	10.03	128	10.11	139	10.22	182
Austria/Hungary	10.01	222	10.13	227	10.13	322
Germany	9.96	29,625	10.02	32,758	10.11	45,411
Russia/Poland	9.90	176	10.06	180	10.12	246
Other	9.99	210	10.05	223	10.12	318
Overall	9.94	56,011	10.03	60,928	10.12	83,548

Notes: Data is from the 1880-1910-1940 linked sample of grandfathers to fathers to sons. The log occupational scores are presented using *occscore* from IPUMS.

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

	Farmer		White-Collar	
Parental Capital (β_1)	0.496	0.294	0.310	0.236
	(0.00652)	(0.00934)	(0.0122)	(0.0126)
Ethnic Capital (β_2)		0.325		0.348
		(0.0101)		(0.0181)
Mean Convergence ($\beta_1+\beta_2$)		0.619		0.584
		(0.0072)		(0.0190)
	Semi-Skilled		Unskilled	
Parental Capital (β_1)	0.157	0.117	0.202	0.122
	(0.00968)	(0.00995)	(0.00870)	(0.00933)
Ethnic Capital (β_2)		0.220		0.249
		(0.0156)		(0.0110)
Mean Convergence ($\beta_1+\beta_2$)		0.337		0.371
		(0.0166)		(0.0114)
	ln(Occ. Sc.), 1950		ln(Occ. Sc.), 1901	
Parental Capital (β_1)	0.413	0.220	0.288	0.168
	(0.0112)	(0.0124)	(0.00973)	(0.0102)
Ethnic Capital (β_2)		0.408		0.316
		(0.0134)		(0.0116)
Mean Convergence ($\beta_1+\beta_2$)		0.628		0.484
		(0.0130)		(0.0124)

Notes: Data is the 1880-1910 link from the 1880-1910-1940 linked dataset. There are 60,928 observations in each regression. The dependent variable is the second generation's outcome, which varies by either occupational category or score in each panel. 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.

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

	Farmer		White-Collar	
Parental Capital (β_1)	0.297	0.186	0.291	0.202
	(0.00613)	(0.00780)	(0.00831)	(0.00878)
Ethnic Capital (β_2)		0.173		0.338
		(0.00800)		(0.0119)
Mean Convergence ($\beta_1+\beta_2$)		0.360		0.540
		(0.0068)		(0.0121)
	Semi-Skilled		Unskilled	
Parental Capital (β_1)	0.0946	0.0747	0.114	0.0831
	(0.00781)	(0.00804)	(0.00839)	(0.00862)
Ethnic Capital (β_2)		0.0999		0.124
		(0.0111)		(0.0112)
Mean Convergence ($\beta_1+\beta_2$)		0.175		0.208
		(0.0119)		(0.0122)
	ln(Occ. Sc.), 1950		ln(Occ. Sc.), 1901	
Parental Capital (β_1)	0.298	0.152	0.304	0.153
	(0.00863)	(0.00932)	(0.00989)	(0.0102)
Ethnic Capital (β_2)		0.307		0.335
		(0.00869)		(0.0113)
Mean Convergence ($\beta_1+\beta_2$)		0.459		0.488
		(0.0096)		(0.0122)

Notes: Data is the 1910-1940 link from the 1880-1910-1940 linked dataset. There are 83,548 observations in each regression. The dependent variable is the second generation's outcome, which varies by either occupational category or score in each panel. 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. Predicted and Actual Correlation between First Generation in 1880 and Third Generation in 1940

	Grandfather's occupation			1st Generation Ethnic Capital			Mean Convergence		
	Predicted AR(1)	Actual	Difference (p-value)	Predicted AR(1)	Actual	Difference (p-value)	Predicted AR(1)	Actual	Difference (p-value)
Farmer	0.0548	0.120	0.0657	0.168	0.141	-0.026	0.223	0.260	0.037
	(0.003)	(0.00665)	(0)	(0.005)	(0.00671)	(0)	(0.005)	0.00548	(0)
White Collar	0.0477	0.117	0.0704	0.268	0.322	0.055	0.315	0.440	0.125
	(0.003)	(0.0119)	(0)	(0.012)	(0.0175)	(0)	(0.013)	0.0183	(0)
Skilled	0.00871	0.0142	0.0058	0.0501	0.0628	0.0126	0.0588	0.0770	0.0182
	(0.001)	(0.00843)	(0.508)	(0.005)	(0.0124)	(0.313)	(0.005)	0.0129	(0.155)
Unskilled	0.0101	0.0339	0.02372	0.0669	0.0541	-0.0126	0.0771	0.0880	0.0109
	(0.001)	(0.00812)	(0.003)	(0.005)	(0.00959)	(0.176)	(0.005)	0.0101	(0.268)
Log (Occ. Score), 1950 basis	0.0336	0.106	0.0729	0.254	0.282	0.026	0.288	0.388	0.100
	(0.003)	(0.00987)	(0)	(0.008)	(0.0103)	(0.016)	(0.009)	0.0101	(0)
Log (Occ. Score), 1901 basis	0.0256	0.0622	0.0356	0.210	0.248	0.037	0.236	0.310	0.074
	(0.002)	(0.0104)	(0)	(0.008)	(0.0115)	(0)	(0.008)	0.0131	(0)

Notes: Data is from the 1880-1910-1940 linked sample. There are 83,548 observations in each regression. The predicted columns assume an AR(1) process using coefficients from Tables 2 and 3. The actual correlation is estimated from Equation (6) after controlling 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. The dependent variable is the third generation's outcome, which varies by either occupational category or score in each row. Standard errors are clustered by G1 father.

Table 5. An AR(2) Model of Ethnic Capital

	Farmer	White-Collar	Semi-Skilled	Unskilled	Log Score, 1950	Log Score, 1901
G2 Parental Capital	0.169 (0.009)	0.192 (0.009)	0.074 (0.008)	0.081 (0.009)	0.144 (0.009)	0.148 (0.010)
G2 Ethnic Capital	0.137 (0.009)	0.296 (0.013)	0.093 (0.012)	0.112 (0.012)	0.245 (0.009)	0.291 (0.012)
G1 Grandfather Capital	0.039 (0.007)	0.047 (0.012)	0.002 (0.008)	0.018 (0.008)	0.041 (0.010)	0.012 (0.010)
G1 Ethnic Capital	0.040 (0.007)	0.152 (0.017)	0.022 (0.013)	0.005 (0.010)	0.113 (0.010)	0.101 (0.012)

Sum of G2 / G1 parental and ethnic capital:

	Farmer	White-Collar	Semi-Skilled	Unskilled	Log Score, 1950	Log Score, 1901
G2 Mean Convergence	0.306 (0.009)	0.488 (0.013)	0.168 (0.013)	0.193 (0.013)	0.390 (0.011)	0.439 (0.013)
G1 Mean Convergence	0.079 (0.007)	0.199 (0.019)	0.024 (0.014)	0.023 (0.011)	0.153 (0.011)	0.112 (0.013)

Notes: Data is from the 1880-1910-1940 linked sample. There are 83,548 observations in each regression. The dependent variable is the third generation (G3) outcomes, which varies across columns. We control for life-cycle effects with a quartic of grandson's age, quartic of father's age, quartic of grandfather's age, quartic of grandson's age interacted with grandfather's outcome, and quartic of grandson's age interacted with father's outcome; these quartics are normalized to age 40. Standard errors are clustered by G1 grandfather.

41

Table 6. Results are robust to quality of links

	Farmer		White Collar	
Parental Capital (β_1)	0.120	0.152	0.117	0.164
	(0.00665)	(0.0143)	(0.0119)	(0.0256)
Ethnic Capital (β_2)	0.141	0.105	0.322	0.313
	(0.00671)	(0.0135)	(0.0175)	(0.0362)
Mean Convergence ($\beta_1+\beta_2$)	0.260	0.257	0.440	0.477
	(0.005)	(0.012)	(0.018)	(0.037)
		Exactly-		Exactly-
Sample	Main	Linked	Main	Linked

	Semi-Skilled		Skilled	
Parental Capital (β_1)	0.0142	0.0189	0.0339	0.0645
	(0.00843)	(0.0194)	(0.00812)	(0.0193)
Ethnic Capital (β_2)	0.0628	0.0720	0.0541	0.0569
	(0.0124)	(0.0280)	(0.00959)	(0.0213)
Mean Convergence ($\beta_1+\beta_2$)	0.0770	0.0910	0.0880	0.121
	(0.013)	(0.031)	(0.010)	(0.023)
		Exactly-		Exactly-
Sample	Main	Linked	Main	Linked

	Log Score, 1950 Basis		Log Score, 1901 Basis	
Parental Capital (β_1)	0.106	0.153	0.0622	0.121
	(0.00987)	(0.0230)	(0.0104)	(0.0230)
Ethnic Capital (β_2)	0.282	0.264	0.248	0.237
	(0.0103)	(0.0238)	(0.0115)	(0.0247)
Mean Convergence ($\beta_1+\beta_2$)	0.388	0.417	0.310	0.357
	(0.010)	(0.023)	(0.013)	(0.031)
		Exactly-		Exactly-
Sample	Main	Linked	Main	Linked

Notes: Data is from the 1880-1910-1940 linked samples. The main sample's results are the actual coefficients reported in Table 4. The exactly-linked sample are those who match exactly in the 1880-1910 and 1910-1940 link on year of birth, first name string and last name string. There are 83,548 in the main sample and 16,760 observations in the exactly linked sample.

Table 7. Correlation between First and Third Generation after accounting for Neighborhood

	Farmer		White Collar	
G1 Grandfather	0.120	0.118	0.117	0.116
Capital	(0.00665)	(0.00708)	(0.0119)	(0.0128)
G1 Ethnic	0.141	0.00958	0.322	0.0563
Capital	(0.00671)	(0.0165)	(0.0175)	(0.0311)
1880 Neighborhood FE	N	Y	N	Y
	Semi-Skilled		Unskilled	
G1 Grandfather	0.0142	0.0141	0.0339	0.0413
Capital	(0.00843)	(0.00922)	(0.00812)	(0.00873)
G1 Ethnic	0.0628	0.0139	0.0541	0.0495
Capital	(0.0124)	(0.0220)	(0.00959)	(0.0194)
1880 Neighborhood FE	N	Y	N	Y
	Log Occ. Score, 1950		Log Occ. Score, 1901	
G1 Grandfather	0.106	0.0972	0.0622	0.0634
Capital	(0.00987)	(0.0102)	(0.0104)	(0.0104)
G1 Ethnic	0.282	0.0554	0.248	0.0528
Capital	(0.0103)	(0.0249)	(0.0115)	(0.0202)
1880 Neighborhood FE	N	Y	N	Y

Notes: Data is from the 1880-1940 portion of the 1880-1910-1940 linked sample. There are 83,548 observations in each regression. The dependent variable is the third generation (G3) outcomes, which vary across panels. We 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 neighbor fixed effect controls for the 1880 enumeration district.

Online Appendix

A. Further Details on Linking

In this section, we further describe linking the 1880 to 1910 Censuses and 1910 and 1940 Censuses.

For each census, we first draw all male individuals under the age of 14 from the 1880 and 1910 full-count censuses, no matter if they are first-generation, second-generation or third-generation Americans. There are 9,595,033 male children under the age of 14 in 1880, and 14,793,768 in 1910. We wish to link these individuals thirty years later to their adult outcomes in either 1910 or 1940. We draw the entire set of boys so that we can drop those with duplicate first name strings, last name strings, race, year of birth, and place of birth. We also extract the set of potential links in the 1910 and 1940 census for male adults between 27 and 47. Note the age range is different from 30 to 44 because we match on a 3-year window.

Before dropping duplicates, we pre-process each census in the following way. First, we keep only first names, or if the first name is only an initial with a longer middle name, the middle name. We then eliminate unusual strings from the first name such as "…", "---" or spaces; these appear if data clerks had difficulty reading the scanned census images. We also convert all nick names to full names in case one would list a name such as "Nick" while a child and then "Nicholas" while an adult. After cleaning these strings, we then create variables that are the NYSIIS coding of the first name and last name. The NYSIIS algorithm is a phonetic algorithm that allows one to match names that are spelled in slightly different ways. For example, John is coded as jan, and Jon is also coded as jan. However, our algorithm will prefer a John-John match over a John-Jon match; this method just gives us a set of potential matches.

After cleaning each census, we drop duplicates who have the same cleaned first name, cleaned last name, race, year of birth and state of birth. For the 1880-1910 link, we include mother and father's country/state of birth as an additional way to separate individuals.

However, before doing this, we need to clean country of birth for the 1910 Census since it has not been cleaned yet by IPUMS, as described in Appendix E. After dropping duplicates, we are left with 8,672,192 boys in 1880 and 13,042,894 boys in 1910, or 90.4 percent of the 1880 starting set and 88.2 percent of the 1910 starting set.

To link to the census 30 years later, we first find all pairwise combinations that match on NYSIIS first name, NYSIIS last name, race and place of birth; we additionally match on mother and father's place of birth between 1880 and 1910. We drop anyone who is not near for year of birth (plus/minus three years). To find the best match, for each potential link we quantify the string distance of first name and last name by the Jaro-Winkler score and the absolute difference in year of birth. We then sum these scores to calculate a match score for each potential link; this mostly follows the process of Pérez (2016). We keep the minimum matching score for each individual in Census X and then Census Y so that each individual is only matched to one other person.

This linking process leaves us with 1,117,989 links between 1880 and 1910, or 12.2 percent of the original amount, and 5,249,259 links between 1910 and 1940, or 35.5 percent of the original sample. A lower linking rate for the 1880-1910 set is likely because of a higher death rate, linking on extra characteristics of mother's and father's place of birth, and lower quality of underlying data.

We are interested in a specific subset of these linked samples, keeping only those with 30-55 year-old grandfathers in 1880, and where grandfather-father-son can be followed from 1880 to 1940. Moreover, we are only interested in foreign-born G1, native-born G2, and native born G3. First, attaching fathers to the 1910-1940 link drops the sample to 4,735,934, to 9.8 percent of links. We then determine if these attached fathers are also sons from the 1880-1910 link, leaving us with 479,804 observations; note that a large reason for the cut off is because fathers must be linked to the earlier 1880 census. Another 45,515 of this set are dropped because a father (or G1 grandfather) cannot be found in the 1880 census; moreover an extra 82,640 are dropped to keep G1 1880 grandfathers in the 30-55 year old age range.

We are left with 317,139 grandfather-father-son links from 1880-1910-1940. However, we have two more restrictions to the dataset. First, we only keep foreign-born G1, native-born G2, and native-born G3; this leads to sample of 92,193 grandsons to grandfathers. Then, we only keep those who have occupations for each generation; this drops to our final sample of 83,548 grandsons linked to grandfathers.

B. Representativeness and Weighting

The linking algorithm does not successfully find people for a random subset of sons in 1880 to 1910 or 1910 to 1940; rather, a link can only be found if someone has a unique combination of first name, last name, state/country of birth and age. Therefore, in cases where people cannot be distinguished from each other, such as for populous states or countries of birth, it is less likely to find someone. Moreover, if an ethnicity has more common names rather than unique names, it will also be less likely to find someone.

We check the biases of the sample by comparing the fathers of successfully linked sons to other fathers in the full-count censuses. We are interested if there are biases by country of birth, skill level, or region in the United States. Table B1 displays the characteristics of the fathers of children linked between 1880 and 1910 and universe of 1880 similarly aged fathers.[49] Many of the biases in representativeness are small, but there are a few significant differences in the linked sample and the universe. First, Germans are much more likely to be linked while Irish are much less likely to be linked. It is unclear what is driving this result, but it may be that Irish were from populous states like New York and Massachusetts, which makes it difficult to determine a unique link. Indeed, Northeast residents are less likely to be matched while Midwest residents are more likely to be matched. Another reason for different linking rates is that the Irish may have more similar names than Germans

Another bias in the representativeness of the sample is that sons with farmer fathers are more likely to be linked than sons with unskilled fathers. One possibility for this relates to differential death rates; it could be that death rates were higher for sons of unskilled fathers such that they were unlikely to be found in the 1910 Census. Some of these biases seem to be correlated: for example, having more farmer fathers in the Midwest is consistent with linking Irish at lower rates. Yet, there are still biases to representativeness within country of origin. For example, within Ireland, we are much less likely to find Northeastern unskilled workers than Midwestern farmers.

The directions of these biases for origin, occupational group and region exist between the 1910 and 1940 link (See Table B2). For this representativeness check in 1910, we compare our set of second-generation fathers with a successfully linked son to other similarly aged fathers of the 1910 census.[50] Now origin is not defined by country of birth but father's country

---

[49] The age is limited to 30-55 years old.
[50] That is, between ages 30 and 44.

of birth. Once again, those with German fathers are more likely to show up in our linked sample, while those with Irish fathers are less likely. Farmers are overrepresented and unskilled workers are underrepresented; similarly, the Northeast is underrepresented and the Midwest is overrepresented.

Fortunately, it is relatively straightforward to fix these biases in representativeness by weighting the sample to match the universe's characteristics. Given that we have full-count censuses in both 1880 and 1910, we have plenty of information to match the distribution in our sample to provide accurate weights. Of course, we can only match on observable characteristics; whether our weights match on unobservable characteristics is unknown. Given that the main biases appear for region, country of birth and occupational group, we weight our linked samples to match these characteristics. That is, we calculate the proportion of the population that is in each region / country of birth / occupation group cell from the full-count census; then we calculate the same proportion in our linked sample, and finally reweight our linked sample to match the population distribution. We group countries with less than 1000 individuals in the linked sample as an "other" country because of issues with small cells.

The above process creates weights for one of the linked sets between either the 1880 and 1910 Censuses or the 1910 to 1940 Censuses. These are shown in the "Single Weighted" or "Weighted" columns in Table B1 and B2. The outcomes for occupational categories, country of origin and residence are now aligned with the characteristics of the full population.

These weights make the sample more representative for one link; however, our main sample is double linked. Thus we cannot use the single weighed outcomes for our grandfather-grandson dataset. To create weights for the linked 1880 to 1940 sample, we pursue an iterative process. First, we create the weights between 1880 and 1910 such that each son has a weight in the 1910 census. Then using these weights, we calculate the proportions in each cell relative to the 1910 census in the same weight described above. The resulting characteristics of the 1880 census when applying these "Double Weights" are show in Table B1, where the linked sample is still representative of fathers in the 1880 Census.

Table B1. Representativeness of 1880 G1 Grandfathers

	Linked	Universe	Single Weighted	Double Weighted
Age	42.51	41.35	42.13	42.20
	(6.680)	(6.909)	(6.642)	(6.640)
White Collar	0.114	0.136	0.136	0.135
	(0.318)	(0.343)	(0.342)	(0.342)
Unskilled	0.287	0.402	0.403	0.394
	(0.452)	(0.490)	(0.491)	(0.49)
Skilled	0.160	0.184	0.184	0.183
	(0.366)	(0.387)	(0.388)	(0.386)
Farmer	0.439	0.278	0.277	0.289
	(0.496)	(0.448)	(0.447)	(0.453)
Northeast	0.267	0.399	0.404	0.345
	(0.442)	(0.490)	(0.491)	(0.475)
Midwest	0.602	0.482	0.477	0.519
	(0.490)	(0.500)	(0.499)	(0.500)
South	0.0767	0.0718	0.0718	0.0924
	(0.266)	(0.258)	(0.258)	(0.290)
West	0.0547	0.0472	0.0479	0.0440
	(0.227)	(0.212)	(0.214)	(0.205)
Log (Occ Score), 1950	9.866	9.946	9.948	9.942
	(0.377)	(0.370)	(0.371)	(0.374)
Log (Occ Score), 1901	9.628	9.649	9.649	9.644
	(0.406)	(0.413)	(0.416)	(0.419)
Germany	0.527	0.361	0.369	0.393
	(0.499)	(0.480)	(0.483)	(0.488)
Ireland	0.169	0.273	0.280	0.258
	(0.375)	(0.446)	(0.449)	(0.437)
England	0.111	0.0923	0.0945	0.0950
	(0.315)	(0.289)	(0.292)	(0.293)
Canada	0.0578	0.0762	0.0781	0.0713
	(0.233)	(0.265)	(0.268)	(0.257)
France	0.0124	0.0190	0.0194	0.0222
	(0.110)	(0.136)	(0.138)	(0.147)
Netherlands	0.0107	0.00926	0.0200	0.0186
	(0.103)	(0.0958)	(0.140)	(0.135)
Switzerland	0.0154	0.0142	0.0146	0.0165
	(0.123)	(0.118)	(0.120)	(0.127)
Russia/Poland	0.00318	0.00388	0.00522	0.00490
	(0.0563)	(0.0622)	(0.0720)	(0.0698)
Norway	0.0277	0.0259	0.0264	0.0255
	(0.164)	(0.159)	(0.160)	(0.158)
Sweden	0.0201	0.0244	0.0250	0.0260
	(0.140)	(0.154)	(0.156)	(0.159)
Scotland	0.0212	0.0243	0.0249	0.0249
	(0.144)	(0.154)	(0.156)	(0.156)
Italy	0.00224	0.00497	0.00563	0.00578
	(0.0473)	(0.0704)	(0.0748)	(0.0758)
Observations	57,610	1,016,713	57,610	57,610

Notes: Universe is 1880 full count of foreign-born fathers. Weighted is to match 1880 universe. Double weighted is also weighting to match 1910 father attributes.

Table B2. Representativeness of 1880 G2 Fathers

	Linked	Universe	Weighted
Age	37.58	36.74	37.50
	(4.334)	(4.123)	(4.326)
White Collar	0.252	0.246	0.248
	(0.434)	(0.431)	(0.432)
Unskilled	0.224	0.318	0.315
	(0.417)	(0.466)	(0.464)
Skilled	0.184	0.204	0.204
	(0.387)	(0.403)	(0.403)
Farmer	0.340	0.233	0.234
	(0.474)	(0.423)	(0.423)
Northeast	0.252	0.316	0.317
	(0.434)	(0.465)	(0.465)
Midwest	0.575	0.486	0.487
	(0.494)	(0.500)	(0.500)
South	0.0846	0.104	0.104
	(0.278)	(0.305)	(0.305)
West	0.0889	0.0938	0.0921
	(0.285)	(0.292)	(0.289)
Log (Occ Score), 1950	9.982	10.01	10.03
	(0.453)	(0.462)	(0.442)
Log (Occ Score), 1901	9.753	9.749	9.773
	(0.441)	(0.425)	(0.419)
Germany	0.538	0.402	0.403
	(0.499)	(0.490)	(0.490)
Ireland	0.163	0.217	0.251
	(0.369)	(0.412)	(0.434)
England	0.112	0.0922	0.0954
	(0.315)	(0.289)	(0.294)
Canada	0.0560	0.0604	0.0698
	(0.230)	(0.238)	(0.255)
France	0.0120	0.0154	0.0218
	(0.109)	(0.123)	(0.146)
Netherlands	0.0108	0.0100	0.0188
	(0.103)	(0.0997)	(0.136)
Switzerland	0.0155	0.0117	0.0165
	(0.124)	(0.107)	(0.128)
Russia/Poland	0.00270	0.00588	0.00401
	(0.0519)	(0.0765)	(0.0632)
Norway	0.0277	0.0267	0.0258
	(0.164)	(0.161)	(0.159)
Sweden	0.0198	0.0188	0.0259
	(0.139)	(0.136)	(0.159)
Scotland	0.0206	0.0264	0.0242
	(0.142)	(0.160)	(0.154)
Italy	0.00227	0.00457	0.00585
	(0.0476)	(0.0675)	(0.0763)
Observations	62,495	809,557	62,495

Notes: Universe is 1910 full count of second-generation fathers. Weighted is to match 1910 universe as described in text.

C. Results When Limiting Sample to Europeans

In this section, we create the tables that link the G1-G2 (Table 2), G2-G3 (Table 3), and G1-G3 (Table 4) using the subset of our sample that is European. The main sources this drops are Canadian and Mexican immigrants. All qualitative conclusions hold.

Table C1. Persistence from the First Generation in 1880 to the Second Generation in 1910

	Farmer		White-Collar	
Parental Capital (β_1)	0.507	0.298	0.306	0.236
	(0.00662)	(0.00972)	(0.0124)	(0.0129)
Ethnic Capital (β_2)		0.330		0.338
		(0.0106)		(0.0189)
Mean Convergence ($\beta_1+\beta_2$)		0.628		0.574
		(0.0074)		(0.0197)
	Semi-Skilled		Unskilled	
Parental Capital (β_1)	0.147	0.108	0.196	0.118
	(0.00993)	(0.0102)	(0.00891)	(0.00958)
Ethnic Capital (β_2)		0.224		0.248
		(0.0162)		(0.0114)
Mean Convergence ($\beta_1+\beta_2$)		0.331		0.366
		(0.0170)		(0.0117)
	ln(Occ. Sc.), 1950		ln(Occ. Sc.), 1901	
Parental Capital (β_1)	0.418	0.224	0.288	0.171
	(0.0103)	(0.0114)	(0.0101)	(0.0106)
Ethnic Capital (β_2)		0.410		0.319
		(0.0118)		(0.0123)
Mean Convergence ($\beta_1+\beta_2$)		0.633		0.490
		(0.0119)		(0.0131)

Notes: This table recreates Table 2, but keeping only European sources.

Table C2. Persistence from the Second Generation in 1910 to the Third Generation in 1940

	Farmer		White-Collar	
Parental Capital (β_1)	0.279	0.191	0.278	0.198
	(0.00368)	(0.00806)	(0.00500)	(0.00905)
Ethnic Capital (β_2)		0.174		0.339
		(0.00838)		(0.0124)
Mean Convergence ($\beta_1+\beta_2$)		0.364		0.537
		(0.0069)		(0.0126)
	Semi-Skilled		Unskilled	
Parental Capital (β_1)	0.0925	0.0753	0.0953	0.0785
	(0.00465)	(0.00827)	(0.00512)	(0.00882)
Ethnic Capital (β_2)		0.0982		0.105
		(0.0116)		(0.0116)
Mean Convergence ($\beta_1+\beta_2$)		0.174		0.183
		(0.0124)		(0.0126)
	ln(Occ. Sc.), 1950		ln(Occ. Sc.), 1901	
Parental Capital (β_1)	0.296	0.149	0.291	0.154
	(0.00518)	(0.00923)	(0.00553)	(0.0102)
Ethnic Capital (β_2)		0.308		0.313
		(0.00894)		(0.0113)
Mean Convergence ($\beta_1+\beta_2$)		0.457		0.467
		(0.0094)		(0.0121)

Notes: This table recreates Table 3, but keeping only European sources.

Table C3. Predicted and Actual Correlation between First Generation in 1880 and Third Generation in 1940

	Grandfather's occupation			1st Generation Ethnic Capital			Mean Convergence		
	Predicted AR(1)	Actual	Difference (p-value)	Predicted AR(1)	Actual	Difference (p-value)	Predicted AR(1)	Actual	Difference (p-value)
Farmer	0.0568	0.122	0.0657	0.172	0.143	-0.026	0.229	0.265	0.036
	(0.003)	(0.00695)	(0)	(0.005)	(0.00706)	(0)	(0.005)	(0.006)	(0)
White Collar	0.0468	0.115	0.0704	0.262	0.312	0.055	0.309	0.428	0.119
	(0.003)	(0.0122)	(0)	(0.012)	(0.0183)	(0)	(0.013)	(0.019)	(0)
Skilled	0.00810	0.0155	0.0058	0.0493	0.0717	0.0126	0.0574	0.0873	0.0299
	(0.001)	(0.00870)	(0.390)	(0.005)	(0.0130)	(0.094)	(0.005)	(0.013)	(0.028)
Unskilled	0.00923	0.0377	0.02372	0.0577	0.0404	-0.0126	0.0669	0.0781	0.0112
	(0.001)	(0.00827)	(0)	(0.005)	(0.00987)	(0.076)	(0.005)	(0.010)	(0.262)
Log (Occ. Score), 1950 basis	0.0333	0.104	0.0729	0.256	0.282	0.026	0.290	0.386	0.096
	(0.003)	(0.00975)	(0)	(0.007)	(0.0105)	(0.019)	(0.008)	(0.010)	(0)
Log (Occ. Score), 1901 basis	0.0264	0.0632	0.0356	0.203	0.234	0.037	0.229	0.297	0.068
	(0.002)	(0.0106)	(0)	(0.008)	(0.0115)	(0)	(0.008)	(0.013)	(0)

Notes: This table recreates Table 4, but keeping only European sources.

D. Neighborhood Effects for Two Generations

In this section we report results when controlling for neighborhood effects for the G1-G2 and G2-G3 links. These recreate the results of Table 7, which shows the link between G1-G3. The results between two generations consistently find a positive effect of ethnic capital despite controlling for neighborhood fixed effects, in contrast to the G1-G3 results in Table 7.

Table D1. Neighborhood Fixed Effects from G1 in 1880 to G2 in 1910

	Farmer		White Collar	
G1 Parental	0.294	0.290	0.236	0.228
Capital	(0.00934)	(0.0104)	(0.0126)	(0.0138)
G1 Ethnic	0.325	0.0673	0.348	0.122
Capital	(0.0101)	(0.0265)	(0.0181)	(0.0313)
1880 Neighborhood FE	N	Y	N	Y
	Semi-Skilled		Unskilled	
G1 Parental	0.117	0.116	0.122	0.120
Capital	(0.00995)	(0.0114)	(0.00933)	(0.0103)
G1 Ethnic	0.220	0.0639	0.249	0.107
Capital	(0.0156)	(0.0267)	(0.0110)	(0.0225)
1880 Neighborhood FE	N	Y	N	Y
	Log Occ. Score, 1950		Log Occ. Score, 1901	
G1 Parental	0.220	0.208	0.168	0.160
Capital	(0.0124)	(0.0132)	(0.0102)	(0.0108)
G1 Ethnic	0.408	0.0826	0.316	0.0479
Capital	(0.0134)	(0.0266)	(0.0116)	(0.0209)
1880 Neighborhood FE	N	Y	N	Y

Notes: This table recreates Table 7, but with the 1880 to 1910 portion of the 1880-1910-1940 dataset. There are 60,928 observations in each regression.

Table D2. Neighborhood Fixed Effects from G2 in 1910 to G3 in 1940

	Farmer		White Collar	
G2 Parental	0.186	0.187	0.202	0.196
Capital	(0.00780)	(0.00973)	(0.00878)	(0.0116)
G2 Ethnic	0.173	0.0484	0.338	0.0664
Capital	(0.00800)	(0.0227)	(0.0119)	(0.0276)
1910 Nbhd FE	N	Y	N	Y
	Semi-Skilled		Unskilled	
G2 Parental	0.0747	0.0823	0.0831	0.0895
Capital	(0.00804)	(0.0107)	(0.00862)	(0.0115)
G2 Ethnic	0.0999	0.0151	0.124	0.0767
Capital	(0.0111)	(0.0247)	(0.0112)	(0.0239)
1910 Nbhd FE	N	Y	N	Y
	Log Occ. Score, 1950 basis		Log Occ. Score, 1901 basis	
G2 Parental	0.152	0.151	0.153	0.149
Capital	(0.00932)	(0.0110)	(0.0102)	(0.0123)
G2 Ethnic	0.307	0.0517	0.335	0.0839
Capital	(0.00869)	(0.0222)	(0.0113)	(0.0241)
1910 Nbhd FE	N	Y	N	Y

Notes: This table recreates Table 7, but with the 1910 to 1940 portion of the 1880-1910-1940 dataset. There are 83,548 observations in each regression.

E. Further Data Details

Coding of country of birth

An issue when placing people into ethnic groups across the 1880 to 1940 period is that country borders change, especially following World War I. However, we define most of our sample based on countries in the 1880 Census or fathers' country of birth in the 1910 census. Sometimes individuals report a sub-geographical area that is not truly a country in 1880. We make the following assumptions when grouping people; however, these issues are not as important because few observations reported different countries.

- Russia includes Russia, Poland, Latvia, Lithuania, Estonia and any Baltic state
- Austro-Hungary includes Austria, Hungary, Czechoslovakia and Yugoslavia

Cleaning the 1910 Census

We use the full-count data in 1880, 1910 and 1940 available from the University of Minnesota Population Center. At the time of writing this paper, the 1910 US Censuses has not been cleaned and the 1940 Census has been cleaned on a preliminary basis. The primary variables we are interested in cleaning are age, occupation, country of birth, relationship status and household head. The process of cleaning the 1910 Census is described in further detail below.

- Own, mother's and father's country of birth

To clean the country of birth strings, we rely heavily on the strings already cleaned by the University of Minnesota Population Center for the 1850, 1880 and 1920 to 1940 full-count data. We create files that yield the most common country of birth codes (BPL) for each country of birth string (BPLSTR). We do this for each full-count Census cleaned by IPUMS, leaving us five different files from the 1850, 1880 and 1920 to 1940 Censuses. Armed with these files, we simply merge them to the uncleaned censuses starting with the 1880 Census since this was before border changes following World War I. For BPLSTR that are unmatched, we simply merge them onto later cleaned census files to update the BPL codes. For this process we merge first to the 1880 or 1850, depending on closeness in time, and then to the 1920 to 1940 Census files. This is because border changes following World War I cause the pre-World War I censuses to be more reliable for assigning BPL codes. After this initial pass, we have cleaned 99 percent of the country of birth strings. Following this, we tabulated a list of remaining strings and cleaned those which appeared more than 100 times. For country of birth strings which appeared less than 100 times, we left their BPL code as missing and dropped them from the

dataset. We clean the 1910 codes for mother and father's country of birth in the same way as for country of birth.

- Occupation

We clean the 1910 occupation strings in a similar way as cleaning country of birth, where we rely on the cleaned full-count censuses OCCSTR and OCC1950 codes. After removing spaces and "-" or "." characters from the occupation string, we merge them with cleaned occupation codes. The primary reason why occupation strings were not matched for these codes were because someone listed two occupations with "&" separating the occupations. For these strings, we take the first occupation listed as the primary occupation.

- Relationship to head

We take the cleaned censuses by 1920-1940 full-count censuses and download the list of relationship strings and coded relationship status. We then merge the 1910 census with relationship strings; this matches with nearly the entire data set since there is not much variation in relationship strings. For relationships that are unmatched, we coded them as other non-family members of the household; often the unmatched strings appear to be occupations by others in the household such as cook or maid. We are primarily interested in the father and son status, so this likely does not miss many pairs of interest.

Additional Appendix References

Pérez, Santiago. "The (South) American Dream: Mobility and Economic Outcomes of First- and Second-Generation Immigrants in 19th-Century Argentina." (2017).