NBER WORKING PAPER SERIES

ETHNIC ENCLAVES AND IMMIGRANT OUTCOMES: NORWEGIAN IMMIGRANTS DURING THE AGE OF MASS MIGRATION

Katherine Eriksson

Working Paper 24763
http://www.nber.org/papers/w24763

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

I thank the NBER and IPUMS for access to full count censuses in 1910, 1920, and 1940. This paper benefited from feedback from participants at "HattonFest", PUC-Chile, UC-Davis, and the Social Science History Association. I acknowledge financial support from the Institute for Social Sciences at UC-Davis. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Ethnic Enclaves and Immigrant Outcomes: Norwegian Immigrants during the Age of Mass Migration

Katherine Eriksson

NBER Working Paper No. 24763
June 2018
JEL No. J61,N31

ABSTRACT

This paper examines the effect of ethnic enclaves on economic outcomes of Norwegian immigrants in 1910 and 1920, the later part of the Age of Mass Migration. Using different identification strategies, including county fixed effects and an instrumental variables strategy based on chain migration, I consistently find that Norwegians living in larger enclaves in the United States had lower occupational earnings, were more likely to be in farming occupations, and were less likely to be in white-collar occupations. Results are robust to matching method and choice of occupational score. This earnings disadvantage is partly passed on to the second generation.

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

I. Introduction

The effect of ethnic enclaves on immigrant wages and assimilation is an empirical question important for current immigration policy. Immigrants tend to choose areas with high concentrations of fellow immigrants. In 2000, the average county's population was 3.4 percent foreign-born, but the average immigrant lived in a county with 21.8 percent immigrants. At the end of the Age of Mass Migration in 1920, the average county had 12 percent foreign-born residents, but the average immigrant lived in a county with 35.7 percent immigrants.[2]

Theoretically, this spatial clustering could have positive or negative effects on wages. If immigrants arrive without local knowledge, fellow immigrants could help with finding a job or learning English. On the other hand, immigrants living in ethnic enclaves may have less incentive to learn about the larger job market or to acquire skills of value outside of the enclave.

The causal identification of ethnic enclaves is challenging for two main reasons: first, immigrants may self-select into ethnic enclaves based on skill; second, even if we could randomly assign immigrants to enclaves of varying sizes, larger ethnic enclaves may have different labor market opportunities than smaller ones. For example, Norwegian enclaves may be more rural with fewer job opportunities outside of farming.

This paper uses the full universe of Norwegian-born immigrants living in the United States in 1910 and 1920 to study the effect of ethnic enclaves on Norwegian immigrant occupational outcomes during the last part of the Age of Mass Migration. I show that larger ethnic enclaves were associated with lower occupational income for Norwegian immigrants arriving after 1900. One standard deviation increase in enclave size reduced immigrant occupational earnings

[2] Authors own calculations from IPUMS (Ruggles et al 2018).

by 10 percent.[3] This result is robust to county fixed effects and adding other countries as a control group.

To better establish causality, I use a new linked sample to construct an instrumental variable to handle self-selection of immigrants into enclaves. I link Norwegian immigrants living in the United States in 1880, 1910, and 1920 back to the previous census in Norway in either 1865 or 1900. This identifies the municipality of birth of the linked migrants. I then construct a chain-migration based instrument for enclave size in 1910 and 1920 by using the distribution of destinations of the earlier wave of migrants from a later migrant's municipality of birth. I show that this predicts actual enclave size well and that the effect of enclave size on occupational income is still negative.

I show that larger enclaves are also associated with a higher probability of marriage, plus an increased probability of working in farming and a decreased probability of being in a white-collar occupation. Finally, I show that the occupational earnings penalty faced by immigrants is passed on to the second generation: sons of Norwegian-born men who grow up in larger enclaves have lower wage earnings as adults, are more likely to be in farming, and are less geographically mobile.

This paper contributes to a literature about ethnic enclaves and immigrant outcomes that has mostly focused on contemporary data, though one exception is LaFortune and Tessada (2017) who look at ethnic networks the Age of Mass Migration, arguing that immigrants used ethnic networks to learn about new labor markets. Work using contemporary data largely uses

[3] One drawback of using a historical period is that individual wage data is not available. Therefore, I look at median occupational income as my outcome variable.

random refugee resettlement to identify causal effects of ethnic enclaves.[4] Edin, Frederiksson, and Aslund (2003) and Damm (2009) find negative selection into ethnic enclaves; after accounting for this selection, ethnic enclaves increase immigrant wages. Other recent work by Battisti, Peri, and Romiti (2016) uses a panel dataset of immigrants in Germany to control for pre-migration characteristics. They find initially positive effects of enclaves followed by negative effects on investment in skills.

II. Norwegian Immigration and Enclaves in the Age of Mass Migration

Norwegian migration to the United States started in 1825 with the first ship that left Norway for the eastern United States. The thirty families settled in Pennsylvania before eventually moving westward to Illinois as the Erie Canal made the Midwestern states viable agricultural areas. Letters home told of fertile soil, but it wasn't until the depression in 1849 and 1850 that large numbers of Norwegians moved. This was followed by the kickoff of a large wave of migration during a series of crop failures in the 1860's.

Norwegian communities sprung up in Wisconsin, Iowa, and Minnesota. The Norwegian community was tight-knit despite being sparsely spread in rural areas. They organized schools, churches, and helped each other at barn-raising events (Semminson, 1980). Semminson (1980) describes how immigrants would send letters or even money and ship tickets back to their hometowns; these relationships resulted in strong correlations over time in the location choices in the United States of families from different towns in Norway.

Figure 1 plots the size of the Norwegian immigrant population and the size of Norwegian enclaves by census year from 1850 to 1940. The Norwegian population in the United States grew

[4] For papers using random resettlement of refugees, see Aizer and Currie (2004), Beaman (2012), Bertrand and Mullainathan (2000), Damm (2009), and Edin, Fredriksson and Aslund (2003). Xie and Gough (2009) find some correlational evidence that enclaves depress wages of immigrants.

steadily from 1850 to 1910 before decreasing slightly by 1940. Norwegians were most likely to live in the same county as their fellow countrymen in 1870 and 1880, with the average Norwegian living in a county with about 17 percent Norwegians; this fell to about 4 percent by 1920.

Figure 2 shows the location of Norwegian immigrants in 1920. The counties with the largest concentrations of Norwegian immigrants are clustered in the upper Midwest: Minnesota, North and South Dakota, Iowa, and Wisconsin. The average Norwegian lived in a county with about 4 percent Norwegian immigrants in 1920, but the largest enclave size was 27.8% in Traill County, North Dakota. The correlation between a county's Norwegian share in 1870 and 1920 is 0.79, suggesting strong persistence over time. Interestingly, the top five counties from 1920 are also the five counties with the most individuals reporting Norwegian ancestry in 2000 (American Factfinder).

Norwegians were not the only ethnic group to cluster in enclaves. Figure 3 shows the average enclave size for a range of European countries in 1920. The Scandinavian countries of Sweden, Norway, and Finland had some of the largest enclave sizes with the average immigrant living with 3-5 percent of his own countrymen. England, Scotland, and Wales had small enclaves by 1920, though Irish enclaves were still large. Other countries with larger enclaves include Italy, Germany, and Russia. Countries with smaller immigrant populations in the United States also had smaller enclaves: Bulgaria, Spain, France, Belgium, and Switzerland all had enclave sizes well under one percent.

There is no obvious pattern in Figure 4 with respect to "new" versus "old" sending countries, but it is possible that the evolution of enclave size depends on the average arrival cohort of a country. Therefore, Figure 3 plots enclave size over time for four sending countries: Norway, Italy, England, and Ireland. England is the longest-standing immigrant group, and, interestingly,

English enclaves were never very large, reaching a maximum of around five percent in 1860-1870. On the other hand, Irish immigrants always lived in large enclaves and the size of these enclaves decreased over time. The average Irish immigrant lived with up to 20 percent of his own countrymen in 1850 and 1860. This decreased to less than four percent by 1940. Italian enclave sizes were small until 1900, after which they grew to six percent through 1930, falling slightly by 1940.

Norwegian enclaves grew large after the first big immigration wave in the 1860's, growing from seven percent in 1850 to over 16 percent in 1870 and 1880. Nonetheless, Norwegian enclaves in 1850 were larger than any time after 1920. Overall, these patterns are consistent with enclaves growing as the size of a group grows, and then falling thereafter as the group assimilates. However, this is not a necessary pattern as shown by England which had small enclaves throughout the nineteenth century. It's possible that discrimination (against the Irish) or the need to learn English drove enclave size in this period.[5]

III. Data

This paper estimates the effect of enclave size on Norwegian immigrant occupational income in 1910 and 1920. I use IPUMS samples and full count US census data from 1910 and 1920, along with a linked sample of Norwegian immigrants matched back to themselves in the 1900 Norwegian census to identify municipality of birth. I also construct a matched sample from

[5] These patterns are similar but not identical to the patterns of segregation over time for these countries found in Eriksson and Ward (2018). Using the method of Logan and Parman (2017), that paper traces segregation patterns over the same period by country of origin. English migrants are also much less segregated than immigrants from almost every other sending country. Italian segregation follows a similar pattern to the enclave measure here. However, Norwegian segregation falls continuously over this period, and Irish segregation is much lower than Norwegian segregation despite similar enclave sizes in 1860 and 1870.

the 1880 US census to the 1865 Norwegian census to identify Norwegian municipality of birth which is used in the instrument described in Section IV.

A. US Census Data

I extract all Norwegian-born men from the men from the restricted use full count census data on the NBER server. To this I add native-born and immigrants from other European countries from the 1910 and 1920 IPUMS 1% (Ruggles et al 2018). I restrict to men between the ages of 21 and 65 who report being in the labor force. The benefit of the IPUMS samples is that occupations are fully coded. In the full count data, up to 17% of men in the labor force have occupation strings that are not yet coded—I assign occ1950 codes by hand for Norwegian-born men, but use the 1% IPUMS samples for other groups.

I assign enclave size at the county level in my primary specification using the population counts by birth country and census year given in ICPSR 2896. Enclave size is measured as the percent of a county born in one's own country of birth; for Norwegians, it is the percent of the county that was born in Norway. I construct a similar measure at the enumeration district level using the Full Count data and assign this to Norwegian individuals.[6]

B. Assigning Occupational Income

The lack of individual wage income poses a challenge for papers using data from censuses before 1940. Even in 1940, income is only collected for wage earners which could exclude large parts of the population, particularly in rural areas. Therefore, the assignment of occupational earnings is an important question (Inwood et al 2018). The most commonly used measure, oc-

[6] Enumeration district is not identified in the IPUMS 1% 1920 sample, so I stick with county as my primary measure of enclave size; furthermore, enumeration districts do not have consistent boundaries over time so I am limited to county with my instrumental variables strategy.

cupational score ("occscore") is based on median earnings by occupation in the 1950 census. The drawback of this measure, however, is that the Great Compression of the 1940's sharply reduced wage inequality (Goldin and Margo 1992); therefore, differences in income across occupations will appear lower than they possibly actually were in older censuses. Second, the relative position of farmers fell dramatically over the first half of the twentieth century, so using occscore might understate earnings for populations which work heavily in farming.

Other papers have used data from earlier periods, some relying on the Cost of Living Survey in 1901 which collected earnings from an urban population (Abramitzky et al 2012). Farmer incomes are then imputed by using a method from the agricultural census which takes farm earnings and subtracts direct costs plus depreciation (see Abramitzky et al 2012 Online Appendix).

In this paper, I use the Cost of Living 1901 earnings as my primary measure. In a robustness table, I show that my results are robust to the following measures: (1) Occscore; (2) a newly constructed occscore from Saavedra and Twinam (2018) which adjusts occscore for age, gender, race, and state of residence ("Lido"); (3) Cost of Living income with farmer and farm laborer income imputed at the county level instead of using one number for all farmers; (4) Occscore adjusted with farmer and farm laborer income imputed at the county level.

I finally note that any results in this paper only capture differences in income across occupations, not within occupations. In particular, it is very likely that farmers living in larger Norwegian enclaves were *more* productive than farmers living in smaller Norwegian enclaves. This should be picked up somewhat by measures (3) and (4) above; in fact, the raw correlation between imputed farm income and Norwegian enclave size is 0.36.

C. Constructing Three Matched Samples

To construct the instrument described in Section IV below, I need to know the Norwegian municipality of birth of Norwegians living in the United States in 1880, 1910, and 1920. I rely on full count censuses from the United States in 1880, 1910 and 1920, as well as the Norwegian full count censuses from 1865 and 1900 from the North Atlantic Population Project (NAPP). I create three linked samples: 1865-1880, 1900-1910, and 1900-1920. I link men aged 18-65 in 1880 and 1910 and aged 21-65 in 1920 backwards to the relevant Norwegian census.

Following Abramitzky et al (2012), I proceed as follows to create these linked samples:

1. After accounting for common nicknames (e.g. Wm becomes William), I standardize first and last names in each dataset using the NYSIIS standardization.
2. I restrict to men who arrived in the United States after 1900 for the 1900-1910 and 1900-1920 linked samples. Unfortunately, the United States census did not start asking year of arrival until 1900, so I cannot restrict to those migrating after 1865 in the 1865-1880 sample.
3. I match individuals across censuses using standardized name and year of birth in an iterative manner: I first look for matches with the correct year of birth; then, I allow individuals to misreport their year of birth in either direction by one year; finally, I allow them to misreport year of birth by up to two years. In each stage, in the case of multiple matches, I flag the observations as unmatched and do not allow them to match again.

In a series of robustness samples, I change the matching procedure (Bailey et al 2017). First, I require uniqueness by standardized name within a five year (plus or minus two years) age band in both datasets. This should reduce false positives at the cost of a lower match rate. I also construct a sample in which I match on raw name strings instead of standardized names; this sample matches less than half as many men as the procedure using standardized names due to phonetic

name spelling differences across different countries' censuses. I also show a robustness sample which only allows individuals to misreport year of birth by up to one year.

Sample sizes and match rates are shown in Table 1 for the primary matched samples. The match rate is the lowest for the 1865-1880 match. There are likely two reasons: first, year of arrival is not reported in the 1880 census, so there are more likely to be non-unique cases in 1880 than in later years when I can drop those arriving before the earlier census; second, literacy rates were lower in 1880 than 1920, so men were likely more likely to misspell their names and to report an age rounded to the closest five or ten years instead of their actual age. The match rate in 1865-1880 is 8.52%, somewhat lower than that found by Abramitzky et al (2012); I am matching ages 18-65 in 1880 instead of a group which is young in 1865 (age 3 to 15 in 1865 in Abramitzky et al 2012), increasing the likelihood that men misreport their age or misspell their name. The rates increase to 14 and 16.1 percent in the 1900-1910 and 1900-1920 samples, respectively; these rates are consistent with the literature.

D. Comparing the Matched Sample to the Population and Summary Statistics

Table 2 presents summary statistics from the 1910 and 1920 Norwegian-born population in the United States and compares this population to the matched sample which I am able to find in Norway in 1900. Column (1) shows means for the population, Column (2) shows the differences between the matched sample and population, and Column (3) shows the differences after reweighting the matched sample to match the population on observable characteristics using inverse probability weights.

Norwegians in 1910 and 1920 lived in counties and enumeration districts where on average 4 percent and 7.1 percent, respectively, of the population were Norwegian-born. Norwegians

arriving in the United States were on average 32 years old by 1910 or 1920. Average income is approximately $570 in 1901 income ($16,400 in 2015 dollars) and the average occscore is 21.3 ($21,000 in 2015 dollars). Over 15 percent of Norwegians are farmers, and 27.4 are either laborers or farm laborers. Finally, 54.7 are married by 1910 or 1920.

Matched samples are by construction non-representative of the population—matching algorithms require individuals to be unique, names to be spelled correctly (up to a standardization), and ages to be reported within two years of the correct age. Therefore, unsurprisingly, the resulting matched sample is usually of slightly higher socio-economic status than the population. In Column (2), individuals in the matched samples live in slightly larger enumeration districts (0.2-0.4 percentage points), are older (0.6 years), have slightly higher occupational income (about 3-12% higher than the population), are more likely to be farmers (2.2 percentage points) and less likely to be laborers (2.2 percentage points), and are less likely to be married. After reweighting in Column (3), these differences all become much smaller and statistically insignificant except for the occscore difference which remains marginally significant and two-thirds as large as before. I show below that my main result is robust to reweighting the data.

IV. Regression Framework and Instrumental Variables Strategy

The main specification estimates the effect of enclave size, measured by the percent of a county or enumeration district's population born in Norway ("% Norwegian"), on occupational income. I run the following specification:

$$(1) \qquad Ln(OccupationalIncome)_i = \beta_0 + \beta_1 \%Norwegian_{ic} + X_i\gamma + e_i$$

for individual i living in county c. The main occupational income variable is taken from the 1901 Cost of Living survey, but I show results using four other occupational income measures.

A causal interpretation of equation (1) is difficult for two reasons: (a) individuals may sort into enclaves based on unobserved skill; for example, if all low-skilled individuals live in large enclaves, we will associate larger enclaves with bad outcomes purely due to selection; (b) larger enclaves have different labor market opportunities; at an extreme, if all Norwegians before 1880 moved to places which are entirely rural farming locations, the occupational choice of those who arrive later (even if they are randomly assigned) will be restricted to farming occupations.

I deal with the problems above in multiple ways. First, I pool 1910 and 1920 census data to enable me to control for a county fixed effect. This would control for any labor market characteristics of counties which are fixed over time and that would determine who chooses to come to counties with bigger enclaves as well as the set of job opportunities available for residents. To better account for differences in county occupational structure, I then assign enclave size at the enumeration district level and include county times year fixed effects. Now the assumption is that labor markets are relatively homogenous within counties but that Norwegians cluster within counties randomly. Next, I add immigrants from other countries, assigning their own enclave size, to more precisely estimate a county fixed effect.

Finally, I use my matched samples to construct an instrument for Norwegian enclave size, based on the destination county of an early wave of immigrants from each later migrant's municipality of birth. This instrument allows me to control for possible selection of low-skilled immigrants into enclaves. The instrument is constructed as follows:

1. Using the matched sample from 1865-1880, I identify the municipality of birth of Norwegian immigrants living in the United States in 1880.

2. Using the matched samples from 1900-1910 and 1900-1920, I identify the municipality of birth of immigrants who arrived after 1900 and who are living in the United States in 1910 or 1920.
3. I construct the instrument as a weighted average of the enclave size of counties of the initial immigrants from an immigrant's municipality of birth. For example, if 50% of pre-1880 migrants from Bergen moved to Houston County, Minnesota which had an enclave size of 8% in 1920 and 50% of migrants moved to Dane County, Wisconsin with an enclave size of 4% in 1920, then an immigrant from Bergen who moved after 1900 would receive a value of 6% (0.5*8 + 0.5*4) for his instrument in 1920. More formally, for individual i from municipality m, I weight the 1910 or 1920 county enclave sizes based on the initial settlement patterns of pre-1880 immigrants in the following way:

$$IV_{im} = \sum_{c} s_{cm} \% Norwegian_{c}$$

where s_{cm} is the share of Norwegian immigrants from municipality m living in county c in 1880.
4. I then use these values of the instrument to predict the actual enclave size of immigrants in 1910 and 1920.

The instrument must be both relevant and exogenous. I show in Table 4 that it predicts actual enclave size well (F>23). The identifying assumption for exogeneity is that any labor market conditions that drew immigrants to certain counties in 1880 are not also drawing immigrants to these counties in 1910 and 1920, except through their effects on enclave size. This assumption may not be tenable if a factor such as land suitability draws migrants to certain counties consistently over time; however, I note that farming as an occupation had become much less

profitable by 1920 than it was in 1880—it is likely that Norwegian migrants would not have entered farming as often as they did if they had not been drawn by previously formed enclaves.

V. Results

In Table 3, I begin by estimating equation (1) using the full population of Norwegian-born men living in the United States in 1910 and 1920 and who arrived after 1900. I regress the log of 1901 Cost of Living survey income on Norwegian enclave size. I add age fixed effects and indicators for six year of arrival bins. In Column (1), I start with state fixed effects and then add county fixed effects in the remaining columns. A one percentage point increase in the proportion of a county that is foreign-born is associated with 1.1-3 percent lower earnings. When measuring enclave size at the enumeration district level, rather than the county level, the effect is smaller at 0.4 percent lower occupational earnings. The average Norwegian lives with about 4 percent other Norwegians, and a standard deviation of this measure is about 3 percentage points, so a one standard deviation change in enclave size would have a large effect of up to a 10 percent reduction in occupational earnings.

Identification in the county fixed effects specification comes from changes in enclave size over time within counties; if the selection of immigrants is constant over time with respect to county characteristics, then this will control for potential selection into enclaves. In the final column, I include immigrants from 15 other European countries, interacting enclave size of each country with an indicator for being from that country. The coefficient for Norwegians is in the same range of above at -2.1 percent. The benefit of using other countries as a comparison is that I now control for anything happening in the counties which draws all groups of immigrants; for example, positive labor market shocks may draw all immigrants to certain counties in each year.

In Table 4, I turn to the matched sample. I start with OLS and then county fixed effects to replicate the results from the unmatched data. I find similar numbers to Table 3—one percentage point increase in enclave size reduces occupational earnings by 2-3 percent. In Column (3), I instrument for enclave size using the strategy outlined in the previous section. The first stage is strong with a coefficient of 0.51 and F-statistic of 84.5. I take this to mean that Norwegian settlement patterns in 1880 well explain settlement patterns by 1910 and 1920. The IV coefficient is larger in magnitude at -3.9 percent. In Column (4), I replicate the IV regression after reweighting the matched sample to be representative of the population, finding almost an identical number.

One drawback of the IV strategy is that I cannot include a county fixed effect in the regression since there is not enough variation in the predicted values across years. An ideal specification would include county fixed effects to control for labor market differences across different enclaves, and also instrument to control for possible individual selection into enclaves. I approximate this the best I can by controlling for county-level covariates in Columns (5) and (6). Specifically, I control for the share urban in the county as well as the share of male employment in the ten occupational categories designated by the first digit of occ1950.[7] Adding these controls weakens the first stage coefficient to 0.293 but it remains statistically significant. The IV coefficient remains -4 percent but is statistically significant only at the 10 percent level. I conclude that labor market differences do not fully account for the lower occupational earnings in larger enclaves.

Table 5 looks at other outcomes, including marriage and occupational categories of farmer, farm laborer, unskilled (farm laborer or laborer), and white collar. I show OLS coefficients in the first row and IV coefficients in the second row. The IV coefficients are almost all

[7] These categories are Professional (0), Farmers (1), Managers, Officials and Proprietors (2), Clerical (3), Sales workers (4), Craftsmen (5), Operatives (6), Service Workers (7), Farm Laborers (8), and Laborers (9).

larger in magnitude than OLS, but they all have the same signs. Larger enclaves increase the probability of being married and being a farmer, farm laborer or unskilled worker, and reduce the probability of working in a white collar occupation. This is all consistent with larger enclaves being more rural farming-oriented communities.

In Table 6 and 7, I look at the robustness of results to the occupational income score and matching method, respectively. In Table 6, patterns are generally consistent across occupation scores, with IV coefficients being the largest. The OLS coefficients range from -1.8 to -4.3 percent while the IV coefficients range from -3.9 to -8.1. The comparison between OLS and county fixed effects estimates do not have any consistent pattern, but they are all negative and significant.

In Table 7, I show the OLS, county fixed effects, and IV estimates as well as the first stage coefficients. Sample sizes vary across the matching methods, with the smallest being when matching by raw names instead of standardized names. This is likely because US enumerators did not use Norwegian-specific spelling but did spell names correctly phonetically (e.g. Eriksen and Eriksson would not match with raw names but would with NYSIIS). All of the first stages are strong with F greater than 40. The coefficients for all three regression specifications are almost identical to those from the primary matched sample.

VI. Enclaves and Second Generation Outcomes

Finally, I look to see whether the occupational earnings disadvantage of fathers is passed on to the next generation. I construct a matched sample of 3 to 15 year old United States born sons who had Norwegian-born fathers in 1920. I follow this group forwards to 1940 to assess their wage and education outcomes when they are 23-35 years old. For a comparison sample, I

also match men born to parents from the fifteen other countries from Abramitzky, Boustan and Eriksson (2014).

The 1940 census was the first census to collect wage earnings as well as years of education, so I have a richer set of variables to consider than simply occupational income. However, the main drawback of the wage income in this census year is that earnings are not reported for self-employed individuals, about 20 percent of my matched sample. My main outcome of interest is the log of wage earnings. I also consider occscore, years of education, urban residence in 1940, whether the child is a farmer, and whether the child moved out of his state of birth by 1940.

In Table 8, I first regress the log of wage earnings in 1940 on the percent Norwegian in the man's county of residence in 1920 during childhood. I find a coefficient similar to what I found for the first generation: a one percentage point increase in enclave size during childhood results in 3 percent lower earnings as an adult in 1940. Obviously, this regression does not pick up any causal effects—one reason to find a negative relationship might be that fathers were negatively selected into larger Norwegian enclaves and this was passed on to their child. Therefore, I next include the log of father's occupational score in 1920. Now, the regression compares men who grew up in different sized enclaves but whose fathers had roughly the same occupation during childhood. This reduces the size of the coefficient to 2.2 percent.

Next, I add men born to parents from fifteen other European countries. I interact enclave size (percent of the population from the father's country in the 1920 county) with an indicator for each country. I report the coefficient for children of Norwegian fathers. In Column (3), the coefficient remains large and negative at 1.9 percent. The benefit of adding these countries, however, is to be able to include county fixed effects to control for different labor market structures across

counties. Therefore, in Column (4), I control for 1920 county fixed effects and see that the coefficient shrinks to -0.5 percent. Including the log of father's occscore in Column (6) does little to change the coefficient.

In Table 9, I look at other outcomes of sons of Norwegian fathers. I follow the specification in Column (6) of Table 8, including all countries, county fixed effects and the log of father's 1920 occscore. I see that children growing up in larger enclaves are more likely to be self-employed and therefore not report wage income. This is likely because they are more likely to be farmers and less likely to be living in an urban area. The occscore in 1940 is only slightly smaller for men growing up in larger enclaves. This is somewhat surprising since most of those missing wage earnings are farmers, a somewhat low status occupation by the occscore measure.

I conclude that growing up in a larger ethnic enclave has negative effects on the second generation. This seems to mostly come through worse labor market opportunities in these counties, coupled with a lower propensity to move away towards other opportunities as an adult.

VII. Discussion and Conclusions

This paper constructs a new linked dataset of Norwegian immigrants during the Age of Mass Migration. Using an instrumental variable based on chain migration, I find that larger ethnic enclaves in the United States were associated with lower occupational earnings; one standard deviation increase in enclave size reduces occupational earnings by up to 10 percent. This penalty is partly passed on to the second generation.

These findings add to our understanding of the consequences of ethnic clustering in a time period with high levels of immigration that have not been reached in the United States until

recently. During a period with no controls on immigration, it appears that geographical clustering did not have positive effects on the immigrants themselves.

I look only at immigrants arriving after 1900. It is possible that ethnic enclaves were useful for Norwegian economic success in earlier periods, or that the older generation of immigrants living in receiving places in the United States benefited from low-skilled labor. Future research should consider the effects of initial settlement patterns of groups of immigrants on the long-run performance of these groups. For example, anecdotal evidence suggests that Norwegian communities invested heavily in schools and social capital. This may not have happened if the initial immigrants had not been as geographically clustered.

References

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

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

Aizer, Anna and Janet Currie. 2004. "Networks or neighborhoods? Correlations in the use of publicly-funded maternity care in California." *Journal of Public Economics* 88.

Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey. 2017. "How Do Automated Linking Methods Perform in Historical Samples? Evidence from New Ground Truth". NBER Working Paper #24019.

Battisti, Michele, Giovanni Peri, and Agnese Romiti. 2016. "Dynamic Effects of Co-Ethnic Networks on Immigrants' Economic Success". NBER Working Paper #22389.

Beaman, Lori. 2012. "Social Networks and the Dynamics of Labor Market Outcomes: Evidence from Refugees Resettled in the U.S." *Review of Economic Studies.*

Bertrand, Marianne, Erzo F. P. Sendhil Mullainathan. 2000. "Network effects and welfare cultures." *Quarterly Journal of Economics* 107.

Damm, Anna Piil. 2009. "Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence." *Journal of Labor Economics* 27(2).

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

Eriksson, Katherine and Zachary Ward. 2018. "The Ethnic Segregation of Immigrants from 1850 to 1940". Working Paper

Goldin, Claudia and Bob Margo. 1992. "The Great Compression: The Wage Structure in the United States at Mid-Century". *Quarterly Journal of Economics* 107(1).

Inwood, Kris, Chris Minns, and Fraser Summerfield. 2018. "How much do Occupational Indices Miss the Mark? Evidence from the Canadian Census". Working Paper

Saavedra, Martin, and Tate Twinam. 2018. "A Machine-Learning Approach to Improving Occupation Income Scores". Working Paper

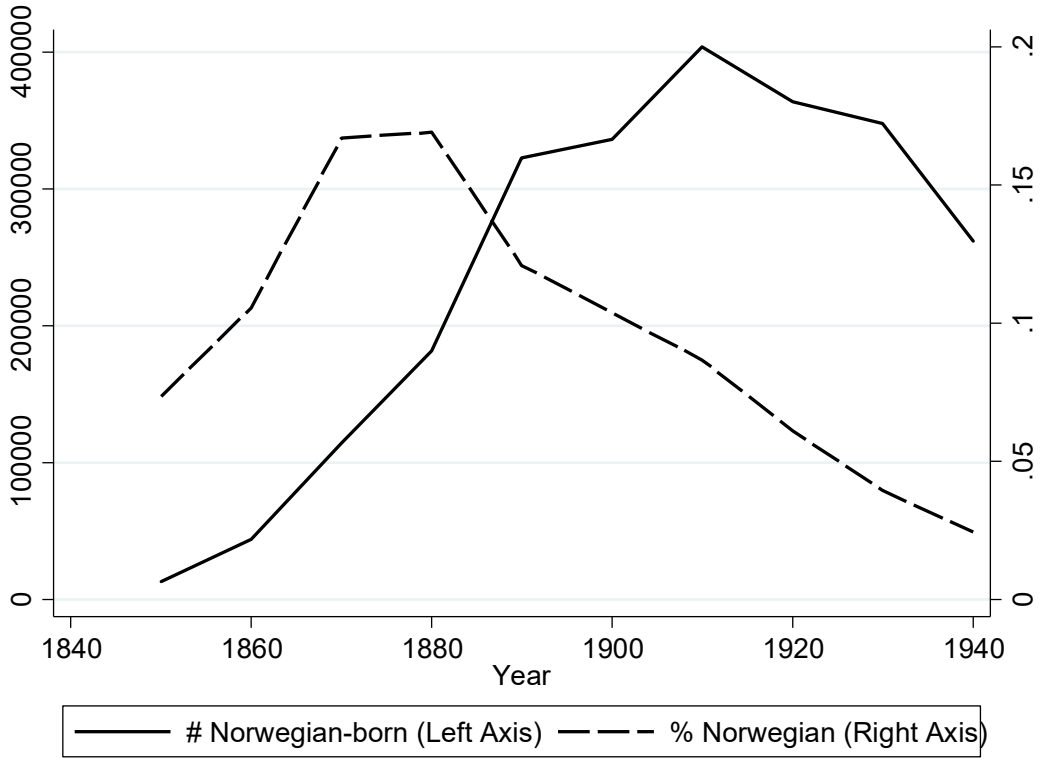
Semmingsen, Ingrid. 1980. *Norway to America: A History of the Migration*. University of Minnesota Press

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. *Integrated Public Use Microdata Series: Version 7.0* [dataset]. Minneapolis: University of Minnesota, 2017. https://doi.org/10.18128/D010.V7.0.

Xie, Yu and Margaret Gough. 2011. "Ethnic Enclaves and the Earnings of Immigrants". *Demography* 48(4).

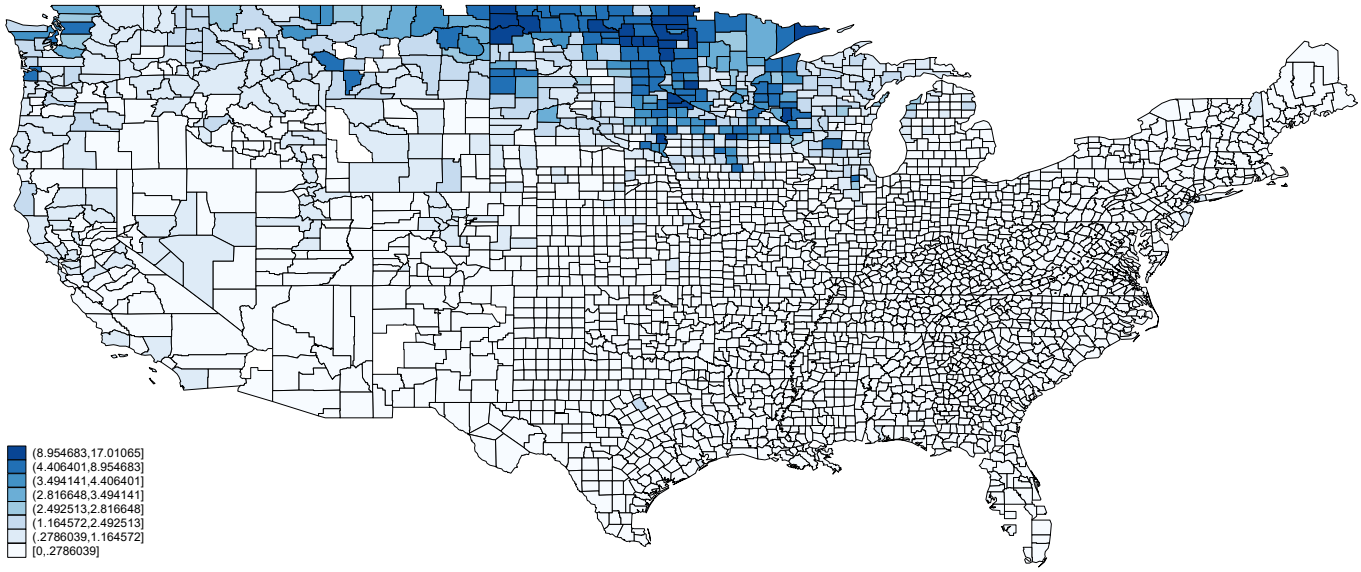
Figures and Tables

Figure 1: Norwegian Born Population and Average Norwegian Enclave Size in the United States, 1850-1940



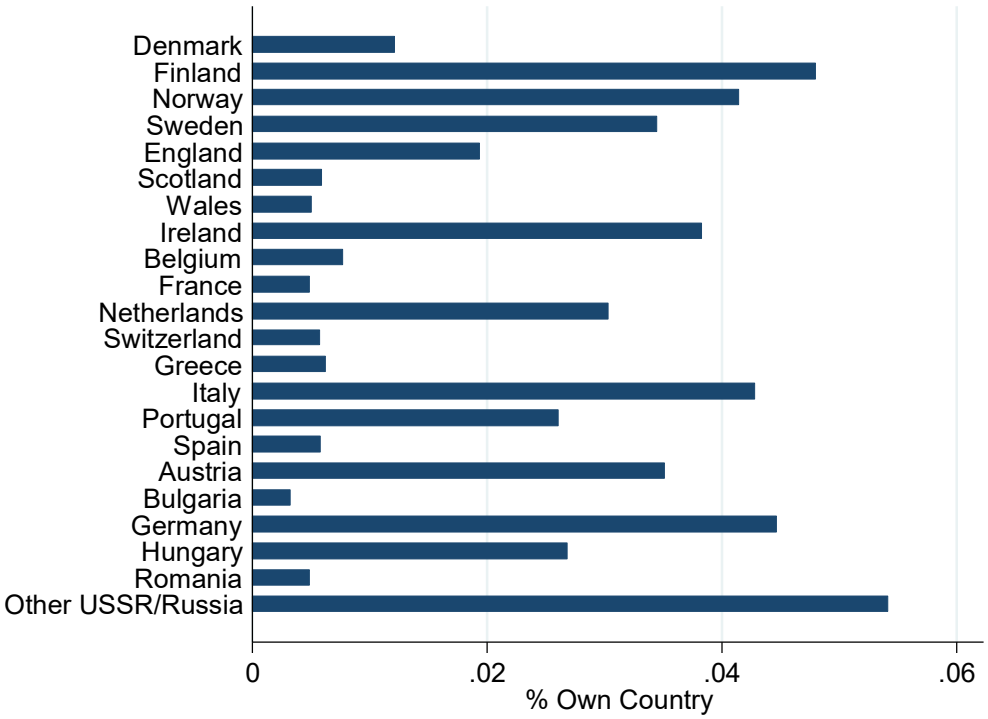
Notes: Figure plots Norwegian-born population size (left-axis) in the United States and average Norwegian enclave size by decade using IPUMS (Ruggles et al 2018). Enclave size is measured as the percent of a county's residence that are Norwegian. The average size is weighted by Norwegian populations to represent the experience of the average Norwegian immigrant in the given year.

Figure 2: Norwegian enclaves, 1920



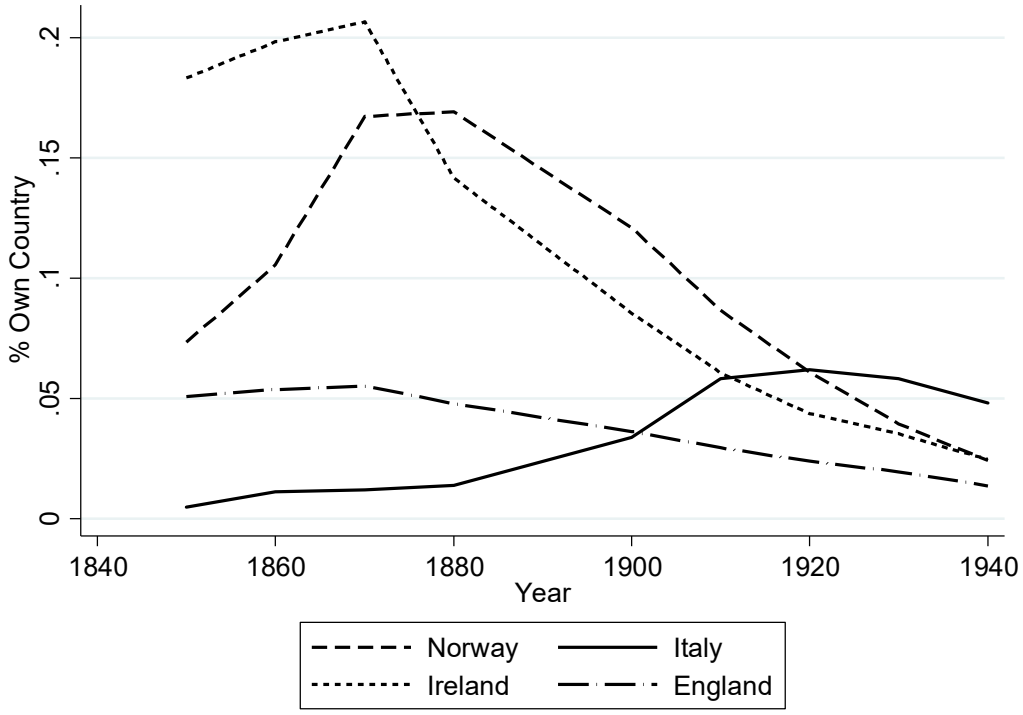
Notes: Figure maps the percent Norwegian-born by county in 1920, using the full count census from IPUMS. The largest bin ranges from 8.2 to 17 percent. The enclave size variable has a standard deviation of 3.34.

Figure 3: Enclaves by country, 1920



Notes: Figure plots average enclave size by sending country using IPUMS (Ruggles et al 2018). Enclave size is measured as the percent of a county's residence that are from each country. The average size is weighted by population to represent the experience of the average immigrant of each group.

Figure 4: Enclave size over time for selected countries, 1850-1940



Notes: Figure plots average enclave size over time for four sending countries. Data comes from IPUMS. Enclave size is measured as the percentage of a county born in the selected country. Means are weighted by population to represent the experience of the average immigrant in each group.

Table 1: Sample Sizes and Match Rates

Sample	Population, US	Matched Sample	Match Rate
1865-1880	89,125	7,593	8.52
1900-1910	66,415	9,323	14.03
1900-1920	74,315	12,002	16.15

Notes: Matched samples match from Norwegian men living in the United States in 1880, 1910, or 1920 backwards to Norway in 1865 or 1900. Sample restricts to men 21 to 65 in the later year. Match conducted based on year of birth plus or minus two years plus exact match on standardized (NYSIIS) first and last name. Matches from 1910 and 1920 to 1900 restrict to men who arrived in the United States after 1900.

Table 2: Summary Statistics in Matched Sample and Population

	(2) Population	(3) Difference: Match - Pop	(4) Difference, Re-weighted
% Norwegian, county	4.064	0.212***	-0.020
		(0.043)	(0.033)
% Norwegian, E.D.	7.145	0.408***	-0.020
		(0.010)	(0.053)
Age	32.11	0.620***	0.045
		(0.067)	(0.065)
Income	570.80	17.00***	0.508
		(1.692)	(1.674)
Occscore	21.29	0.275***	0.173*
		(0.079)	(0.090)
Farmer	0.152	0.022***	0.002
		(0.003)	(0.003)
Laborer	0.274	-0.022***	0.002
		(0.004)	(0.004)
Married	0.547	-0.051***	-0.004
		(0.004)	(0.004)

Notes: N = 19,745/133,456 (matched/population). Column (1) presents means from the Norwegian-born population aged 21 to 65 living in the United States in 1910 and 1920. Column (2) shows the coefficient from regression the variable on an indicator for being in the matched sample. Column (3) reweights Column (2) using an inverse probability weight procedure which uses a probit regression to predict matched status based on the variables in the table, plus a polynomial of order four in the continuous variables.

Table 3: Effects of Norwegian Enclave Size on Income, Unmatched Data
Outcome = Log of Occupational Income, 1901 Cost of Living Survey

	(1)	(2)	(3)	(4)
% Norwegian	-0.011***	-0.030***	-0.004***	-0.021***
	(0.002)	(0.004)	(0.001)	(0.002)
Enclave	County	County	ED	County
Mean % Nor.	4.145	4.145	7.145	4.145
Countries	Norway	Norway	Norway	All
Fixed Effects	State	County	County	County
R^2	0.094	0.170	0.184	0.165
N	132,164	132,164	132,164	181,390

Notes: Table shows coefficient on %*Norwegian* from Equation (1) in text. Regressions use unmatched Full Count data from 1910 and 1920 for Norwegian men, coupled with IPUMS 1% samples from 1910 and 1920 for other sending countries. Robust standard errors in parentheses are clustered by county. Regressions control for age fixed effects and indicators for six year of arrival bins. Ages are restricted to 21 to 65.

Table 4: Effects of Norwegian Enclave Size on Income, Matched Data
Outcome = Log of Occupational Income, 1901 Cost of Living Survey

Sample	Full			Full, IPW	Full, include county characteristics	
Method	OLS	County FE	IV	IV	OLS	IV
% Norwegian	-0.020***	-0.029***	-0.039**	-0.037**	-0.006*	-0.040*
	(0.003)	(0.007)	(0.010)	(0.010)	(0.001)	(0.022)
1st stage coeff.			0.511***	0.499***		0.293***
1st stage F			84.45	78.53		23.78
Mean % Norw.	4.145	4.145	4.145	4.145	4.145	4.145
R^2	0.145	0.254	0.110	0.099	0.185	0.254
N	17,552	17,552	17,552	17,552	17,752	17,752

Notes: Regressions use matched data and control for age dummies plus indicators for six year of arrival bins. Age is restricted to 21 to 65. Columns (5) and (6) include the share urban within a county and the share of male employment in the ten occupational categories designated by the first digit of occ1950. Standard errors are clustered by county.

Table 5: Effect of Norwegian Enclave Size on Other Outcomes

	Married	Farmer	Farm Laborer	Unskilled	White Collar
OLS	0.007***	0.025***	0.021***	0.021***	-0.013***
	(0.002)	(0.004)	(0.002)	(0.003)	(0.002)
IV	0.044***	0.017***	0.035***	0.057***	-0.034***
	(0.008)	(0.005)	(0.005)	(0.008)	(0.006)
N	17,613	17,613	17,613	17,613	17,613

Notes: Regressions use matched data and control for age dummies and six categories for year of arrival. Age is restricted to 21-65. Coefficients presented are the coefficients on % Norwegian, measured at the county level. Standard errors are clustered by county.

Table 6: Robustness of main results to occupation score

	OLS	County FE	IV
Income	-0.020***	-0.029***	-0.039***
N=17,552	(0.003)	(0.007)	(0.010)
Occscore	-0.043**	-0.008*	-0.081***
N=17,585	(0.004)	(0.005)	(0.001)
Lido	-0.028***	-0.017***	-0.043***
N=17,374	(0.002)	(0.006)	(0.007)
Income + farm adj.	-0.018***	-0.029***	-0.044***
N=17,552	(0.004)	(0.008)	(0.010)
Occscore + farm adj.	-0.034***	-0.027***	-0.072***
N=17,575	(0.005)	(0.009)	(0.012)

Notes: Regressions replicate Table 4 with different occupation score measures. Income refers to 1901 Cost of Living income. Occscore refers to the IPUMS occscore based on 1950 median occupational earnings. Lido refers to age/race/state-adjusted occscores from Saavedra and Twinam (2018). The last two rows replace farm income with county-level farm income estimated from the 1920 Census of Agriculture. All outcome variables are logged. Coefficients presented are from *%Norwegian* in Equation (1).

Table 7: Robustness to Matching

	Standard	5 year age band	Raw Names	+/- 1 year in age
OLS	-0.020***	-0.020***	-0.021***	-0.021***
	(0.003)	(0.003)	(0.003)	(0.003)
County FE	-0.029***	-0.020**	-0.036***	-0.029***
	(0.007)	(0.009)	(0.011)	(0.007)
IV	-0.039***	-0.038**	-0.039***	-0.041***
	(0.010)	(0.009)	(0.010)	(0.009)
1st stage coeff.	0.511***	0.581***	0.533***	0.429***
1st stage F	84.45	40.74	46.69	58.67
N	17,552	8,984	7,496	14,969

Notes: Regressions replicate Table 4 with different matched samples. Coefficients presented in first three rows are from *%Norwegian* in Equation (1). Column (1) reproduces the main results. Column (2) requires individuals to be unique within a five year age band in each census year. Column (3) matches based on raw names instead of using the NYSIIS standardization. Column (4) restricts men to report an age within one year of the other census.

Table 8: Effects of Norwegian Enclaves on Income of the Second Generation
Outcome = Log of wage income, 1940

	(1)	(2)	(3)	(4)	(5)
% Norwegian	-0.030***	-0.022***	-0.019***	-0.005***	-0.006***
	(0.006)	(0.004)	(0.001)	(0.001)	(0.001)
Ln(OccscoreDad)		0.452***			0.199***
		(0.026)			(0.003)
Countries	Norway	Norway	All	All	All
Fixed Effects	None	None	None	County	County
R^2	0.1259	0.1636	0.0962	0.1288	0.1342
N	18,864	18,864	468,101	468,101	468,101

Notes: Regressions use matched sample of sons of Norwegian-born fathers from 1920 to 1940. Enclave size is assigned based on childhood county of residence. Regressions control for age fixed effects as well as father year of arrival bins. Columns (3)-(5) interact enclave size with indicators for each father's country of birth and present the interaction for Norwegians.

Table 9: Effects of Norwegian Enclaves on Other Outcomes of the Second Generation

	(1) Occscore	(2) =1 if no Wage Inc.	(3) =1 if Urban	(4) Education	(5) Farmer	(6) Migrant
% Norwegian	-0.001*	0.003***	-0.004***	0.003	0.001***	-0.010***
	(0.000)	(0.001)	(0.001)	(0.004)	(0.000)	(0.001)
Countries	All	All	All	All	All	All
Fixed Effects	None	None	None	None	County	County
R^2	0.2150	0.1259	0.1636	0.0962	0.1288	0.1342
N	582,543	614,092	614,092	614,092	614,092	614,092

Notes: Sons of Norwegian immigrants aged 3-15 in 1920 are matched to 1940. Regressions control for age, birth state, and father's birth country fixed effects as well as log of father's occscore in 1920.