Title
Essays on Education and Immigration throughout the 20th Century

Author
Eriksson, Katherine Amelia

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on Education and Immigration throughout the 20th Century

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Economics

by

Katherine Amelia Eriksson

2013

ABSTRACT OF THE DISSERTATION

Essays on Education and Immigration throughout the 20th Century

by

Katherine Amelia Eriksson

Doctor of Philosophy in Economics

University of California, Los Angeles, 2013

Professor Leah Boustan, Chair

This dissertation is comprised of three chapters tied together under the broad umbrella of economic history. The first chapter examines the effect of access to schooling on black crime in this historic period. A large gap in incarceration rates between black and white men has been evident since the early 20th century. I use the construction of 5,000 new schools in the US south, funded by northern philanthropist Julius Rosenwald between 1913 and 1932, as a quasi-natural experiment which increased the educational attainment of southern black students. I match a sample of male prisoners and non-prisoners from the 1920-1940 Censuses backwards to their birth families in previous Census waves in order to assign exposure to Rosenwald schools based on birth cohort and childhood county of residence. I find that one year of access to a Rosenwald school decreased the probability of being a prisoner by 0.04-0.10 percentage points (10-15

percent of the mean). Using other data from archival and government sources, I show that Rosenwald schools affected juvenile crime and all categories of adult crime. I investigate the channels through which Rosenwald schools reduced crime, including educational attainment, school quality, and migration responses. Finally, I consider heterogeneous effects by household background and county of birth socio-economic characteristics.

The second chapter examines immigrant assimilation in the early 20th century US. During the Age of Mass Migration (1850-1913), the US maintained an open border and absorbed 30 million European immigrants. Prior cross-sectional work on this era finds that immigrants held lower-paid occupations than natives upon first arrival but experienced rapid convergence. In newly-assembled panel data, we show that, in fact, immigrants did not face a substantial initial earnings penalty and experienced occupational advancement at the same rate as natives. Cross-sectional patterns are driven by biases from declining arrival cohort quality and departures of negatively-selected return migrants. We show that these findings vary substantially across sending countries and explore potential mechanisms.

The third chapter uses an exogenous change in the language of instruction in South African schools in 1955 to examine the effect of mother-tongue versus "market" language instruction on long-term educational and economic outcomes. Using the 1980 South African census, a difference-in-difference framework allows me to estimate the effect of increasing mother-tongue instruction for black students from four to six years. I find small positive effects on wages which I interpret as evidence of increases in human capital; these effects might have been larger in the absence of labor market discrimination against blacks under *apartheid.* I find positive effects on the ability to read and write, negative effects on the ability to speak English and Afrikaans, and positive effects on educational attainment. I examine heterogeneous effects

by region. This paper is relevant to language policy in post-colonial countries as well as Spanish speaking areas of the United States.

The Dissertation of Katherine Amelia Eriksson is approved.

Dora Costa

Christian Dippel

W. Walker Hanlon

Leah Boustan, Chair

University of California, Los Angeles

2013

This dissertation is dedicated to my parents: my mother who knew I was an economist at age 3, and my father who has supported every dream of mine over the years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I acknowledge and appreciate financial support from the Center for Economic History at UCLA and the California Center for Population Research.

Chapter 2 is a co-authored work with Leah Boustan (UCLA) and Ran Abramitzky (Stanford). We gratefully acknowledge financial support from the National Science Foundation, Center for Economic History at UCLA and the California Center for Population Research.

VITA

Prior Education

MA, Economics, UCLA, December 2008

MS, Agricultural Economics, Virginia Polytechnic Institute and State University (Virginia Tech), August 2007. *Thesis*: "Exploring changes in poverty in Zimbabwe between 1995 and 2001 using parametric and nonparametric quantile regression decomposition techniques"

BA, Philosophy, Politics, and Economics (PPE), Somerville College, University of Oxford, July 2006
First Class Honors

BS, Mathematics and Philosophy (double major), Virginia Tech, May 2004

Awards

UCLA Elise Burman fellowship, 2012-2013

UCLA Center for Economic History fellowship, fall 2012

UCLA Center for Economic History research grant, 2012

UCLA Departmental Progress Award, 2008

California Center for Population Research trainee fellowship, 2007

Somerville College award for First Class honors degree, July 2006

Virginia Tech Philosophy Department book award for best graduating senior, May 2004

Phi Beta Kappa member, 2004

Virginia Tech Math Department scholarship, 2003, 2004

Virginia Tech Dean's scholarship, 2003

Chapter 1: Access to Schooling and the Black-White Crime Gap in the Early 20th Century US South: Evidence from Rosenwald Schools

1.1 Introduction

In the contemporary United States, black men are disproportionately more likely than white men to be arrested for crime and to be incarcerated. This racial gap in incarceration has been apparent in the data since the late nineteenth century. In 1910, blacks made up only thirteen percent of the population but twenty-seven percent of those in prison.[1] In the South in 1910, blacks were 30 percent of the population yet were 60 percent of those incarcerated (US Census Report 1910).[2] This gap could have many causes, including discrimination, differences in family background, lack of job opportunities for blacks, higher urbanization rates of blacks, and differences in education.

This paper explores the role of one of these factors – disparities in education – in explaining the historical roots of the racial gap in crime. In particular, I analyze the relationship between educational attainment and the probability of incarceration between 1920 and 1940 in the southern United States. I use the construction of almost 5,000 schools in 14 southern states for rural black students between 1913 and 1931, sponsored in part by northern philanthropist Julius Rosenwald, as a quasi-natural experiment which increased the educational attainment and literacy of blacks born in the South.

[1] In 2009, this disparity was even larger. Blacks made up 13.9 percent of the population and 39.4 percent of those incarcerated. This is likely due to the emergence of a series of factors which have led to higher black incarceration rates such as stricter penalties for drug offenses.

[2] In the North, blacks were 1.8 percent of the population but 9.6 percent of those incarcerated.

I build a new dataset that matches the full universe of southern-born male prisoners from the 1920-1940 US Censuses, along with a comparison sample of non-prisoners, to their birth families in a previous census earlier in life. Based on their location in childhood, I assign to these men their likely exposure to a Rosenwald school, defined as the number of years between the ages of seven and thirteen that a Rosenwald school was present in his county. I also construct an alternative measure of exposure, which is weighted by the probability that any child could have obtained a seat in the school in each year. That is, if a child had a school in his county for one year, but it could only serve half of the students, he would be assigned one half of a year by this measure.

I find that education reduces crime later in life among adults. Exposure to a Rosenwald school for one year during childhood reduces the probability of committing a crime for blacks by between 0.05 and 0.1 percentage points or 10-15 percent of the mean incarceration rate for blacks. This finding is robust to different comparison groups and alternative specifications. I find that most of the effect is among students whose county received a school when they were between seven and nine rather than between ages ten and thirteen.[3]

I use county fixed effects in my estimation and utilize white men as a control group. Identification of the effect of Rosenwald schools on crime depends on variation in school construction across cohorts and races within a given county. Potential threats to identification can arise from characteristics of counties that change over time in a way that was correlated with both the construction of new schools and with subsequent criminality. I address this concern by constructing a different measure of exposure which is based on the distance of a household from

[3] This fits with a pattern of early childhood interventions being more important than ones in later childhood (Cunha, Heckman, Lochner, and Masterov 2006). Alternatively, it could be that children whose counties did not receive a school until later in childhood "missed the boat".

a school within a county; this design compares children within a certain distance from the school to those who did not grow up close to a school but who grew up in the same county.

A second concern might be that parental background differs among those who have exposure to schools and those who do not; for example, more literate parents could be migrating into certain counties as Rosenwald schools are being built. I provide evidence that this is likely not true but I also present estimates which compare brothers within the same family to control for family background differences.

Within North Carolina, I collect by hand more information about my census sample; I collect literacy, home ownership, and occupation of the child's father as well as the literacy of prisoners born in North Carolina.[4] The first lets me look at how effects differ by household socio-economic characteristics while the second allows me to determine whether one channel through which Rosenwald schools affect crime is through literacy. I do find that literacy is an important channel through which crime decreases, but that the returns to school also increased with exposure to Rosenwald schools. I also find differing effects by household background; schools are more important and have a larger effect if households are poorer.

Finally, I collect data from non-census sources in North Carolina to look at juvenile crime and categories of adult crime. I find that juvenile court cases decrease for black children in counties with exposure to Rosenwald schools. One year of exposure reduces juvenile court cases by about 15 percent of the mean. I collected archival data from the state prison registers in the North Carolina State Prison which include information about type of crime committed. Within a multinomial logit framework, I show that Rosenwald schools reduce all categories of crime and

[4] Literacy and level of schooling are not available in the census indices which I use in my primary analysis. I collect literacy by hand for my prisoner sample in North Carolina in 1920 and 1930, and collect years of school in 1940.

that the effect is not restricted to the types of "nuisance" crimes for which blacks were regularly imprisoned in the Jim Crow South (Naidu 2010).

This paper contributes to two literatures. The first concerns the convergence in wages and other outcomes between blacks and whites over the 20th century. The educational attainment of three years on average between blacks and whites at the beginning of the 20th century, which was caused by poor investments in black schools and the legacy of slavery, declined between 1910 and 1940 (Margo 1990, Aaronson and Mazumder 2011). These improvements in black educational attainment have contributed to a decline in the black-white wage gap (Heckman, Lyons and Todd 2000, Smith and Welch 1977). My estimates suggest that the black-white incarceration gap should have declined by fifty percent over the 20th century due to these relative increases in black educational attainment. The fact that black incarceration rates have not only remained persistently high but also have increased further since the mid-1970s, suggests that other factors have counteracted the forces of educational convergence.

This paper also speaks to work on the social returns to education. Other research has shown that education contributes to improvements in health, more targeted fertility, and increases in voting and civic behavior (Lleras-Muney 2005, Clark and Royer 2009, Aaronson et al 2012, Milligan et al 2004). One of the social returns to education is a significant reduction in criminality. The relationship between education and crime has been extensively studied in a contemporary context, with the most relevant being Lochner and Moretti (2004) who find in a contemporary context that reductions in the probability of incarceration for blacks and whites following an additional year of school.[5] Lochner and Moretti, as well as multiple other papers,

[5] Lochner and Moretti (2004) is the best known study; their results have been replicated and expanded in Sweden, the UK, and other European counties (Hjarlmarsson et. al 2010, Machin et. al 2011, Meghir et al. 2011). Other work has looked at the relationship between school quality and crime (Deming 2012) and finds a significant effect. Anderson (2012) shows that juvenile crime decreases with higher minimum dropout ages for students.

focus primarily on secondary schooling, using variation induced by compulsory schooling laws, whereas I focus on elementary school. If there are decreasing returns to education, we might expect results to be stronger at these younger ages. I find that one year of school reduces the probability of being incarcerated by 0.25-0.5 percentage points which is similar to Lochner and Moretti who find a social return to school of 0.4 percentage points. Note, however, that the mean incarceration rate in contemporary data is 3.6 percent whereas in my historical context it is only 0.8 percent. Additionally, papers looking at other social returns to school have found that the institutional context matters; I focus on a poorer, highly unequal society and so my results are potentially more applicable to developing countries today.[6]

Crime has a variety of social costs. In 2007, the Department of Justice estimated that victims of crimes face $15 billion in losses and that the prevention and punishment of crimes cost $179 billion for local, state and federal governments. Committing crimes also has long-run costs for the criminal; individuals with a criminal record often find it difficult to get a job, and those jobs they do get likely pay lower wages (Western, Kling, and Weiman 2001). These consequences are concentrated in black communities; a black male's lifetime risk of being murdered is 1 in 21 as opposed to 1 in 131 for white males (Sampson and Wilson 2005). The concentration of the consequences of crime in black areas coupled with the higher offender rates among blacks results in crime contributing to black-white economic inequality.

The structure of this paper is as follows. Section II provides historical background about black/white differences in crime and in schooling through the 20th century. In section III, I describe the data and exposure to Rosenwald schools. Section IV discusses my identification strategy and potential threats to identification. Section V presents results from my primary

[6] The gap between white and black schooling and white and black crime rates in the early 20th century US south is very similar to those rates today in South Africa; South Africa's recent history of *apartheid* resulted in similar differences in school resources available to blacks and whites as recently as 1994.

sample and additional data. Section VI examines robustness and alternative specifications. I conclude in Section VII.

1.2 Historical Background

A. Incarceration rates by race and region over time

The gap between white and black incarceration rates was evident as early as 1890. In 1890, blacks were 3.2 times more likely to be incarcerated than whites (Petersilia and Reitz 2012).[7] Contemporary observers attributed this difference in crime rates to a variety of factors: poverty, discrimination, lack of education, "moral degradation", etc (Sellin 1900). Figure 1 shows the evolution of black and white state and federal prison incarceration rates from 1926 to 1986 (Department of Justice, 1986). Panel A reveals that incarceration rates were higher for blacks than for whites in all years, with approximately eight per 10,000 blacks incarcerated nationally through the early 1960s compared to four whites. The racial gap is present in both the North and the South, despite higher incarceration rates in the North in the 1920s and 1930s. These graphs demonstrate that the current black-white incarceration gap is not a recent phenomenon but, rather, has been present throughout the 20th century. Any explanation of differences in crime rates needs to take into account the historical patterns of incarceration.

The debate about whether, and to what extent, education reduces crime goes back to the early 20th century and was in fact a central topic of concern at that time. Scholars such as John Roach Stratton (1900) argued that the "race problem," i.e. the high crime rates and "immorality" of blacks, could not be solved by education. Stratton thought that the positive correlation between increasing black crime and increases in black education between the end of the Civil War and 1900 showed that education actually increased criminality. He argued that allowing

[7] In 2010, this number was 6.4. It grew from 3.2 in 1890 to 4.3 in 1923, and was as high as 7.1 in 2000.

blacks to gain education and move from farms to cities to find work increased crime at no real benefit to anyone. In fact, Governor Vardaman of Mississippi used this reasoning when restricting funds for black schools in 1904 (Hollandworth 2008). On the other side of the argument, Booker T. Washington and W. E. B. Du Bois sought to explain higher black criminality with low wages and discrimination. The issue of high black incarceration rates was one motivation for Booker T. Washington's interest in improving black schools, out of which grew the Rosenwald Initiative.

B. Black schooling and Rosenwald schools

Blacks in the birth cohorts of 1880 and 1910 completed on average three fewer years of education than whites. Motivated by his concerns about the low funding levels for black education, Booker T. Washington, principal of the Tuskegee Institute in Alabama, reached out to Northern philanthropist and businessman Julius Rosenwald.[8] Rosenwald agreed to fund a pilot program supporting the construction of six black schools in 1913-14, with the promise of up to 100 more. The original schools were built primarily in Alabama; by 1920, the program supported 716 schools in eleven southern states. The Rosenwald Fund explicitly targeted rural students.

Rosenwald believed that communities must "buy-in" to, or invest in, any endeavors. This view, coupled with Washington's belief in black self-reliance, led to the use of a matching grant approach, whereby local communities had to raise anywhere from 75 up to 90 percent of the funds for a new school. After the schools were built, they were reliant on the local community and the state for funding. The early schools received about 25 percent in grant money, whereas this number fell to 10-15 percent by the later years of the program. The program ended in 1931

[8] The Rosenwald School Initiative was not the only black schooling initiative in this time. The Jeanes Fund provided teacher training. Other philanthropic interventions are described in Donohue, Heckman and Todd (2002).

with Rosenwald's death and the decreased value of Fund assets after the collapse of the stock market. In Section IV, I analyze the location and timing of school construction. In addition to building 4,983 schools by 1931, the Fund also provided some money for teacher training schools, homes, and shops.

By the end of the program, 76 percent of counties in the 14 southern states had a school and 92 percent of black students in these states lived in a county with a school. Aaronson and Mazumder (2011) estimate that the schools could serve 36 percent of rural black students. They show that the Rosenwald schools were a significant contributor to the narrowing of the black-white schooling gap by 1940. Black cohorts born through 1909 received three fewer years of school than white cohorts. This gap narrowed to half a year for cohorts born after 1940. The gap remained constant for cohorts born through the 1950s. Aaronson and Mazumder estimate that Rosenwald schools increased school attendance by about 5 percentage points. They calculate that through this increase in enrollment, Rosenwald schools account for 40 percent of the reduction in the racial schooling gap.

1.3 Data

A. Primary sample

My measure of crime is the propensity of men who grew up in a given county or a town to be incarcerated later in life.[9] I calculate this measure from US Census data for the years 1920-1940. To do so, I assemble a sample of prisoners and non-prisoners in each relevant census. Prisoners are defined as men reporting a relationship to household head of "Prisoner" in the full

[9] The FBI Uniform Crime Statistics do not become available for a substantial number of counties until the 1960's. Crime statistics are only available from census reports or for major cities prior to the start of the FBI UCR. This is the first paper to collect individual data by race and for the full country. Moehling and Piehl (2011) collect individual data for immigrants living in select states in a similar time period.

census indexes from FamilySearch.org. A robustness check also includes men whose relationship to the household head is listed as "Inmate."[10] Traditionally, measures of crime do not include inmates of county jails, so in my analysis for North Carolina, I restrict to those incarcerated in State or Federal Prisons.[11] Non-prisoners are found either in the IPUMS 1% samples of 1920 and 1930 or from the full index of the 1940 Census, which has only recently been made publically available. I restrict the sample to men between ages 18 and 35 who were born in the South.

I identify the town or county in which sample individuals grew up by matching all men to the relevant census one or two decades earlier to find the individual living in their birth family. Men less than 26 years old are matched to the previous census while those between 27 and 35 years old are matched over a twenty year period.[12] To match individuals, I follow the procedure pioneered by Ferrie (1996) and used in Abramitzky, Boustan, and Eriksson (2012). I first standardize first and last names using the NYSIIS algorithm (Atack and Bateman 1992) which spells names with the same phonetic sound identically. Individuals are then matched by first name, last name, state of birth, race, and age across Census waves. I allow individuals to misreport their age by up to two years in either direction. Inherent in any matching procedure is a trade-off between sample size and accuracy. A second sample prioritizes accuracy by requiring

[10] "Inmate" can refer to inmates of prisons but also can refer to inmates of mental institutions or other institutionalized group quarters (e.g. orphanages). To minimize the probability of falsely categorizing an individual as a criminal, I do not use inmates in my primary sample.

[11] The DOJ omits local jails and county detention centers from most statistics about crime, arguing that since most people are only in local jails for a few days to a few months, counting inmates at any one point in time will understate the number incarcerated in a given year. Furthermore, many inmates of jails are there for minor crimes (e.g. vagrancy) as opposed to more serious crimes warranting incarceration in state prisons.

[12] This cutoff was determined by calculating the age at which approximately 90% of individuals appear to be living at home in the 1910 census. The percentage living at home at young ages increases with later censuses. In 1910, 88.6% of black children and 92.4% of white children report a relationship to household head of "child", "grand-child", "sibling", or "other relative". Results are robust to using a cutoff of 14 or 18 years old between the two matched samples.

individuals to be unique by name/birthplace/race within a 5 year age band. Accuracy may also come at the cost of representativeness if the uniqueness of an individual's name is correlated with his socio-economic status. The match rates in my study are consistent with the literature, averaging around 20 percent over a single Census decade and 19 percent over a 20 year period. Matching details and sample sizes are available in Appendix A.

In North Carolina, I collect additional information about my prisoners and non-prisoners. I collect the name of the prison by hand and restrict my dataset to consider only state and federal prisoners since inmates of local jails could be there for a number of reasons not related to criminality (e.g. vagrancy, etc). I also hand-collect literacy for the full sample of prisoners born in North Carolina. To allow for heterogeneous effects of the school-building program on crime, I collect household characteristics for the sample born in North Carolina. These variables include parents' literacy, home ownership, number of siblings, and the occupation of the household head.

Summary statistics are shown in Table 1. Black individuals are three times more likely than whites to be incarcerated in my sample (0.6 percent versus 0.2 percent). Prisoners tend to less literate, have less literate parents, and are more likely to have parents who are laborers rather than owner-occupier farmers. Finally, more prisoners are found outside of their state of birth than non-prisoners, but this pattern is mainly driven by migration of white prisoners.

B. Rosenwald schools

Information about the Rosenwald school program is taken from Aaronson and Mazumder (2012). The dataset of 4,983 schools was compiled from school-level index cards archived at Fisk University. Information available includes school name, county location, year of construction, and some information about funding sources and the size of the school. The earliest

Rosenwald schools were located in Alabama in 1913; by 1932, schools had been built in fifteen Southern states.[13]

The school-level cards do not, for the most part, contain information on the school's address within a county. For the North Carolina sub-sample of 775 schools, I used historical society documents and school directories (North Carolina Department of Education 1937, 1942) to determine the town in which the school was located. I was able to locate 720 schools. This procedure is limited to North Carolina because it is the only southern state with readily available school directories. One might worry that North Carolina is not representative of the south.[14] North Carolina was in fact one of the poorest states in 1910, with the lowest per capita income and low literacy rates. I expect that results might be stronger in this state than in other states.

C. Calculating Exposure to Rosenwald Schools

Exposure to a Rosenwald school varies across towns and counties in the South as well as by birth cohort. Children who were born before 1906 would have been too old to attend even the first Rosenwald school. As the program expanded, later birth cohorts enjoyed a higher likelihood of attending a Rosenwald school. However, the Rosenwald program did not build schools uniformly throughout the South; out of 1,340 counties in the states with Rosenwald schools, 65 percent received at least one school. Aaronson and Mazumder (2012) find no correlation between pre-existing black socioeconomic characteristics and the placement of Rosenwald

[13] Only a few schools were built in Missouri so I omit Missouri from the analysis. My analysis therefore includes 14 states.

[14] In 1910, North Carolina's per capita income was only $123, similar to South Carolina's $126 but lower than all other Southern states (Klein 2009). However, North Carolina's black pupil expenditures tracked those of other states between 1910 and 1950 while expenditures on white students were significantly lower than other states (Margo 1990). Furthermore, North Carolina had one of the lowest literacy rates in the region, with 81.6 percent of white men and 49.6 percent of black men between 20 and 64 being literate in 1900 (IPUMS 2010).

schools in a county but do find a relationship between white literacy and school construction; I discuss these and other potential threats to identification in the next section.

For the full southern sample, I assign to individuals a measure of their likely exposure to a Rosenwald school based on their age and their county of residence during childhood. Following Aaronson and Mazumder, I calculate two measures of exposure. The first is a simple count of the years between ages 7 and 13 in which a child had a Rosenwald school in his county; this varies between zero and seven. This measure is referred to as "school in county" in result tables. The second measure, which takes into account that Rosenwald schools were not large enough for all students, is the sum of the proportion of black students in the county who could be served by a Rosenwald school in the years during which the child was between 7 and 13. This measure is smaller than the first. The second measure is used in most analysis and is referred to as "likely seats".

I take advantage of the specific school locations in the North Carolina sample to create a potentially more exogenous measure of exposure. One concern with the full southern sample is that the placement of Rosenwald schools may not be truly exogenous. Even if the timing of school construction is not correlated with county-level characteristics, it could have been affected by unobservable factors like the degree of racial antipathy. This measure is calculated as the number of years a student had a school within a 5 mile radius of their town between the same ages as above. In this *within* county analysis, I compare individuals who were the same age and lived in the same county but who grew up different distances from a school.

D. Additional North Carolina Data Sources: Prison Registers and Juvenile Court Cases

One drawback to using Census data on incarceration to measure crime rates is that they do not include the severity of the crime or length of the sentence. Furthermore, a count of individuals in prison at a point in time captures the flow of new criminals, not the stock of all individuals who have been engaged in criminal activity in the past (who may have since been released from prison). For these reasons, I collected the full universe of prisoners admitted to the State Prison in North Carolina between 1910 and 1955 from the prison registers at the North Carolina State Archives. Available variables include date of admission, length of sentence, crime, county of birth, county of sentencing, age, gender, race, education, occupation, height, and weight; for legal reasons, I was not allowed to copy the names of prisoners. I create a 10 percent random sample from this group of more than 40,000 prisoners, resulting in a sample size of 2,765. I pool this data with my sample of non-prisoners and calculate exposure to a Rosenwald school based on the reported birth county and age as above. Outcome variables in the analysis include an indicator for being a prisoner, sentence maximum and minimum length[15], and the type of crime. The drawback of this data source is that it does not account for out-migration from North Carolina; approximately 1/4 of prisoners born in North Carolina were in jail outside of North Carolina in my primary sample.

One additional benefit of this data is that I can look at racial patterns or biases in sentencing for the same crime. I regress minimum or maximum sentence length on indicators for being black and available demographic characteristics. Without controlling for severity of crime, sentences of blacks are 1.1 years longer than for whites. However, this difference goes away

[15] Results from sentence length are not reported because there does not appear to be an effect of exposure on sentence length, *conditional* on being convicted and sentenced to the state prison.

when including type of crime fixed effects, suggesting that higher sentences for blacks are driven not by discrimination but by the fact that blacks commit more severe crimes.[16]

I also collect complete county-race-year counts of inmates admitted to the State Prison from the State Prison reports between 1930 and 1955. I assign exposure as the weighted average of exposure of individuals based on the age distribution of admissions. That is, I calculate the number of years individuals of each age would have had a school in their county and then take the weighted average based on the age of admissions. The number of admissions increases almost tenfold between 1933 and 1935 I restrict to post-1935.[17] Prior to 1931, data is not available broken down by race.

Because juvenile delinquents were so infrequently incarcerated, they are unlikely to show up in the census data.[18] Children had to be at least 16 years of age to be sent to a state prison. I use county-level data on juvenile court cases by race coded by Wiley Britton Sanders (1945) to estimate the effect of the Rosenwald program on juvenile delinquency. I have eight data points per county: separate counts for black and white children for the time periods 1919-29[19], 1929-34, 1934-39 and 1939-44. Ideally, one would use annual data to trace out the effect of a new Rosenwald school but the underlying data from which Sanders constructed his series appears to

[16] Note, however, that this does not provide definitive evidence that there is no discrimination in sentencing. It could be that blacks and whites commit the same crimes but that black crimes are "labeled" as more severe crimes in the prison register to justify higher sentences for blacks. If crimes had standard sentences, then a judge would have to convict a black person of a more serious crime to increase his sentence length.

[17] This seems to be because in the early periods the state prison reports only reported people who were sentenced to the state prison and not to road camps; data after this includes prisoners at state prison camps and road construction gangs outside of the state prison which made up a large proportion of prisoners.

[18] Juvenile court systems were not organized on a systematic basis until the 1930's and any incarceration facilities were small at best. For examples, North Carolina had only two "training schools" to which judges could send juvenile delinquents. One for whites, called the Morrison Training school, was established in 1918 but had only 30 places. It wasn't until 1925 that a facility was built for black children. The Stonewall Jackson school had room for 35 students.

[19] I divide these counts by two to get a comparable 5 year count.

be lost. I calculate average exposure for each county in the time period and assign this to the relevant data point. Specifically, the exposure measure is calculated as the number of years in this time period that the county had a school.

E. Heterogeneity

In addition to collecting information about household background characteristics, I collect data from multiple other sources to examine whether the effect of Rosenwald schools depends on local county characteristics. Compulsory schooling laws might augment the effect of new school construction. Rosenwald schools likely made schools available in areas where children could not attend school so I expect that a Rosenwald school in an area with compulsory schooling would have a larger effect on literacy/human capital and therefore a larger effect on crime than one somewhere without compulsory schooling laws.

I also look at effects that differ by county economic characteristics such as share of employment in farming or manufacturing. Finally, I use data on lynching and voting behavior to see whether the effects differ between more and less "racist" counties.[20]

1.4 Estimation Strategy

A. Reduced Form Estimation

I estimate the effect of being exposed to a Rosenwald school on the probability of being a criminal, measured by incarceration, later in life. My estimation strategy exploits differences across counties in the number of Rosenwald schools as well as variation across cohorts within a county in exposure to a Rosenwald school at relevant ages. I also contrast black and white students in the same county and cohort. All regressions also include dummy variables for single

[20] Heterogeneity results are not complete yet.

years of age by race to allow for different white and black age-crime profiles. I create sample weights to account for the full population of prisoners but only one percent sample for the comparison sample; prisoners receive a weight of one while the comparison group receives a weight of 100.

My preferred estimation strategy within North Carolina makes use of variation in the length of exposure to Rosenwald schools and in the exact placement of these schools within counties. Here, *Exposure* is equal to the number of years that an individual could have attended a Rosenwald school within five miles of their childhood home. *Exposure* is equal to zero for men who lived too far away from a Rosenwald school or who were too old to take advantage of their local school. My main estimating equation (1) is therefore:

$$prisoner_{ict} = \alpha_c + \gamma_t + \beta_1 black_i + \beta_2 exposure_{ict} + \beta_3 black_i * exposure_{ict} + \epsilon_{ict}$$

where *Prisoner* equals one if the individual is incarcerated. This specification allows me to include county fixed effects.

The coefficient of interest in this equation is β_3 which measures the additional effect of a year of exposure to a local Rosenwald school on black youth over any effect there may be on white youth. The total effect of exposure on black incarceration is $\beta_2+\beta_3$.

This equation uses multiple sources of variation to address potential concerns about bias in the estimate of β_3. I compare blacks and whites to control for any shocks that are correlated with Rosenwald school construction but which affected both races equally. I look within a county with county fixed effects to control for any differences between counties which received schools and those that did not. The estimate will only be biased if there are local events that are correlated with the timing of construction of Rosenwald schools and with trends in incarceration. I use county fixed effects because, without them, we might expect the estimate to be biased

downwards (more negative) since it is plausible that counties with Rosenwald schools had positive attitudes towards their black populations which might also translate into lower incarceration rates. Furthermore, using whites as a comparison group for blacks allows me to wash out anything which affects both blacks and whites and which is correlated with Rosenwald schools and crime. Any confounding factors must affect only black residents but not neighboring whites. Finally, note that Rosenwald school exposure is measured in childhood but crime is measured at least ten years later so any county-level confounding factor would have to be constant throughout this gap.[21]

B. First stage estimation

My main interest is in the direct effect of Rosenwald schools on crime. One likely channel through which Rosenwald schools reduced criminality is by increasing the literacy rates and educational attainment of its black pupils. As a result, I also present quasi-first stage estimates of the effect of Rosenwald schools on both literacy and educational attainment for my sample. These results complement Aaronson and Mazumder (2012), who show that Rosenwald schools increased school enrollment of affected cohorts and improved educational attainment of WWII enlistees.

My first stage equation follows the format of equation (1) above. I estimate:

$$education_{ict} = \alpha_c + \gamma_t + \beta_1 black_i + \beta_2 exposure_{ict} + \beta_3 black_i * exposure_{ict} + \epsilon_{ict}$$

where β_3 is the effect of Rosenwald schools on black literacy versus white literacy. The sum of $\beta_2+\beta_3$ is the overall effect of Rosenwald schools on black literacy. *Education* is measured

[21] Note that household characteristics are constant from the perspective of the individual so if parental education is correlated with opening of the schools, then there could still be a problem. I address this by (a) controlling for parental education and other characteristics in some regressions; (b) arguing in Appendix B that education of blacks does not seem to predict when a Rosenwald school opens; (c) constructing within household estimates of the return to Rosenwald schools.

either as literacy or as years of education in 1940. For 1940, I translate years of education into literacy based on the assumption that individuals with three or more years of schooling qualify as literate.[22]

I use the first stage estimate to calculate a Wald estimate of the effect of literacy on incarceration. That is, what is the predicted difference in incarceration rates between someone who is literate and someone who is not? In order for this to be interpreted as an IV estimate, I must be willing to assume that Rosenwald schools only affected incarceration through literacy or learning. In fact, access to schooling may have reduced criminality in other ways, principally by keeping children occupied during the day.

C. Identification

One main concern about identification may be that counties who received schools are fundamentally different from those which did not receive schools. I address this concern by using county fixed effects so that something which would cause my estimates to be biased must be changing over time within a county and be correlated with the introduction of schools *and* with crime or incarceration. However, these changes would have to affect exposed and non-exposed cohorts differentially. For example, we might think that some counties have racist attitudes which would lead them not to build schools and to also tend to incarcerate black men more often.

Following Aaronson and Mazumder, I test whether a variety of county-level variables predict school locations. I find that counties with higher literacy rates of blacks and whites or with a higher percentage of black farmers tend to have schools and to receive them earlier than

[22] Collins and Margo (2006) argue that literacy is the equivalent of 1-3 years of school. Results are not sensitive to the choice of two or three years as the cutoff.

other counties. There is some evidence of a positive correlation between the percentage foreign born and receiving a school. See Appendix B for details.

We might also be worried that similar concerns plague the timing of school building within counties. This problem would not be eliminated with county fixed effects. Particularly in North Carolina, almost all counties ended up with at least one school by the end of the program, so it might not be across-county differences that are a problem, but that certain counties decided to build schools earlier or later. In Appendix B, I show that that black literacy, increases in the share foreign born, and increases in population density predict the length of time from the beginning of the program in 1913 to the date a county first built a school. I control for census year county characteristics in some regressions and results are not affected.

A final identification issue is that it is possible that families migrated towards counties with Rosenwald schools, either because they anticipated good schools for their children or as a response after the schools were built. I try to address this by assigning exposure based on the earliest census in which I can find the child. If the family migrated later, then the individual still gets assigned exposure from the previous county. Further, I regress literacy of men ages 30-50 on Rosenwald exposure in the past ten years; the idea is that if families migrated towards Rosenwald schools, then average literacy of men who are likely to be parents of exposed children should increase as Rosenwald exposure increases. I find no evidence that this was the case; I also find no evidence that average incomes of the same ages increased.

1.5 Results

A. Reduced form results

My empirical analysis begins in Table 2 by presenting estimates from equation (2) for the full sample of southern counties. In columns (1) and (2), I define my first measure of exposure as "likely seats", referring to the number of years a school was in a county, weighted by the probability of having a seat. In columns (3) and (4), exposure is "presence of school" which is measured as the number of years a student had a school anywhere in his county between the ages of 7 and 13.

Using the first measure of exposure, one year of exposure to a Rosenwald school reduces the probability of being a prisoner by 0.096 percentage points, or just under 20 percent of the average propensity of a black individual to be imprisoned. This number falls to 0.05 percentage points when using the second measure of exposure. This pattern is what we would expect if we think about these two measures as "intent to treat" measures. The first measure is twice as likely to lead to an extra year of schooling (Aaronson and Mazumder 2011). Columns (2) and (4) restrict the sample of prisoners to those living in enumeration districts that are likely to contain state and federal prisons, namely, those enumeration districts where there are 10 or more prisoners in the census.[23] Estimates decrease slightly, suggesting that part of the effect might be coming from inmates of local jails.

Table 3 restricts the analysis to the North Carolina subsample. As noted above, North Carolina is poorer than most southern states at this time and had higher exposure to Rosenwald schools. Results are similar with both measures of county-level variation and with local-level variation. Recall that the local variation measures a child as exposed if they lived within a five

[23] The name of the prison or institution is not available in my FamilySearch.org data. I have manually looked up the institution for all individuals in the North Carolina sub-sample but not in the full sample.

mile radius of a Rosenwald school. One year of exposure to a school reduces the probability of being a criminal by between 0.04 and 0.05 percentage points. This is equivalent to reducing the probability of being incarcerated by about 10% on a base of 0.55 percent for blacks. In all regressions, exposure of whites to Rosenwald schools is small and insignificant, indicating that Rosenwald schools did not affect white crime.

B. "First stage" estimates and Wald Estimates

The above results suggest that Rosenwald schools reduced the criminality of black students later in life. The most likely mechanism for this decline in crime is that educational attainment raises the opportunity cost of engaging in criminal activity by increasing literacy rates and therefore wages. Other mechanisms include the effect of education on migration, if youth move to locations with lower crime rates, or the "incapacitation" effect, whereby staying in school keeps children occupied, preventing them from entering a life of crime. I look at these explanations below, but in this section I estimate the effect of Rosenwald schools on literacy. I run an identical regression to above but the dependent variable is whether the individual is literate; results are reported in Table 4. In the full southern sample, exposure to one more year of a Rosenwald school raises the probability of being literate by between 3.1 and 3.8 percentage points. These numbers are larger than the number from Aaronson and Mazumder who find that the probability of being literate increased by 1.8 percentage points for one year of exposure[24]. Literacy is not yet available for the North Carolina subsample.[25]

[24] This is most likely because Aaronson and Mazumder also use urban students as a comparison group. They find effects on these urban students so adding the two coefficients (black*exposure*rural and black*exposure) would give a similar number.

[25] I am in the process of collecting literacy of all prisoners in North Carolina.

I calculate Wald estimates of the social return to literacy in Table 5. The Wald estimate is the reduced form coefficient divided by the first stage.[26] By this estimate, teaching a black man to read would reduce his likelihood of later-in-life incarceration by between 1.2 percent and 3.8 percentage points. In comparison, Lochner and Moretti find that an additional year of schooling reduces the likelihood of imprisonment by only 0.4 percentage points. If, instead, I consider the effect of the Rosenwald program on years of educational attainment, I find that an additional year of schooling reduces imprisonment by 0.25 to 0.4 percentage points,[27] a value that is much closer to Lochner and Moretti.

These estimates would be valid two-sample instrumental variables estimates if the only mechanism through which Rosenwald schools affected crime was through literacy. However, it is likely that the Rosenwald program affected criminality through multiple channels; work is ongoing to disentangle these mechanisms. Wald estimates are presented here simply to give an idea of the magnitude of the coefficients that are estimated in the reduced form analysis.[28]

C. North Carolinian additional data sources

The benefit of the Census data used thus far is that it provides a picture of all prisoners in the full South. A drawback is that it does not contain information about type of crime or sentence length, two characteristics that provide indirect information on whether and to what extent the racial gap in incarceration is simply due to the discriminatory Jim Crow legal system, rather than

[26] I calculate standard errors using the delta method. Standard errors presented assume that the coefficients are uncorrelated but the standard errors are similar and the coefficients are still significant letting the correlation range between -1 and 1.

[27] These Wald estimates comes from dividing the first stage coefficient (0.04-0.1) by the 0.2 years of schooling first stage coefficient.

[28] I also plan to add results considering heterogeneous effects of exposure, as well as estimates which compare within households to account for any omitted household background characteristics. Data collection for both of these is currently ongoing.

to "true" differences in criminal activity. One may be concerned that exposure to Rosenwald schools reduced discrimination, which then led to less imprisonment in the local area. Another potential issue is that census data measures a stock of crime in a census year, not the flow of prisoners into prisons; if whites get shorter sentences than blacks, then we might overstate black crime. The first alternative hypothesis is unlikely because prisoners are rarely incarcerated in their county of birth; about 20 percent of prisoners are outside of their state of birth with another 15 percent also within the same state but outside of the county of birth.

In this section, I present estimates from different data sources that address concerns about discrimination in the southern legal system or measurement problems in census data. First, the state prison reports provide a flow measure of incarceration in each year. Another benefit of this data is that state prisoners are more likely to have committed "real crimes," rather than, for example, being imprisoned due to enforcement of vagrancy laws, which was more likely to happen in local jails. Second, I examine county-level counts of court cases; this is a measure of the number of individuals *accused* of a crime, rather than only those convicted and sentenced for a crime. If the main effects derive from discrimination in conviction rates, due to partial juries, or sentencing lengths, due to biased judges, I will not find an effect of Rosenwald schools on the case load. Following this, I look at juvenile court case data; because juveniles are rarely incarcerated but instead are put on probation, Census data does not allow me to consider the contemporaneous effect of Rosenwald schools on youth activity. Finally, I incorporate micro-data from the state prison registers about the type of crime committed to determine whether education is more likely to prevent violent or property-based crime.

i. State Prison Reports and Attorney General Reports

Results from the State Prison Reports are shown in Table 6. Average exposure and sample means are given at the bottom of the table. I create covariates based on the population census data in the decennial years and linearly extrapolate between years. Controls include the total black and white male populations, the total county population, the share of the population that is black, and the share of the population that lives in urban areas. I use the total number of prison intakes in a county and the log of intakes of the specific race as the outcome variables.[29] Exposure is a weighted average based on the age distribution of prisoners and the timing of school openings in counties.[30] I find that exposure to Rosenwald schools for one year would reduce the black-white gap in incarceration at the county level by 15 prisoners for a mean of 93 prisoners in column (1) or by 37 percent for a mean gap of 66 percent in column (2).

An alternative measure of crime is the number of individuals accused of a crime in a given county in each year. I digitize county level court cases from the Attorney General reports from 1913-1935 (the data is not available in later years) and find that court cases against blacks also decrease by a similar magnitude with average exposure in a county. Results are shown in Table 7.

The results from the State Prison reports suggest that what I find in census data is not due to discrimination in sentence lengths or due to petty crimes. In fact, exposure to Rosenwald schools decreases the probability of being incarcerated in state prison. The fact that court cases go down as well suggests that arrests decrease with Rosenwald exposure. While I cannot rule out

[29] County of birth is not available in this data so I use county of arrest. To the extent that there is migration from rural counties to urban counties, I will understate exposure since urban areas had fewer Rosenwald schools.

[30] Because the mean of the dependent variable increases by a factor of ten between 1933 and 1935, I restrict to post-1935 in my analysis.

that court cases are related to discrimination at the arrest stage, this provides indirect evidence that my results are not due to discrimination at the sentencing stage.

ii. Juvenile Court Data

I consider contemporaneous effects of Rosenwald schools on the criminal activity of youth in Table 7. The observed decline in adult crime might arise because additional years of schooling increase the opportunity cost of crime. Alternatively, the main effect could be due to persistence of criminality: offenders who commit crimes at young ages often commit crimes later in life as well. If Rosenwald schools reduce crime among juveniles, these individuals might also be less likely to commit crimes later. I regress the number of juvenile court cases for blacks or whites in a given county within a five year period on exposure to Rosenwald schools and the same set of county controls as above. Exposure is measured as the number of years over the previous five year period the county had a school. Exposure to Rosenwald schools has a negative and significant effect in all specifications. The effect of Rosenwald schools on white outcomes is insignificant and almost always small as we would hope. The effect of exposure on black juveniles is a decrease of 9.4 court cases per five years from the average of about 50 court cases.

The magnitude of the effect on juvenile crime is very similar to that on adult crime in the census. If anything, it is larger. About 36 percent of juvenile offenders commit a crime at some point later in life (Illinois 2007). If we assume that these will happen in the next ten years and will require on average five years of incarceration (the average sentence length in the North Carolina state prison), then census data should capture 18 percent of them; incarceration in census data therefore should decrease by about 2 percent of the mean (=0.18*0.1). That means that the decrease in incarceration in census data could be partly due to the fact that kids have

been kept out of a life of crime early on. However, 80% of the 10 percent decrease in crime is likely due to less non-juvenile offenders.

iii. State Prison Registers

In Table 8, I turn to individual-level prison register data. It is hard to assess the social return to the decline in crime documented in the main results without knowing the distribution of crimes affected. If educational attainment only reduces rates ofpetty theft, the potential return to the Rosenwald Program would be lower than if schooling reduced violent or other severe crimes. Furthermore, any decreases in crime could have come from reductions in the number of black individuals picked up for vagrancy (Naidu 2010).

I estimate a multinomial logit by pooling my state prison data with the comparison samples of non-prisoners. I categorize the types of crime into four groups: property crimes, violent non-fatal crimes, violent fatal crimes, and other crimes. Property crimes include robbery, burglary, and housebreaking. Violent non-fatal crimes are mainly rape and assault. Fatal crimes are manslaughter and murder. Within the "other" category, there are many white-collar crimes such as forgery and embezzlement.

I find that exposure to a Rosenwald school decreases the likelihood that black men commit any of these crimes to a roughly equal degree. The relative risk ratios are all less than 1, indicating that a year of exposure to Rosenwald schools reduces that category of crime relative to the category of not being in prison.[31] These results suggest that education reduces crime of all types.

[31] I have also carried out similar analysis using only the prison population. It seems like the crimes that are reduced the most are property crimes and non-fatal violent crimes. Breaking into more categories makes the estimates less precise but there is some weak evidence that embezzlement and forgery increase with exposure, consistent with the idea that schooling increases skills that are useful for certain types of crimes.

1.6 Robustness and Alternative Specifications

In this section I turn to robustness measures and alternative specifications. I first discuss potential concerns about the sample and sources of variation that are driving my results. Then I investigate non-linear effects of exposure and the hypothesis that there is a critical "age window" in which exposure to a Rosenwald school has the strongest effect on later-life outcomes. All regressions use the "likely seats" measure of exposure which is the number of years a child had a school in his county, weighted by the probability of having a seat available.

Table 9 contains results for five robustness samples. One potential worry is that there are county time trends that might affect schooling and lower crime later in life. For example, if parents are becoming more literate over time in Rosenwald counties, then this could explain part of the effect. In columns (2) and (3) I address this issue. In column (2), I restrict to ages less than 30 since 30 to 35 year olds might not be a good comparison group for an 18 year old. In column (3), I only keep a sample of exposed individuals and those who "just missed" exposure in the sense that they were too old but there was a school built between the ages of 14 and 17. The effect of exposure stays the same in these two samples so it does not appear that the results are sensitive to the ages of the comparison sample.

Another issue could be that if there are peer effects which *do* affect the "just missed" group through their slightly younger peers being more educated, then the estimate of the effect of exposure might be too small. This could work through either direct peer effects as children or higher wages in Rosenwald counties for everyone as a function of the average increase in literacy due to Rosenwald schools.[32] I therefore restrict to those who were exposed and those

[32] I plan to examine the county-level effects of Rosenwald schools in more detail in another paper. Work to examine this issue for this paper is ongoing.

who did not just miss. That is, either individuals were exposed for a certain number of years between ages 7 and 13 or they did not have a school in their county until after age 18. The coefficient remains the same, indicating that the effect does not come only through peer effects or effects on county-level wages.

Another question is whether whites are a good comparison group for blacks. If there were things affecting the county judicial system which affected whites and blacks differently, then maybe whites are not a good comparison. I argue that, first, the coefficient on exposure for whites is never statistically or economically significant. Second, in column (5), I restrict to blacks only and regress the probability of being a prisoner on exposure alone. Note that in this case, I cannot include county-level fixed effects because much of my variation is coming off of cross-race variation. With state-level fixed effects, the coefficient is much the same as before albeit insignificant.

Finally, one might wonder whether the variation that is driving the results is from counties with Rosenwald schools compared to counties without schools. In column (6), I show that this is not the case; if anything, the coefficient is larger when restricting to counties which built a school at some time.

The results thus far have relied on the assumption that the effect of exposure is linear. In Table 10, I consider alternative specifications. First, I look at whether the result is robust to simply comparing those who are exposed by any amount to those that are not. The average effect of any exposure is a 0.22 percentage point decrease in the likelihood of being a prisoner for blacks. This number is larger than the previous 0.09 percentage points per year of exposure, suggesting that the effect is not coming all from exposed versus unexposed groups.

Second, I allow the effect of one year of school to vary based on the age of the individual. It might be that older children have "missed the boat" if a school is built in their county at age 10 or 11. I do find some evidence that this is the case in column (2) where I include the exposure between age 7 and 9 and between ages 10 and 13 separately. The coefficient on the first time period is negative and significant but the coefficient on ages 10 to 13 is small and marginally significant. This suggests that new schools had a larger affect at younger ages. I also allow the effect of a year of school, at any age, to be non-linear. In column (3) we see that there do not appear to be diminishing effects.[33]

1.7 Conclusion

This paper considers the social returns of a program which increased the schooling of black children between 1920 and 1940. Rosenwald schools increased schooling and literacy, which resulted in lower crime rates. I find social returns to school of about a 0.4 percentage point decrease in crime, which is similar to the literature.

My results are relevant to a developing economy today, such as South Africa, where inequality between blacks and whites is similar with respect to incarceration rates, education, and wages. This is the first paper to consider social returns to education in a historical, developing economy environment; previous literature has considered social returns to school primarily in the United States and Europe where inequality is lower, institutions are stronger, and incomes and education levels are higher.

Further work will examine migration responses to Rosenwald schools and the relationship between migration and black-white crime gaps. The historical gap between blacks and whites in

[33] I have also tried regressions with dummy variables for exposure at each age and dummy variables for one through seven years of exposure. Both regressions are imprecisely estimated but there is a similar pattern of higher values at younger ages and diminishing returns to exposure after 5 years.

incarceration is not well explained and is hardly documented in current literature. Education is one factor that helps explain this gap, but the evolution of the gap over the 20th century is not well understood and is a subject for future research.

Appendix A: Matching

The matching procedure starts with the base year of 1920, 1930, or 1940 and matches backwards to either 10 or 20 years prior. The procedure is as follows:

(1) I begin by standardizing the first and last names of men in the later year to address orthographic differences between phonetically equivalent names using the NYSIIS algorithm (Atack and Bateman 1992). I also recode any common nicknames to standard first names (e.g. Will becomes William). I restrict my attention to men in 1900 that are unique by first and last name, birth year, race, and state of birth. I do so because, for non-unique cases, it is impossible to determine which of the records should be linked to potential matches in the earlier year.

(2) I match observations backwards from the later year to the earlier year using an iterative procedure. I start by looking for a match by first name, last name, race, state of birth, and exact birth year. There are three possibilities:

(a) if I find a *unique* match, I stop and consider the observation "matched";

(b) if I find multiple matches for the same birth year, the observation is thrown out;

(c) if I do not find a match at this first step, I try matching within a one-year band (older and younger) and then with a two-year band around the reported birth year; we only accept unique matches. If none of these attempts produces a match, the observation is discarded as unmatched.

My matching procedure generates a final sample of 28,455 prisoners and 36,972 non-prisoners. I can successfully match 20.3 percent of all prisoners and 21.5 of non-prisoners. The difference between these match rates is most likely attributable to lower education levels among prisoners (e.g. they are less likely to report the correct age or to spell their names consistently

enough to match). I have lower match rates for blacks than for whites, which is also consistent with lower education levels.

Individuals can fail to match due to (i) non-unique name-birthplace-race combinations; (ii) multiple matches in the earlier year; (iii) mis-reporting of age; and (iv) complete misspellings of the name. Note that mortality cannot account for any failure to match due to starting with the later year and matching backwards. Non-unique combinations in 1940 account for 40% of match failures. Multiple matches account for 12%. Allowing individuals to match within a 10 year age range gains an additional 10% so these are likely misreported ages. Finally, individuals who cannot be found because of differences in name spellings account for the remaining 38%. Note that this could be because the individual misspelled their name (likely correlated with education, etc) or because the enumerator or the modern transcriber misspelled it (random). In fact, the correlation between the census indexes of Ancestry.com and FamilySearch.org is quite low, suggesting that a lot of the differences in spelling are due to transcription differences in the index.

Appendix B: Locations of Rosenwald Schools

A. Relationship between county characteristics and Rosenwald schools

I consider the relationship between pre-Rosenwald county characteristics and the existence of a school. In Table B.1, I regress four outcomes on a set of race-specific socioeconomic variables and other demographic county variables. The outcomes are (a) whether the county ever had a Rosenwald school; (b) The total number of schools built in the county by the end of the program in 1931; (c) whether the county had a school before 1919; and (d) the number of years until the first school was built for counties which did get a school. I use literacy, school enrollment, and average occupational score as my socioeconomic variables. I also consider the effect of population density, the share of the population living in urban areas, the county's share of black and foreign born men, and the percentage of households that own a home. To account for the idea that distance from the Tuskegee Institute or other important historical black institutions might matter for the outcome variables, I include the distance from the center of the county to the Tuskegee Institute, Montgomery, AL, Nashville, TN, and Hampton, VA. Regressions also include state fixed effects and industry share controls by race. Finally, I consider the level of all of these variables in 1900 and 1910 and also the change between 1900 and 1910; that is, the share of the population that is urban might matter, but urbanizing counties might also be more or less likely to build a school.

In Panel A, I find a positive relationship between black literacy and ever having a Rosenwald school. A ten percentage point increase in black literacy in 1910 would increase the probability of having a school by 2.2 percentage points. The coefficients on white literacy are almost twice as large but insignificant. There is also some evidence that counties with higher population density and urbanizing counties with increasing urban share between 1900 and 1910

were more likely to get a school. In Panel B, the number of schools built by 1931 is positively correlated with population density, black literacy, the percent of farmers who are black, and the total number of farms in the county. Counties which built a school before 1919, in Panel C, tended to have a lower percentage of black farms, higher *white* literacy, a higher percentage of black men, and lower percentage of black owner-occupier farmers. Counties where home ownership or the percentage of black farmers were growing were less likely to get a school before 1919. Finally, in Panel D, we see that counties with increasing population density or increases in the share foreign born waited fewer years before a school was built. Those with higher black literacy also waited fewer years. The percentage of black farmers is positively correlated with the number of years until the first school was built.

B. Family migration decisions and Rosenwald schools

Another concern is that families who were more literate or of higher socio-economic status may have migrated towards counties with Rosenwald schools. In that case, we would expect average income and literacy of individuals not directly affected by the Rosenwald schools to increase in a county *after* the introduction of the schools. That is, average literacy in a county with high exposure to the schools should increase more than that of counties with low exposure to the school if literate parents migrate towards schools for their children. If literate parents have children who are unlikely to be criminals, this might bias my results.

To examine this, I regress literacy of men ages 30 to 50 on indicators for county Rosenwald exposure in various years of the previous decade. Because the data is noisy, I group it into three categories: exposure 1-3 years earlier (*rose1)*, exposure 4-6 years earlier (*rose2*), and exposure 7-9 years earlier (*rose3*). I interact each exposure measure with a dummy variable for

being black. I would worry about selective migration if these interaction terms are significant. The results are shown in Table B.2. I find no significant evidence that there was selective in-migration to Rosenwald counties of more literate families between census years. There is some weak evidence (p=0.09) that Rosenwald exposure in the previous three years is associated with higher occupational scores of men in this age range, but this test is less clean; it is reasonable that Rosenwald schools could have increased wages for men so this is not direct evidence of selective migration

Table B.1: Correlation between 1900 and 1910 county characteristics and Rosenwald schools

	(a) Ever had a Rosenwald school			(b) Number of schools, 1931		
	1900	1910	Change	1900	1910	Change
Pop density	0.075	0.329**	0.085	3.038***	5.273**	3.071*
	(0.137)	(0.115)	(0.076)	(0.980)	(1.981)	(1.581)
% Black	1.305	1.144**	0.371	3.753	-0.088	13.373**
	(0.363)	(0.424)	(0.439)	(2.971)	(3.195)	(5.057)
% Urban	-0.0005	0.116	0.526***	1.111	1.282	7.438**
	(0.109)	(0.087)	(0.161)	(2.105)	(1.708)	(2.831)
% Foreign-born	-0.245	0.572	0.843	-4.473	-2.940	9.643
	(0.669)	(0.662)	(0.846)	(4.871)	(5.905)	(16.76)
% Own home	0.564	0.715***	-0.128	-3.556	2.404	4.180
	(0.103)	(0.154)	(0.259)	(3.195)	(3.421)	(3.134)
% Lit White	0.714	0.745	-0.869	1.427	1.051	-6.861*
	(0.369)	(0.455)	(0.623)	(5.291)	(7.369)	(3.668)
%Lit Black	0.462	0.224**	-0.138	3.056*	3.507*	2.449
	(0.114)	(0.095)	(0.101)	(1.551)	(1.881)	(1.743)
# farms (1000s)	0.0053	0.004**	0.005	0.124	0.124***	0.264***
	(0.0021)	(0.001)	(0.003)	(0.034)	(0.020)	(0.056)
% farms black	-0.344	-0.436	-0.497***	3.396***	10.712	-10.949***
	(0.151)	(0.319)	(0.116)	(2.054)	(3.971)	(2.116)
% black farmers who own	-0.083	-0.062	-0.046	-1.425	-0.754	0.365
	(0.087)	(0.046)	(0.084)	(0.852)	(0.683)	(0.724)
% white farmers who own	-0.228	-0.513*	-0.061	7.556**	-1.375	-7.571
	(0.221)	(0.235)	(0.262)	(2.645)	(3.291)	(4.419)
Sample Mean	0.65	0.65	0.65	3.495	3.495	3.495
N	1,088	1,154	1,067	1,088	1,154	1,067
R-squared	0.3998	0.3929	0.3593	0.4214	0.4548	0.4226

Notes: All regressions control for industry shares by race and include state fixed effects. They also control for distance from the center of the county to Tuskegee Institute, Hampton, Virginia, Montgomery, AL, and Nashville, TN. Standard errors are clustered at the state level.

Table B.1: Correlation between 1900 and 1910 county characteristics and Rosenwald schools

	(c) Has a school by 1919			(d) Years until first school		
	1900	1910	Change	1900	1910	Change
Pop density	0.162	0.277	0.125	0.261	2.460	-5.988**
	(0.110)	(0.167)	(0.083)	(1.809)	(3.648)	(2.298)
% Black	0.781***	0.340	0.205	-5.061	-2.664	-1.297
	(0.259)	(0.377)	(0.402)	(4.319	(2.905)	(1.946)
% urban	0.166	0.217**	0.278	-0.292	-0.429	-1.329
	(0.122)	(0.097)	(0.192)	(0.931)	(0.764)	(2.231)
% Foreign-born	-0.487	-0.298	0.448	5.780	1.365	-13.026*
	(0.301)	(0.355)	(1.235)	(4.107)	(6.544)	(7.144)
% Own home	0.151	-0.011	-0.482**	-2.026	-2.034	2.500
	(0.112)	(0.289)	(0.222)	(1.845)	(1.683)	(1.891)
% Lit White	0.225	0.534***	-0.504	1.211	2.851	4.051
	(0.231)	(0.170)	(0.391)	(3.207)	(2.441)	(5.819)
%Lit Black	0.154	0.093	0.196	-2.179	-3.932*	-3.458*
	(0.112)	(0.158)	(0.195)	(2.541)	(1.866)	(1.930)
# farms (thousands)	0.005**	0.004**	0.006**	-0.052***	-0.042***	-0.058*
	(0.003)	(0.002)	(0.003)	(0.014)	(0.013)	(0.029)
Proportion farms black	-0.166*	0.033	-0.459***	0.157	-1.842	3.031***
	(0.085)	(0.265)	(0.129)	(1.475)	(2.414)	(0.598)
% black farmers who own	-0.141**	-0.016	-0.012	-0.670	-0.573	-0.084
	(0.059)	(0.044)	(0.039)	(0.808)	(0.541)	(1.639)
% white farmers who own	0.232	-0.110	-0.337	-2.091	-0.202	2.516
	(0.188)	(0.213)	(0.188)	(1.643)	(1.845)	(2.148)
Sample Mean	0.225	0.225	0.225	3.153	3.153	3.153
N	1,088	1,154	1,067	782	767	781
R-squared	0.2963	0.3003	0.2789	0.2682	0.2685	0.2553

Notes: All regressions control for industry shares by race and include state fixed effects. They also control for distance from the center of the county to Tuskegee Institute, Hampton, Virginia, Montgomery, AL, and Nashville, TN. Standard errors are clustered at the state level.

Table B.2: Selective migration of literate and higher occupational status families to Rosenwald Counties

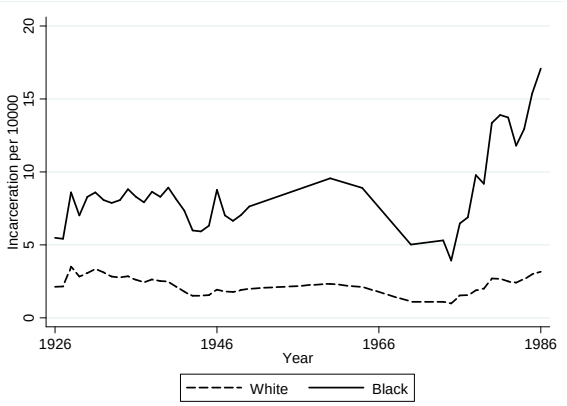
Outcome:	Literacy	Literacy	Occupational Score	Occupational Score
Black*Rose1	0.011	-0.017	0.927	2.097*
	(0.018)	(0.018)	(0.769)	(1.221)
Black*Rose2	-0.005	0.031	-0.145	-0.900
	(0.029)	(0.032)	(0.855)	(1.245)
Black*Rose3	0.004	-0.037	-0.632	-0.619
	(0.025)	(0.036)	(0.502)	(0.686)
Rose1	-0.002	0.006	0.042	-0.746
	(0.007)	(0.006)	(0.257)	(0.490)
Rose2	0.013	0.004	-0.249	0.846
	(0.010)	(0.010)	(0.405)	(0.505)
Rose3	-0.019**	-0.011	0.169	-0.302
	(0.008)	(0.010)	(0.358)	(0.302)
Ages	30-50	30-40	30-50	30-40
R-squared	0.1473	0.1443	0.1183	0.1001
N	51,850	29,807	51,850	29,807

Notes: regressions include fixed effects for age, black*age, year, state*year, and county. I restrict to men living in rural areas in the census year since schools were targeted to rural areas.

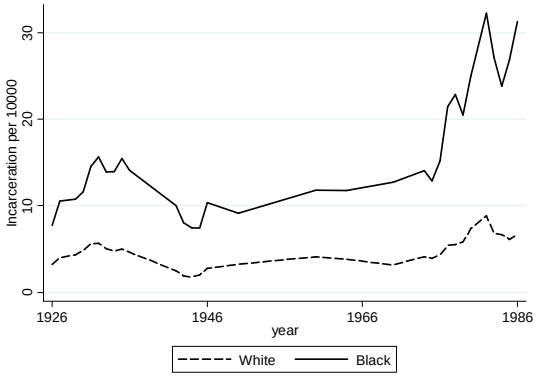
Figures

Figure 1.1: Incarceration Rates by Race and Region, 1926-1986

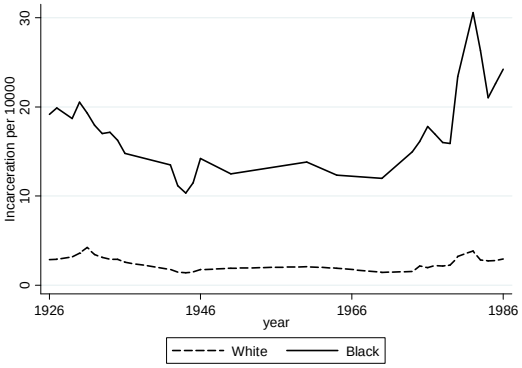
A: Whole US



B: Southern States



C. Northern States



Tables

Table 1.1: Summary Statistics

	Prisoner	Non-Prisoner
Full South		
Percentage (Black subsample)	0.688	99.31
Percentage (White subsample)	0.186	99.81
Age	25.18	25.09
	(4.651)	(4.949)
Average Exposure (coverage)	0.270	0.410
	(0.800)	(1.029)
Average Exposure	1.318	1.894
	(2.833)	(2.815)
Percentage living outside state of birth	21.19	14.34
North Carolina		
Literate	0.822	0.921
	(0.382)	(0.270)
Both Parents Literate	0.688	0.778
	(0.463)	(0.415)
Father = Owner-Occupier Farmer	0.259	0.340
	(0.439)	(0.474)
Father = Tenant Farmer	0.283	0.275
	(0.451)	(0.447)

Notes: N=76,639. North Carolina statistics are calculated from available (incomplete) data. North Carolina subsample restricts to those living in North Carolina in the childhood census.

Table 1.2: Reduced Form Results: Whole South

Exposure:	Likely Seats		Presence of School	
	Prisoner	Prisoner, likely state & federal	Prisoner	Prisoner, likely state & federal
Black*Exposure	-0.0939***	-0.0760***	-0.0427***	-0.0327***
	(0.0122)	(0.0120)	(0.0050)	(0.0049)
Exposure	0.0031	-0.0009	0.0020	-0.0027
	(0.00003)	(0.0041)	(0.0022)	(0.0022)
Black	0.5139***	0.4339***	0.5967***	0.4946
	(0.0003)	(0.0352)	(0.0039)	(0.0371)
County FE	Y	Y	Y	Y
Age, Age*Black FE	Y	Y	Y	Y
Mean Exposure	0.410	0.410	1.717	1.722
Sample Mean	0.550	0.553	0.553	0.556
Number of counties	1,321	1,112	1,321	1,112
N	65,885	62,351	65.885	62,351

Notes: Coefficients are multiplied by 100 to be expressed as percentages instead of proportions. Regressions include year, and state*year fixed effects. In columns (2) and (4), I restrict to enumeration districts in which there were at least 10 prisoners in the census of that year; these prisons were more likely to be state and federal prisons than county jails. I restrict to ages 18-35.

Table 1.3: Reduced Form Results, North Carolina

	(1)	(2)	(3)
Black*Exposure	-0.0495***	-0.0462***	-0.0469*
	(0.0112)	(0.0111)	(0.0464)
Exposure	-0.0002	-0.0076	0.0011
	(0.0631)	(0.0500)	(0.0566)
Black	0.2611***	0.2581***	0.0492***
	(0.0612)	(0.0613)	(0.0045)
Mean Exposure	3.469	1.001	0.6827
Sample Mean	0.0032	0.0032	0.0032
Exposure	Likely Seats	School in County	Local Radius
Black*age and age FE	Y	Y	Y
County FE	Y	Y	Y
Number of counties	100	100	100
N	9,008	8,363	8,145

Notes: Coefficients are multiplied by 100 to represent percentages instead of proportions. Regressions include age, black*age, county, year, and state*year fixed effects. I restrict to ages 18-35 and to those living in North Carolina in the childhood census.

Table 1.4: Effect of Rosenwald Exposure on Literacy

Sample:	(1) Full Ipums	(2) North Carolina Ipums	(3) North Carolina Ipums
Black*Exposure	0.0368**	0.0390**	0.0218*
	(0.0145)	(0.0155)	(0.0132)
Exposure	-0.0006	0.0020	0.0010
	(0.0010)	(0.0031)	(0.0043)
Black	-0.1245***	-0.1323***	-0.1295***
	(0.0341)	(0.0568)	(0.0556)
Mean Exposure	1.023	1.234	0.655
Sample Mean, black	0.842	0.816	0.818
Exposure	Likely Seats	Likely Seats	Local Radius
R-Squared	0.1044	0.0955	0.0801
N	31,456	4,563	4,345

Notes: Coefficients are multiplied by 100 to be expressed as percentages instead of proportions. Regressions include year and state*year fixed effects. I restrict to ages 18-35.

Table 1.5: Wald Estimates of the Effect of Literacy on Incarceration

	(1) Full South	(2) North Carolina	(3) North Carolina
Literate	-2.471***	-1.210**	-1.512*
	(1.104)	(0.0667)	(0.887)
Mean Literacy, black	0.842	0.816	0.816
Exposure Measure	Likely Seats	Likely Seats	Local Radius
Black*age and age FE	Y	Y	Y
County FE	Y	Y	Y
N	31,466	4,563	4,355

Notes: Coefficients are multiplied by 100 to be expressed as percentages instead of proportions. Regressions include year and state*year fixed effects.. I restrict to ages 18-35. Wald Estimates are calculated as the reduced form coefficient on black*exposure divided by the first stage coefficient on the same variable. Standard errors are calculated using the delta method. Sample sizes are presented for the second stage sample.

Table 1.6: Effect of Rosenwald Schools on County/Race Admissions to State Prison
Dependent variable= Number of prisoners from county in each year

	(1)	(2)
Outcome:	Level	Log
Black*exposure	-5.305***	-0.193***
	(1.094)	(0.054)
Exposure	2.349	0.052
	(1.937)	(0.041)
Black	40.71***	0.543***
	(8.711)	(0.156)
Exposure	Likely Seats	Likely Seats
Years	1935-50	1935-50
Sample Mean	73.26	3.420
Average exp.	2.335	2.310
R-squared	0.7350	0.7876
N	3,200	3,064

Notes: Regressions include county and race-specific year fixed effects. Standard errors are clustered at the county level. Controls are included in all regressions: total black and white male populations, total population, share urban in the county, and share black in the county.

Table 1.7: Effect of Rosenwald Exposure on Attorney General Reports of County Court Cases

Dependent variable= Number of court cases in county by year

	(1) Level	(2) Log
Black*exposure	-17.60***	-0.297**
	(6.791)	(0.135)
Exposure	7.803	0.098*
	(4.977)	(0.053)
Black	-23.43***	-0.345***
	(6.429)	(0.135)
Exposure	Likely Seats	Likely Seats
Sample Mean	69.68	3.644
Average exp.	0.496	0.495
R-squared	0.6986	0.5987
N	3,600	3,534

Notes: Regressions include county and race-specific year fixed effects. Standard errors are clustered at the county level. Controls are included in all regressions: total black and white male populations, total population, share urban in the county, and share black in the county.

Table 1.8: Results from North Carolina Juvenile Crime Data

	(1) Level	(2) Log
Black*Exposure	-9.484***	-0.344***
	(2.564)	(0.046)
Exposure	1.802	-0.026
	(5.62)	(0.101)
Black	17.21*	0.465***
	(9.272)	(0.166)
Exposure	Likely Seats	Likely Seats
Sample Mean	50.54	2.833
Avg. Exposure	0.425	0.421
R-squared	0.8529	0.7349
N	800	733

Notes: All regressions include county and race-specific year fixed effects. Controls include total black male population, total white male population, total county population, county share urban, and county share black in the respective year.

Table 1.9: Effects on Type of Crime, North Carolina State Prison Registers

Outcome:	Not in prison	Property	Violent-fatal	Violent-nonfatal	Other
Black*Exposure	1	0.707***	0.808*	0.799*	0.754**
		(0.063)	(0.112)	(0.109)	(0.101)
Exposure	1	0.716**	0.743**	0.707***	0.672***
		(0.044)	(0.082)	(0.087)	(0.080)
Black	1	10.129***	271.64***	146.85***	0.355
		(6.329)	(327.76)	(172.35)	(0.459)
Sample proportion		0.667	0.103	0.112	0.118

Notes: N=7,074. Coefficients are Relative Risk Ratios from one multinomial regression. The proportion of crimes in each cell is given in the last line. Controls include age, black*age, and county and year fixed effects.

Table 1.10: Robustness, Full Sample

Restriction	Coefficient on Black*Exposure
1. None	-0.0939*** (0.0122)
2. Ages 18-30	-0.0911*** (0.0124)
3. Exposed and "Almost Exposed"	-0.0909*** (0.0124)
4. Exposed and "Never Exposed"	-0.0951*** (0.0156)
5. Black Only	-0.1357 (0.1477)
6. Rosenwald Counties Only	-0.1077*** (0.0146)

Notes: Regressions include age, black*age, state*year, year, and county fixed effects. Restrictions are explained in detail in the text.

Table 1.11: Alternative Specifications

	(1)	(2)	(3)
Black*Exposed	-0.2318***		
	(0.0308)		
Exposed	0.0309**		
	(0.0105)		
Black*EarlyYears		-0.2205***	-0.2208***
		(0.0306)	(0.0316)
Black*LaterYears		-0.0704	-0.3047***
		(0.0803)	(0.0582)
EarlyYears		0.0237	0.0324
		(0.0156)	(0.0166)
LaterYears		-0.0114	0.0011
		(0.0108)	(0.0201)
Early Years =		Ages 7-9	Exposed 1-3 years
Later Years =		Ages 10-13	Exposed 4-7 years

Notes: N=64,885. Fixed effects include county, age, age*black, state*year, and year. "Exposed" is defined as one if exposure is greater than zero.

References

Aaronson, Daniel and Bhashkar Mazumder. (2011) "The Impact of Rosenwald Schools on Black Achievement," *Journal of Political Economy*, University of Chicago Press, vol. 119(5): 821-888.

Anderson, D. Mark. "In School and Out of Trouble? The Minimum Dropout Age and Juvenile Crime". *Review of Economics and Statistics. forthcoming.*

Angrist, Joshua D. and Alan B. Krueger. (1992) "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples," *Journal of the American Statistical Association* 87 (418): 328-336.

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): 1832-1956.

Atack, Jeremy, Fred Bateman, and Mary Eschelbach Gregson. (1992) "'Matchmaker, Matchmaker, Make Me a Match': A General Personal Computer-Based Matching Program for Historical Research." *Historical Methods* 25(2): 53–65.

Cahalan, Margaret Werner. (1986) "Historical Corrections Statistics in the United States, 1850-1984." Rockville, MD: U.S. Department of Justice, Bureau of Justice Statistics. http://www.ncjrs.gov/pdffiles1/pr/102529.pdf.

Clark, Damon and Heather Royer. (2010) "The Effect of Education on Adult Health and Mortality: Evidence from Britain", Working Paper.

Collins, William J. and Robert A. Margo. (2006) "Historical Perspectives on Racial Differences in Schooling in the United States," *Handbook of the Economics of Education*.

Cunha, Flavio, Heckman, James J., Lochner, Lance J. and Masterov, Dimitriy V. (2006). " Interpreting the Evidence on Life Cycle Skill Formation ," In *Handbook of the Economics of Education*, edited by E. Hanushek and F. Welch. Amsterdam: North Holland, Chapter 12: 697-812.

Currie, Janet and Enrico Moretti. (2003) "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings," *Quarterly Journal of Economics*, 118(4): 1495-1532.

de Walque, Damien. (2007) "Does Education Affect Smoking Behaviors?: Evidence Using the Vietnam Draft as an Instrument for College Education," *Journal of Health Economics*, 26: 877-95.

Dee, Thomas S. (2004) "Are there Civic Returns to Education?" *Journal of Public Economics*, 88(9-10): 1697-1720.

Deming, Daniel. (2012) “Better Schools, Less Crime”. *Quarterly Journal of Economics*, forthcoming.

Donohue, John, James J. Heckman and Petra E. Todd. (2002) "The Schooling of Southern Blacks: The Roles of Legal Activism And Private Philanthropy, 1910-1960," *The Quarterly Journal of Economics*, 117(1): 225-268.

Ferrie, Joseph. (1996) “A New Sample of Males Linked from the Public Use Micro Sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census Manuscript Schedules.” *Historical Methods* 29: 141–56.

Glied, Sherry and Adriana Lleras-Muney. (2008) “Health Inequality, Education and Medical Innovation” *Demography*, 45(3): 741-761.

Heckman, James, Thomas M. Lyons and Petra E. Todd. (2000) “Understanding Black-White Wage Differentials, 1960-1990”, *The American Economic Review*: *Papers and Proceedings of the One Hundred Twelfth Annual Meeting of the American Economic Association*. 90(2): 344-349.

Hjalmarsson, Randi, Helena Holmlund and Matthew J. Lindquist. (2011) "The Effect of Education on Criminal Convictions and Incarceration: Causal Evidence from Micro-data," CEPR Discussion Paper 8646, November.

Hollandsworth, James. (2008) *Portrait of a Scientific Racist: Alfred Holt Stone of Mississippi*. LSU press: New Orleans, Louisiana.

Japelli, T, J Pischke, and N Souleles. (1998) “Testing for Liquidity Constraints in Euler Equations with Complementary Data Sources,” *The Review of Economics and Statistics*. 80: 251-262.

Klein, Alex. (2009) “Personal Income of U.S. States. Estimates for the period 1880-1910”. Warwick Economic Research Papers, no. 916, Department of Economics, University of Warwick.

Margo, Robert. (1990) *Race and Schooling in the South, 1880-1950: An Economic History*, NBER Books, National Bureau of Economic Research, Inc.

Moehling, Caroline and Anne Morrison Piehl. (2011) “Assimilation of Immigrants and Incarceration, 1900-1930.” Manuscript.

Langan, Patrick. (1991) *Race of Prisoners Admitted to State and Federal Institutions, 1926-86*. Department of Justice: Washington, D.C.

Levitt, Steven D. (2004) “Understanding why crime fell in the 1990’s: Four factors that explain the decline and six that do not”. *Journal of Economic Perspectives.* 18(1): 163-190.

Lleras-Muney, Adriana. (2005) “The Relationship Between Education and Adult Mortality in the United States,” *Review of Economic Studies*, 72(1): 189-221.

Lochner, Lance and Enrico Moretti, (2004) "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports," *American Economic Review*, 94(1): 155-189.

Manacorda, M. and E. Moretti. (2006) “Why Do Most Italian Young Men Live With Their Parents? Intergenerational Transfers and Household Structure”, *Journal of the European Economic Association*, 4 (4): pp 800-829.

Machin, Stephen, Olivier Marie, and Sunčica Vujić. (2011) “The Crime Reducing Effect of Education,” *Economic Journal* 121, 463-484.

McCrary, Justin and Heather Royer. (2009) “The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth,” NBER Working Paper No. 12329.

Meghir, Costas and Mårten Palme. (2005) “Educational Reform, Ability, and Family Background,” *American Economic Review* 95(1): 414-424.

Meghir, Costas, Mårten Palme, and Marieke Schnabel. (2011) “The Effect of Education Policy on Crime: An Intergenerational Perspective,” Research Papers in Economic No. 2011: 23, Department of Economics, Stockholm University.

Naidu, Suresh. (2010) “Recruitment Restrictions and Labor Markets: Evidence from the Post-Bellum U.S. South” *Journal of Labor Economics*. April.

North Carolina Department of Public Instruction. (1937) *Educational Directory of North Carolina.* Raleigh, NC.

North Carolina Department of Public Instruction. (1942) *Educational Directory of North Carolina.* Raleigh, NC.

Petersilia, Joan and Kevin R. Reitz. (2012) *Oxford Handbook of Sentencing and Corrections.* Oxford University Press: pp 6-7.

Sampson, Robert J. and William Julius Wilson. (2005) “Toward a Theory of Race, Crime, and Urban Inequality” in *Race, Crime, and Justice: A Reader*. Ed: Shaun Gabbidon: pp 177-189.

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

Sanders, Wiley Britton. (1933) *Negro Child Welfare in North Carolina: A Rosenwald Study.* University of North Carolina Press: Chapel Hill, North Carolina.

Sanders, Wiley Britton. (1948) *Juvenile Courts in North Carolina.* University of North Carolina Press: Chapel Hill, North Carolina.

Sanders, Wiley Britton. (1968) *Negro Child Welfare in North Carolina*. University of North Carolina Press: Chapel Hill, North Carolina.

Schmidt, John, Kris Warner, and Sarika Gupta. (2010) "The High Budgetary Cost of Incarceration." The Center for Economic and Policy Research, Washington DC.

Sellin, Thorsten. (1928) "The Negro Criminal: A statistical note". *Annals of the American Academy of Political and Social Science.* November: 52-64.

Siedler, Thomas. (2007) "Schooling and Citizenship: Evidence from Compulsory Schooling Reforms," Discussion Papers of DIW Berlin 665, DIW Berlin, German Institute for Economic Research.

Siedler, Thomas. (2010) "Schooling and Citizenship in a Young Democracy: Evidence from Postwar Germany," *Scandinavian Journal of Economics*, 112(2): 315-338.

Silles, Mary A. (2009) "The Causal Effect of Education on Health: Evidence from the United Kingdom," *Economics of Education Review*, 28: 122-128

Smith, James P. and Finis R. Welch. (1977) "Black-White Male Wage Ratios: 1960-70". *The American Economic Review*. 67(3): 323-338.

Solon, Gary and Atsushi Inoue. (2010) "Two-Sample Instrumental Variables Estimators," *Review of Economics and Statistics* 92, August: 557-561.

Stephenson, Gilbert. (1917) "Education and Crime Among Negroes". *South Atlantic Quarterly*, 16: 16-20.

Stratton, John Roach. (1900) "Will Education Solve the Race Problem?" *The North American Review*. 170(523): 785-801.

Western, Bruce, Jeffrey R. Kling, and David Weiman. (2001) "The Labor Market Consequences of Incarceration". *Crime and Delinquency*. 47(3): 410-427.

Washington, Booker T. (1900) "Will Education Solve the Race Problem: A Reply". *The North American Review*. 170(525): 221-232.

Chapter 2: A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration

2.1 Introduction

This paper assembles new panel data to study the assimilation of European migrants in the US labor market in the early twentieth century and the selection of return migrants from this migrant pool. Our findings challenge the conventional wisdom about immigrant assimilation in the Age of Mass Migration (1850-1913). Prior studies, which were based on cross-sectional data, found that immigrants held substantially lower-paid occupations than natives upon first arrival but experienced rapid convergence with natives over time. In contrast, the typical immigrant in our panel data did not face a large initial earnings penalty and moved up the occupational ladder at the same rate as natives.[34] Comparing assimilation patterns in the panel and cross sectional data, we conclude that the apparent convergence in the cross section is driven by a decline in the productivity of immigrant arrival cohorts over time and the departure of negatively-selected return migrants.

Gaining an accurate view of assimilation and selection of return migrants in the Age of Mass Migration is important for our understanding of both the past and the present. First, this era was one of the largest migration episodes in modern history, sizeable enough to influence labor supply and economic development on both sides of the Atlantic. European countries lost up to a third of their population through emigration and, by 1910, 22 percent of the US labor force was

[34] We follow the historical literature and focus on occupation-based earnings because individual earnings were not recorded in population Censuses before the mid-twentieth century. We thus match each individual's recorded occupation to the median earnings in his occupation in 1950. We note that this unavoidable reliance on occupation-based earnings suggests an important limitation: our measure only captures convergence between occupations and cannot speak to within-occupation income convergence.

foreign born (compared with 17 percent today). Second, the US maintained an open border policy for European migrants in this period and had yet to develop a comprehensive welfare state, allowing us to assess how the processes of labor market assimilation would look like in the absence of immigrant selection policies or government support. Finally, beliefs about immigrant assimilation during this period shaped migration policies that are still with us today. The widespread view that migrants from eastern and southern Europe could not assimilate triggered restrictive migration quotas, first passed in the 1920s, that select migrants on the basis of nationality and skill.

Our current understanding of assimilation in the Age of Mass Migration is based on cross-sectional data (Blau 1980; Hatton 1997; Minns 2000). While these studies contributed greatly to our understanding of this era, we illustrate in Figure 1 how inferring assimilation from a cross section is subject to bias due both to differences in the productivity of arrival cohorts and to return migration. The figure normalizes natives' earnings to 100 in all years and depicts four types of migrants. Migrants A and B arrived to the US in 1895 and earned 100 and 80 respectively throughout their stay in the US, while migrants C and D arrived to the US in 1915 and earned 60 and 40 respectively . Furthermore, migrant B is assumed to return to his home country in 1909, illustrating one possible case of negatively-selected return migration.

Now imagine that the researcher only has a single cross section of data, say the 1920 Census. In this case, she will compare migrant A, who arrived in the US in 1895 and remained in the US in 1920, to migrants C and D, who arrived in 1915, and conclude that, upon arrival migrants faced earning penalty relative to natives but completely closed the earning gap after 25 years in the US. However, this conclusion would mistake a combination of arrival cohort differences and negatively-selected return migration for true migrant assimilation (this point was

first made by Douglas, 1919 and was developed by Borjas, 1985).[35] The bias due to differences in arrival cohorts can be avoided if the researcher has access to repeated cross sections, say the 1900 and 1920 Censuses. However, Figure 1 demonstrates that, even with repeated cross sections, inferences on migrant assimilation can still be biased by the process of return migration (this point was made in Jasso and Rosenzweig, 1988). In the 1900 census, the 1895 migrant arrival cohort contains both the high-skilled migrant A and the lower-skilled migrant B. By the 1920 census, migrant B had returned to his home country, leaving only migrant A behind. The apparent increase in wages over time in the 1895 arrival cohort is driven solely by a compositional change in the migrant pool rather than by the true assimilation of those who remained in the US.

Panel data, which follows individual migrants over time in the US, measures immigrant assimilation free from these two forms of bias (see: migrant A in Figure 1). Adopting this approach, we construct a large panel dataset of 24,000 native-born Americans and immigrants from 16 European sending countries by matching men by name, age, and place of birth between the 1900, 1910, and 1920 US Censuses. Assembling such panel data is possible because US Census policy makes complete individual records (including names) publicly available after 72 years. In particular, we match immigrants and US natives from the 1900 Census manuscripts to the 1910 and 1920 Census manuscripts using the genealogy website Ancestry.com.

Moreover, by contrasting the assimilation patterns in the repeated cross-section and panel data, we can infer the selection of return migrants relative to migrants who remained in the US long term. In particular, as is illustrated in the southeast corner of Figure 1, differences in the

[35] In an early paper in this literature, Chiswick (1978) found that immigrants in the 1970 cross section experienced faster wage growth than the native-born and overtook natives within 15 years of arrival. Borjas (1985) demonstrated that, in this period, half of the apparent convergence in a cross section is driven by changes in cohort quality over time.

assimilation profile estimated from repeated cross-section and panel data are due to selective return migration.[36] Importantly, we are also able to compare this indirect approach, in which we infer the selection of return migrants, with a direct approach. Specifically, we compare return migrants in a special supplement to the 1910 Norwegian Census to Norwegian migrants who remain in the US; this direct evidence on the selection of return migrants validates our indirect approach.[37]

Over 25 percent of migrants returned to Europe during this era (Gould, 1980; Bandiera, Rasul and Viarengo, 2010). Return migrants may have been negatively selected, as in the example above, if migrants who were not successful in the US returned home. Furthermore, many migrants in this era employed a deliberate strategy of temporary migration to the New World (Piore, 1980; Wyman, 1996). These temporary migrants may have been negatively selected on observable characteristics if they remained in low-paid occupations during their short sojourn (Dustmann, 1993) or may have been positively selected if productive migrants reached their "target savings" goal faster (Galor and Stark, 1991).

Our comparison of assimilation patterns in cross-sectional and panel data reveals patterns akin to the hypothetical illustration in Figure 1. We find that immigrants in each cross section held lower-paid occupations than natives upon first arrival and appear to completely converge with natives over time. The occupation-based earnings gap between immigrants and natives is

[36] Differences in the assimilation profile in the repeated cross-sections and the panel could also be due to other sources of selective attrition, namely selective mortality or selective name changes. We discuss these possibilities in section III.D.

[37] Our indirect approach of comparing panel and cross sectional data is similar to Lubotsky (2007), who uses Social Security earnings records to build such a panel for the contemporary period. He finds that around 40 percent of the observed convergence between immigrants and natives in repeated cross-sectional data can be attributed to negatively-selected return migration. For other panel analyses of immigrant assimilation in the contemporary period, see Borjas (1989), Hu (2000), Edin, Lalonde and Aslund (2000), Duleep and Dowhan (2002), Constant and Massey (2003), Eckstein and Weiss (2004) and Kim (2011). Zakharenko (2008) provides descriptive evidence that return migrants leaving the US are negatively selected.

only half as large when using repeated, rather than single, cross sections, suggesting a decline in cohort quality. In the panel data, immigrants' initial penalty disappears almost entirely (completely disappearing in some specifications). We conclude that the apparent convergence in a single cross-section is driven by a decline in the quality of immigrant cohorts over time and the departure of negatively-selected return migrants.

Our analysis is also the first to reveal a large degree of heterogeneity in the assimilation and selection of return migrants across sending countries. We show that immigrants from five sending countries, including the English-speaking countries of England, Scotland and Wales, held significantly higher-paid occupations than US natives upon first arrival, while immigrants from other sending countries started out in equivalent or lower-paid occupations. Yet, regardless of starting position, immigrants from all countries experience occupational upgrading at the same rate as natives, thereby preserving the initial gaps between immigrants and natives over time. We explore potential mechanisms for variation in immigrant performance and find suggestive evidence that immigrants from countries that were more economically developed or more culturally similar to the US are those who were most successful in the US labor market.

Finally, we study how the children of immigrants who came during this era performed in the US labor market. Occupational convergence between immigrants and natives may take more than one generation. On one hand, these second generation migrants spoke English better than their parents did and, having grown up in the US, they might have been more exposed to US norms and culture. On the other hand, occupational differences could persist over generations if, for example, second generation migrants grew up in migrant enclaves, inherited skills from their parents, or used their parents' networks to find jobs. We find persistence in occupations across generations rather than extensive assimilation in the second generation: when migrants from a

certain sending country outperformed US natives, so did second generation migrants, and vice versa.

The remainder of the paper proceeds as follows. Section 2 discusses the historical context and related literature. Section 3 describes the data construction and the matching procedures. Section 4 presents our empirical strategy and main results on immigrant assimilation and the selection of return migrants. We estimate the occupation-based earnings penalty (or premium) as well as the earnings distribution for the typical immigrant. Section 5 contains direct evidence on the selection of return migrants from the 1910 Norwegian Census. In Section 6, we show country-by-country results on assimilation and return migration consider possible mechanisms for the cross-country variation in immigrant performance. Section 7 analyzes the performance of second generation immigrants relative to their parents, Section 8 provides a robustness analysis for our main findings, and Section 9 concludes.

2.2 Historical context and related literature

A. Historical context

The US absorbed 30 million migrants during the Age of Mass Migration (1850-1913). By 1910, 22 percent of the US labor force was foreign-born. The foreign-born share of the labor force was even larger outside of the South (29.8 percent), especially in urban areas (38.3 percent).[38] Initially, migrants hailed from countries in northern and western Europe. By 1880, migrant sending countries had shifted toward the poorer regions of southern and eastern Europe (Hatton and Williamson, 1998). Not only were these new immigrants culturally, linguistically and religiously distinct from previous waves, but they were also more likely to be low skilled.

[38] Authors' calculations using the 1910 Integrated Public Use Microdata Series (IPUMS).

For example, in 1900, only 51.2 of Italian immigrants could read and write, compared with 92.7 percent of the German born.[39]

Many native-born residents expressed concerns about the concentrated poverty in immigrant neighborhoods and the low levels of education among immigrant children. Newcomers often lived in overcrowded city tenement buildings with poor ventilation and sanitation (Muller, 1993). Children from immigrant families were more likely than children of the native born to leave school at young ages in order to work in textile factories and other manufacturing industries (Moehling, 1999). Progressive reformers believed that immigrants' behaviors could be changed and championed a series of private initiatives and public legislation, including child labor laws and compulsory schooling requirements, to aid immigrant communities (Lleras-Muney, 2002; Carter, 2008; Lleras-Muney and Shertzer, 2011). Nativist politicians and commentators instead believed that new arrivals would never be able to assimilate into American society (Higham, 1988; Jacobson, 1999).

Concerns about immigrant assimilation prompted Congress to convene a special commission in 1907 to study the social and economic conditions of the immigrant population. The resulting 41-volume report, which was published in 1911, concluded that immigration, particularly from southern and eastern Europe, was a threat to the economic and social fabric of the country. Members of the commission particularly singled out the trend of temporary and return migration as an impediment to assimilation. Two authors of the report, Jeremiah Jenks and W. Jett Lauck, later summarized this view, writing:

> "if an immigrant intends to remain permanently in the US and become an American citizen, he naturally begins at once… to fit himself for the conditions of his new life…If, on the other hand, he intends his sojourn in this country to be short… the acquisition of the English language will be of little consequence… The chief aim of a person with this

[39] Over 70 percent of German immigrants were literate as early as 1850.

> intention is to put money in his purse… not for investment here but for investment in his home country"
>
> (quoted in Wyman, 1996, p. 99-100).

The Immigration Commission report provided fuel for legislators seeking to restrict immigrant entry (Benton-Cohen, 2010). In 1917, Congress succeeded in passing a literacy test (after three prior attempts failed), which required potential immigrants to demonstrate the ability to read and write in any language (Goldin, 1994). In 1924, Congress further restricted immigrant entry by setting a strict quota of 150,000 arrivals per year, with more slots allocated to northern and western European countries.

B. Related literature: Immigrant assimilation in the early 20th century

Immigration to the United States picked up again after passage of the 1965 Immigration and Nationality Act, which not only increased the number of visas granted, but also shifted the emphasis for admission from country-specific quotas to preferences based on immigrant skills and family reunification with US citizens. Within a few years of this historic legislation, a literature emerged in economic history re-assessing immigrant performance in the labor market of the early twentieth century.[40] The earliest studies in this area (re-)analyzed the aggregate wage data published by the Immigration Commission and find that immigrants caught up with the native-born after 10 to 20 years in the US (Higgs, 1971; McGoldrick and Tannen, 1977; Blau, 1980).

A second generation of scholarship examined individual-level wage data from surveys conducted by State Labor Bureaus (Hannon, 1982; Eichengreen and Gemery, 1986; Hanes,

[40] In a related body of work, Ferrie (1997, 1999) measures immigrant assimilation in the Antebellum period. Lieberson (1980) and Alba and Nee (2003) are two core references in the sociological literature on immigrant assimilation.

1996). The first analyses of these sources found substantially lower rates of earnings growth for immigrant workers; in some cases, immigrants appear to have experienced no wage convergence with native workers at all. Although differences between these sources present something of an empirical puzzle, Hatton (1997) argues that this discrepancy is due to specification choice. He re-analyzes the state data with two simple modifications and finds that immigrants who arrived at age 25 fully erased the wage gap with natives within 13 years in the US.[41]

A more recent work on immigrant assimilation incorporates data from the federal Census of Population. Unlike the State Labor Bureau surveys, which are confined to specific industries in particular locations (Michigan, Iowa and California), the Census offers complete industrial and geographic coverage. However, in lieu of individual-level wage data, the Census only contains information on occupation. Relying on the 1900 and 1910 Census cross-sections, Minns (2000) finds partial convergence between immigrants and natives outside of the agricultural sector.[42] Immigrants eliminate 30 to 40 percent of their (between-occupation) earnings deficit relative to natives after 15 years in the US.

Overall, the existing literature suggests that immigrant workers experienced substantial occupational and earnings convergence with the native-born in the early twentieth century. In three different datasets – the Immigration Commission reports, state- and industry-level surveys, and the 1900 and 1910 Censuses – immigrants appear to eliminate between 40 and 100 percent of the earnings gap with natives after 15 years in the US. However, all these analyses compare

[41] In particular, Hatton (1997) allows for differences in the return to experience for younger and older workers and separates immigrants who arrived as children from those who arrived as adults. The convergence figure reported in the text is based on Hatton (1997, Table 4, columns 1 and 3). Because Hatton estimates different returns to experience parameters for immigrants and the native born, the size of the initial wage gap varies by age. For this calculation, we consider an immigrant who arrives at age 25, at which point the implied wage gap with natives is 0.275, a gap which is erased after the immigrant spends 13 years in the US.

[42] Consistent with our results, Minns finds that the full immigrant population actually earn as much as (or more than) natives. The immigrant deficit explored in his paper is present only outside of the agricultural sector.

earnings in a single cross-section, a method that suffers from two potentially important sources of bias: selective return migration, and changes in immigrant cohort quality over time.[43]

2.3 Data and matching

A. Matching men between the 1900, 1910 and 1920 US Censuses

This section describes the construction of our new panel dataset that follows native-born workers and immigrants through the US Censuses of 1900, 1910 and 1920. We restrict our attention to men between the ages of 18 and 35 in 1900, an age range in which men are both old enough to be employed in 1900 and young enough to still be in the workforce in 1920. We further limit the immigrant portion of the sample to men who arrived in the US between 1880 and 1900. For comparability with the foreign born, 95 percent of whom live outside of the South, we exclude native-born men residing in a southern state and all black natives regardless of place of residence.[44]

We identify a sample of men in the base year (1900) from two Census sources. For large sending countries (listed in Table 1, panel A), we rely on the 1900 5 percent Integrated Public Use Microdata Series (IPUMS) (Ruggles, 2010) to find immigrants from large sending countries and to randomly select a sample of 10,000 native-born men. To ensure a sufficient sample size for smaller sending countries (Table 1, panel B), we instead compile the full population in the

[43] We note that Minns (2000) acknowledges the potential bias from changes in the quality of immigrant arrival cohorts. Hatton (1997) partially addressed the shift in sending countries by separately analyzing assimilation profiles by country of origin for three sending countries (Britain, Ireland and Germany).

[44] We also tried including native-born men living in the South into the sample. Because men who live in the South held lower-paid occupations, the immigrant earnings premium increases by around $1,000 in both the repeated cross-section and the panel. Yet the extent of convergence in both samples and the comparison between immigrants in the cross-section and panel (relative to natives) is preserved.

relevant age range in 1900 from the genealogy website Ancestry.com. Altogether, we identify immigrants from 16 sending countries.[45]

We search for viable matches for these men in 1910 and 1920 using the iterative matching strategy developed by Ferrie (1996) and employed more recently by Abramitzky, Boustan and Eriksson (forthcoming) and Ferrie and Long (2011). Figure 2 illustrates our matching procedure by showing one observation in our dataset. The Census manuscript of 1900 reveals that Alexander James was born in Wales in 1871 and moved to the US in 1893. In the US, Alexander worked as a coal miner in 1900. Based on his name, age, and country of birth, we find Alexander James in the 1910 Census. He was still working as a miner. When we find Alexander again in 1920, he had become a foreman, i.e. he had moved up the occupational ladder.

More formally, our matching procedure proceeds as follows:

(1) We begin by standardizing the first and last names of men in our 1900 samples to address orthographic differences between phonetically equivalent names using the NYSIIS algorithm (see Atack and Bateman, 1992). We restrict our attention to men in 1900 who are unique by first and last name, birth year, and place of birth (either state or country) in our sample. We do so because, for non-unique cases, it is impossible to determine which of the records should be linked to potential matches in 1910 and 1920. Table 1 presents information about the number of potential matches by country.
(2) We identify potential matches in 1910 and 1920 by searching for all men in our 1900 sample in the 1910 and 1920 Census manuscripts available from Ancestry.com. The Ancestry.com search algorithm is expansive and returns many potential matches for each case, which we cull using the iterative match procedure described in the next step.[46]

[45] We include men from all European sending countries with at least 3,000 migrants living in the US in 1900, with the exception of Poland, Czechoslovakia and the Netherlands, which made the cut but were nevertheless excluded. Individuals born in Polish or Czech territory were allowed to report these locations as their place of birth on the 1900 Census. This option was removed from the 1910 Census and then restored in 1920 after both countries gained their independence in 1918. Migrants from the Netherlands reported varied birthplaces on the Census, rendering it difficult to find these individuals in Ancestry.com and follow them over time.

[46] The Ancestry.com search engine aims to maximize potential 'hits' under the assumption that individual users can identify their relatives from a longer list by hand. To this end, it uses many approaches to convert names into their phonetic equivalents and applies a very lax matching rule. For small sending countries, we instead match the complete 1900 population to the complete 1910 and 1920 populations obtained from Ancestry.com.

(3) We match observations forward from 1900 to either the full population (for small countries) or to the set of potential matches (for large countries) in 1910 and 1920 using an iterative procedure. We start by looking for a match by first name, last name, place of birth (either state or country) and exact birth year. There are three possibilities: (a) if we find a *unique* match, we stop and consider the observation "matched"; (b) if we find multiple matches for the same birth year, the observation is thrown out; (c) if we do not find a match at this first step, we try matching within a one-year band (older and younger) and then with a two-year band around the reported birth year; we only accept unique matches. If none of these attempts produces a match, the observation is discarded as unmatched.
(4) After matching each sample in 1900 separately to 1910 and 1920, we create our final dataset by restricting to men who were located both in 1910 and 1920.

The second and third columns in Table 1 present match rates and final sample sizes for each sending country and for native born men. Our matching procedure generates a final sample of 22,070 immigrants and 1,891 natives. We can successfully match 19 percent of all native-born men forward from 1900 to both 1910 and 1920. For the foreign born, the average match rate across countries is lower (10 percent), which is expected given that a sizeable number of migrants return to Europe between 1900 and 1920. These double match rates are similar to those in Ferrie (1996) and Abramitzky, Boustan and Eriksson (forthcoming).

B. Occupation and earnings data

We observe labor market outcomes for our matched sample in 1900, 1910 and 1920. Because these Censuses do not contain individual information about wages or income, we assign individuals the median income in their reported occupation.[47] Table 2 reports the ten most common occupations for our sample of matched natives and foreign born workers. Although the top ten occupations are similar for both groups, migrants to the US were less likely to be farmers

[47] For observations taken from the 1900 IPUMS (the native born and immigrants from large sending countries), we use the occupation recorded in the digitized micro data. For the remaining countries in 1900 and for all countries in 1910 and 1920, we collect the occupation string by hand from the historical manuscripts on Ancestry.com. We then standardize occupation titles to match those identified in the 1900 IPUMS. Our final sample has 1,426 native-born men and 18,249 immigrants with non-missing occupation data.

(18.1 versus 24.8 percent) and more likely to be mine operatives (3.3 versus 1.4 percent). The native born were more likely to be salesmen and clerks, two occupations with high returns to fluency in English. Other common occupations in both groups include managers, operatives, and general laborers.[48]

Our primary source of income data is the "occupational score" variable constructed by IPUMS. This score assigns to an occupation the median income of all individuals in that job category in 1950. For ease of interpretation, we convert this measure into 2010 dollars. Using this measure, our dataset contains individuals representing around 125 occupational categories. Our unavoidable reliance on median earnings by occupation prevents us from measuring the full convergence between immigrants and natives. In particular, we are able to capture convergence due to advancement up the occupational ladder (between-occupation convergence), but we cannot measure potential convergence between immigrants and natives in the same occupation.[49] A further concern with the IPUMS 'occupation score' variable is its anchoring to occupation-based earnings in the year 1950. The 1940s and 1950s was a period of wage compression (Goldin and Margo, 1992). If immigrants were clustered in low-paying occupations, the occupation score variable may understate both their initial earnings penalty and the convergence implied by moving up the occupational ladder. We address this concern by using occupation-based earnings from the 1900 Cost of Living survey as an alternative dependent variable.

[48] Men who were not employed at the time of the survey reported their last-held occupation. 1910 was the only census in our time period to ask about unemployment. In that year, native-born men of native parentage (age 18-60) had an unemployment rate of 4.4 percent, while 5.7 percent of foreign born were unemployed. This differential unemployment likely contributed to the true earnings gap between immigrants and natives.

[49] We use the 1970 IPUMS to assess the share of total wage convergence between immigrants and natives that takes place between versus within occupational categories. The 1970 Census is the first to record both wage data and year of immigration or years spent in the US for the foreign born. We exclude immigrants who arrived after 1965, the year of major immigration policy change. Immigrants experience 9 log points of total wage convergence relative to natives after spending 30 years in the US and 3 log points of convergence when using an occupation-based measure of earnings. We conclude that our method is likely to capture around one third of total wage convergence between immigrants and natives. We note that this exercise is subject to all the problems we mentioned previously of inferring convergence from a cross sectional data.

C. Comparing matched samples with the full population

Our matched sample may not be fully representative of the immigrant and native born populations from which they are drawn. In particular, men with uncommon names are more likely to be successfully linked between Censuses, and the commonness of one's name could potentially be correlated with socio-economic status. We assess this possibility by comparing men in the cross-sectional and panel samples in 1920. By definition, men in both the panel and repeated cross-sections must have survived and remained in the US until 1920. Thus, by 1920, up to sampling error, any difference between the panel and the repeated cross-sections is due to an imperfect matching procedure.

Table 3 compares the mean occupation score of men in our cross-section and panel samples in 1920. We consider natives and the foreign born separately and re-weight the matched sample to reflect the distribution of country of origins in the 1920 population.[50] Immigrants in the matched sample slightly out-earn their native counterparts by 1920 ($23,500 vs. $23,200). Among natives, the difference in the mean occupation score in the matched sample and the population in 1920 is small ($53) and statistically indistinguishable from zero. In contrast, immigrants in the matched sample have a $369 advantage over immigrants in the representative sample. Country-by-country comparisons reveal that this gap is generated by five sending countries: Belgium, France, Ireland, Italy and Norway. Results are robust to dropping these five countries from the analysis. Overall, we have little concern that the matching procedure generates a systematic bias.

[50] We need to re-weight the matched sample because our universe of potential matches is drawn from 5 percent samples for large countries and from 100 percent samples for smaller countries. We weight according to the 1920 cross-section to reflect the fact that migrants in the panel sample remain in the US until 1920.

D. Other sources of selective attrition

We infer selection of return migrants relative to migrants who settled in the US long term indirectly, by comparing occupational upgrading patterns in the repeated cross-section versus the panel data. Any difference between the panel and the repeated cross-sections is due to selective attrition from the cross-sections, which is arguably mostly due to selective return migration.[51] However, any form of selective attrition from the repeated cross-sections (such as selective mortality) could drive differences between the panel and the repeated cross-sections.

Selective mortality is not a likely concern. Mortality in 1900 for this age group (ages 15-45) was fairly low and fairly uniform across sending countries. The Irish were slightly more likely to die (8 per 1000) and the Russian were slightly less likely to die (3 per 1000), but mortality among people from other nationality and US natives were all around 5-6 per 1000 (figures by Marriam, 1903, based on 1900 Census). Furthermore, note that, while selective mortality is a potential concern for both native- and foreign-born men, selective return migration is not an issue for the native born; few native-born men emigrated away from the US. Therefore, one way to test for the presence of selective mortality in our sample is to compare the occupation-based earning patterns of native-born men in the repeated cross-section versus the panel data. We find that the occupation-based earnings of natives are similar in the repeated cross-sections and the panel in all years, suggesting that selective mortality is a non-issue (at

[51] During this period, some immigrants engaged in circular migration, migrating to the US and returning to Europe multiple times (Piore, 1980; Wyman, 1996). Circular migrants will enter the panel sample only if they happen to live in the US on the Census years; otherwise, they will be treated as temporary migrants.

least for the native born).[52] We note that this test for selective mortality relies on the assumption that native- and foreign-born men were subject to the same mortality process.

Selective name changes by immigrants are also not a likely concern. First, most name changes occurred upon entry to the US and were processed by state or federal officials (for example, at Ellis Island). Any such change would have taken place before we first observe migrants in the 1900 Census and would thus affect neither the panel nor the cross-sectional data. Second, men who changed their name between Censuses are not likely to affect the results. To see this, note first that even though such men would never be included in the panel sample, they would stay in the repeated cross-sections before and after their name change. We can thus test whether migrants in the repeated cross-section are less likely to have a "foreign" name (an indication that they may have changed their name).[53] Indeed, we find that foreign-born men in the panel sample have slightly more "foreign" names than their foreign-born counterparts in the cross-section, which is consistent with the fact that men who change their name after arriving in the US do not enter the panel. Yet the small difference in the "foreignness" index is associated with only a $60 difference in occupation-based earnings (in 2010 dollars) and so is not quantitatively large enough to affect the results.

2.4 Immigrant assimilation in panel data

A. Estimating equation

[52] We regress occupation-based earnings score on a dummy for being in the panel sample for the native born. In 1900, for example, the coefficient on this dummy variable is -0.212 (s.e. = 0.294). After adjusting for age differences between the two samples, the difference falls further to -0.130 (s.e. = 0.288). This finding is consistent with the presence of a minimal relationship between socio-economic status and health in the early twentieth century (Frank and Mustard, 1994; Hummer and Lariscy, 2011).

[53] The "foreignness" index is constructed by first calculating the probability of being foreign born conditional on having a given first name (and, separately, a given last name) in the 1900-20 IPUMS samples. The "foreignness" index is then the sum of the two probabilities; the index varies between zero and two. Foreign-born men in the cross-section (panel sample) have an index value of 1.13 (1.23).

Our main analysis compares the occupational mobility of native-born and immigrant workers. We estimate:

$$Occupation\_score_{ijmt} = \gamma_{t-m} + \lambda_m + \eta_t + \alpha_j + \beta_1 Age_{it} + \beta_2 I\left[Age_{it} \geq 35\right] + \beta_3 Age_{it} I\left[Age_{it} \geq 35\right] + \varepsilon_{ijmt} \quad (1)$$

where *i* denotes the individual, *j* denotes the country of origin, *m* is the year of arrival in the US, *t* is the (Census) year, and *t-m* is thus the number of years spent in the US.[54] *Occupation score* is a proxy for labor market earnings that varies between (but not within) occupations. The coefficients β_1 through β_3 relate years of labor market experience to the worker's position on the occupational ladder. Following Hatton (1997), we allow the slope of the experience profile to vary by age to account for steep returns to labor market experience for young workers in the early twentieth century.

The vector γ_{t-m} separates the foreign-born into five categories according to time spent in the US (0-5 years; 6-10 years; 11-20 years; 21-30 years; 30 or more years). Equation 1 includes a dummy variable for each time interval, with the native born constituting the omitted category.[55] The sign and magnitude of the coefficient on the first dummy variable (0-5 years) indicates whether immigrants received a premium upon arrival to the US, whereas the difference between this indicator and the remaining dummy variables reveal whether immigrants eventually catch up with or surpass the earnings of natives. Our main specification divides the foreign born into two year-of-arrival cohorts (pre-1890 arrivals versus those who arrived after 1890) to allow for differences in earnings capacity by arrival cohort (Section IV.C explores the sensitivity of the results to the choice of the number of arrival cohorts).

[54] In contrast to the existing literature, we include country fixed effects in all specifications. As a result, we do not rely on variation in typical sending countries across arrival years but instead compare immigrants from the same country of origin who arrive in different years.

[55] The rates of convergence for immigrants in the cross-section and the panel are similar if, instead, we estimate a quadratic in years spent in the US (see online Appendix A).

We begin by estimating a cross-section version of equation 1 with data from the 1900, 1910 and 1920 IPUMS and Census year fixed effects (results are similar when looking at each single cross-section separately, see online Appendix B). We then add the arrival cohort dummy (λ_m) to estimate the repeated cross-section model. The coefficient on the arrival cohort dummy indicates changes in cohort quality over time. Comparing the cross-section and the repeated cross-section allows us to infer how much of the initial occupational penalty can be attributed to differences in the quality of arrival cohorts.

We next compare the repeated cross-section with the panel. The repeated cross-sectional data follows arrival cohorts, rather than individuals, across Censuses. Therefore, comparing the estimates in the repeated cross-sections and panel data allow us to infer whether and to what extent return migrants were positively or negatively selected from the immigrant population. In 1900, the cross-sectional data includes both temporary and permanent migrants. Over time, the temporary migrants return home, leaving only permanent migrants in the cross-section by 1920. In contrast, the panel is restricted to permanent migrants in all years. If we observe more (less) convergence in the cross-section than in the panel, we can infer that the temporary migrants are drawn from the lower (upper) end of the occupation-earnings distribution, thereby leading their departure to increase (decrease) the immigrant average.

In particular, we estimate a single regression that pools the 1900-1920 cross-sections with the matched panel sample. We allow the variables of interest, including the arrival cohort (λ_m) and years spent in the US (γ_{t-m}) fixed effects, to have separate coefficients in the cross-section and panel samples.[56] As before, we reweight observations in the panel sample by country of birth to be representative of the full population, both native- and foreign-born, in 1920.

[56] Note that, by pooling the two data sources, we constrain the year, country of origin, and age effects to be common across the two samples. Results are similar when we run equation 2 separately for the panel and the repeated cross

We emphasize that measuring the assimilation process requires a panel dataset that follows individuals over time. With the panel sample, we can estimate an assimilation profile for permanent migrants – defined here as migrants who remain in the US for at least 20 years – who demonstrate continued participation in the US labor market (specifically migrants who migrated between 1880 and 1900 and remained until 1920). The repeated cross-sections instead allow us to observe occupational outcomes for the typical migrant in the US labor market, consisting of both permanent migrants and temporary migrants who will later return to their home country. These patterns are of interest in themselves because they represent the experience of the average migrant in the US at a point in time.

B. Occupational convergence in cross-section and panel data

In this section, we estimate equation 1 with the full sample of immigrant and native-born workers. We show that: (1) In the cross-section, immigrants initially hold lower-paid occupations but converge upon natives over time. (2) Following arrival cohorts from 1900 to 1920 in the repeated cross-sections lowers the initial migrant disadvantage. (3) Permanent immigrants (as represented by the panel data) hold somewhat higher-paid occupations than natives upon first arrival and experience similar occupational upgrading over time. We conclude that the apparent immigrant disadvantage in a single cross-section is driven by the lower quality of later arrival cohorts (1900 versus 1880) and the negative selection of temporary migrants who eventually return to Europe.

Table 4 presents estimates of equation 1 for the cross-section, the repeated cross-sections and our newly-constructed panel sample. The coefficients on the "years in the US" dummy

---

section or when we restrict the arrival cohort effects to be the same in both samples (results are shown in online Appendix C).

variables indicate the gap between immigrants of a given vintage and the native born. In the cross-section, new immigrants hold occupations that earn $1200 in 2010 dollars below natives of similar age and appear to completely make up this gap over time (column 1). Columns 2 and 3 report coefficients from a specification that pools the cross-section and panel. In the repeated cross-section, immigrants who arrived after 1890 had significantly lower occupation-based earnings than did earlier arrivals, receiving an arrival cohort penalty of $750. Thus, simply by controlling for arrival cohort in column 2, the occupation score gap between recently-arrived immigrants and natives shrinks to $300. In other words, *even within sending countries*, around three-quarters of the initial gap in the pooled cross-section is due to the lower occupational skills of immigrants who arrived after 1890. In the repeated cross-section, immigrants again appear to completely close this (smaller) occupation gap with natives after spending time in the US.

Coefficients for the panel data are reported in column 3. In this subsample, we find *no initial occupation score gap* between immigrants and natives. If anything, immigrants start out $300 ahead of natives, although this difference is not statistically significant. Comparing the two samples suggests that the initial earnings gap in the repeated cross-section is capturing the negative selection of immigrants who end up returning to Europe (temporary migrants).

The differences in the initial immigrant-native gaps and implied rates of convergence between the cross-section and panel samples are underscored in Figure 3. This figure graphs the coefficients on the five ‘years in the US’ dummy variables in the pooled cross-section, the repeated cross-sections and the panel dataset. In graphical form, it is even easier to see that, in the cross-section, immigrants face an occupation score gap relative to natives upon first arrival, but are able to erase this gap over time. In contrast, immigrants in the pre-1890 arrival cohort arrived with a much smaller occupation score gap relative to natives. Finally, permanent

immigrants in the panel data hold somewhat higher-paying occupations than do natives, even upon first arrival, and retain this slight advantage over time.

In section VIII, we explore a series of alternative specifications and measures of occupation-based earnings. Overall, we show that specification choice has little effect on either the estimated rate of convergence or the implied selection of return migrants. Adjusting the occupation-based earnings measure increases the initial earnings gap that immigrants face, sometimes substantially, but has little effect on either estimated convergence or selection of return migrants.

C. Earnings distribution of natives and immigrants

On average, immigrants earn less than natives upon first arrival in the cross-section. Table 5 reports percentiles of the earnings distribution for natives and for recent immigrants (those who arrived within the past 10 years) in both the cross-section and panel samples.

The earnings distribution reveals why immigrants in the cross-section face an earnings penalty relative to natives upon first arrival. Although immigrants earn more than natives at the low-end of the earnings distribution, natives hold higher-paid occupations than immigrants at the 75^{th}, 90^{th} and especially the 99^{th} percentiles. The weight placed on the high-end of the earnings distribution in the levels specification explains why the immigrant occupation penalty in levels becomes an occupational premium in logs.

The better performance of long term immigrants relative to natives even upon first arrival is also apparent in the earnings distribution. Permanent immigrants out-earn natives at percentiles below the median. By extension, one can readily see the earnings advantage of permanent immigrants relative to the total migrant pool. Permanent immigrants have a smaller

left-tail in the earnings distribution, earning more than the typical migrant at the 10th percentile of the occupation-based earnings distribution.

2.5 Direct evidence on Norwegian return migration

Thus far, we have inferred the selection of return migration to Europe indirectly, by comparing cross-section and panel data. This section directly examines the selection of men who returned from the US to Norway, and compares the direct and indirect evidence on selection of return migrants in the case of Norway. Return migration was sufficiently high that the 1910 Norwegian Census added a supplement for individuals who had spent some time in the US. Return migrants were asked to report the date on which they left for US and the date on which they returned, as well as the occupation they held in the US. We use these data to compare the occupational distribution of Norwegian migrants who stayed in the US with those who returned to Norway.

In the 1910 Norwegian Census, occupations are coded according to the Historical International Standard Classification of Occupations (HISCO). For comparison, we convert these values into US Census occupation codes and then into 1950 income. We focus on men between the ages of 18 and 55 in 1910 who migrated to the US between 1880 and 1900 and who returned to Norway between 1900 and 1910 (if they returned). We observe the occupations held in the US by return migrants in the year before their return (sometime between 1900 and 1910), and the occupations of Norwegian migrants who stayed in the US in the 1910 US Census.

We pool 957 migrants in the US and 3,100 return migrants in Norway and regress occupation-based earnings on a dummy for having returned to Norway and a polynomial in age. The coefficient on being a return migrant is -$1659 (s.e. = 225) in 2010 dollars. In other words,

return migrants held lower-paid occupations than migrants who remained in the US. This magnitude is remarkably similar to our inference on the extent of negative selection among return migrants generated by comparing cross-section and panel (-$1757, Figure 6 below).

2.6 Heterogeneity by sending country

A. Variation in convergence and selectivity of return migration

The typical permanent immigrant in the panel sample holds a slightly higher-paid occupation than the average native, even upon first arrival. However, this pattern masks substantial heterogeneity across sending countries. Figure 4 illustrates cross-country variation in the occupation-based earnings of immigrants relative to the native born, both upon first arrival and after 30 or more years in the US. The grey bars indicate that six of the 16 countries in the current sample hold occupations that pay significantly less than those held by the native born upon first arrival. The size of this occupation-based earnings penalty varies from $1000 (Finland) to $4000 (Portugal) in 2010 dollars. In contrast, immigrants from three English-speaking countries (England, Scotland and Wales), a developed country in Western Europe (France) and one country from the new immigrant stock (Russia) arrived with statistically-significantly more occupation-based skill than the typical native-born worker. The remaining five countries exhibit little difference in earning power relative to natives (Austria, Germany, Ireland, Italy and Sweden).[57] We explore possible explanations for these cross-country differences in Section VI.

[57] Consistent with Table 5, panel H, the average relative immigrant earnings by country declines when we include state-by-urban fixed effects. However, the order and statistical significance of the country-specific immigrant earnings penalties are primarily robust to include state-by-urban fixed effects. Exceptions are the earnings premium of Russian and Scottish migrants, which disappears with these added controls, and the neutral earnings of Irish immigrants, which becomes a $1,700 earnings penalty (in 2010 dollars).

The black bars in Figure 4 demonstrate that, on the whole, permanent immigrants experience little occupational growth relative to natives after spending time in the US. That is, permanent immigrants move up the occupational ladder at the same pace as the native born. Migrants from eight countries experience between $500 and $1000 of convergence relative to natives over this period, while migrants from seven countries actually experience up to $1000 of divergence relative to natives. We note that none of these patterns are statistically significant. Immigrants from Finland are the only group that exhibits a statistically-significant amount of divergence, falling further behind natives by over $2000 from their year of first arrival.

The average immigrant in the cross-section and panel samples differ both because of declines in arrival cohort quality and negatively-selected return migration. Figures 5 and 6 report variation in the direction and magnitude of these two biases by country-of-origin. We begin by estimating a version of equation 1 with four arrival cohorts (see Table 5, Panel B). Figure 5 reports differences by country between immigrants who arrived between 1880 and 1884 and those who arrived between 1895 and 1900. Countries like Russia and Italy whose immigration waves only began in large numbers in the early 1880s are among those with the largest decline in immigrant arrival cohorts over this period, perhaps because positively-selected "pioneer" migrants are replaced by the more typical migrant over time. However, old immigrant groups like the English and the Irish experience large declines in arrival cohort quality as well during this time.

Figure 6 explores heterogeneity in the implied selection of return migrants by sending country. In particular, we report the difference between immigrants' occupational upgrading relative to natives in the cross-section versus the panel sample by sending country; recall that a negative value indicates that return migrants are negatively selected. The figure reveals

statistically-significant negative selection in the return migration flow back to five sending countries (England, Italy, Norway, Russia and Switzerland) and significant positive selection to one country (Finland). The return migrant flow to the remaining ten countries is neutral.[58]

Russia is a particularly interesting case. Figure 4 shows that Russian migrants performed well in the US upon first arrival and Figure 6 suggests that return migrants to Russia were particularly negatively selected. These patterns can be explained by the ethnic composition of the Russian migration. The Russian migrant flow is made up of two groups, Jews and non-Jews, who were primarily Poles and other non-ethnic Russians. The Jewish immigrants were both higher skilled and less likely to return to Russia than their non-Jewish counterparts (Perlmann, 1999). In fact, only 7.1 percent of Russian Jews returned to Europe compared with 87 percent of Russian non-Jews (Gould, 1980). Therefore, the return migrant flow is made up primarily of low-skilled non-Jewish Russians.

B. Explaining cross-country variation in immigrant performance

Figure 4 documents substantial variation in the performance of immigrants from different sending countries in the US labor market. This section explores the relationship between economic and cultural characteristics of source countries and the initial earnings penalty (or premium) that immigrants from these countries face in the US. We emphasize that, because of the small sample size (16 countries) and lack of exogenous variation, these relationships are

[58] The height of the bars in Figure 6 represent the product of the selectivity of return migrants and the return migration rate. We use return migration rates by country reported either in Gould (1980) or in Bandiera, Rasul and Viarengo (2011) to back out the actual selection term. Gould (1980) reports return migration rates for Russian Jews and non-Jews separately (7.1 percent and 87 percent); we use the weighted average. Because there is little cross-country variation in the rates of return migration, the resulting picture is nearly identical to the pattern reported in Figure 6 in both cases (not shown). The one exception is that return migrants to Russia look even more negatively selected when we use the Bandiera, et al. (2011) return migration rates.

merely suggestive. Nonetheless, it is interesting to document the source country characteristics that predict success in the New World.

In particular, we regress the earnings penalty (or premium) of recently-arrived immigrants relative to natives on a set of economic characteristics for the sending country in 1880 and on measures of the linguistic, cultural and religious difference between the source country and the US. Results are reported in Table 6.

We begin in column three by regressing the earnings penalty on each sending country characteristic one-by-one. Immigrants from countries with a higher share of the labor force working in agriculture or a lower real wage hold lower-paid occupations relative to natives when they arrive in the US. Residents of these poorer, more agricultural countries may develop fewer skills in their native countries; alternatively, immigrants from these countries could be negatively selected from the sending population. In contrast, immigrants from countries that share a language, cultural background or religious affiliation with residents of the US are more successful in their new destination. Hailing from a country with a similar culture could help immigrants assimilate in the US or may prevent them from facing discrimination in the labor market. Population pressure and health conditions in the source country, as measured by the rates of natural increase and of infant mortality, have no relationship with subsequent immigrant outcomes.

With only 16 country-level observations, we have limited degrees of freedom. Yet in columns 4 and 5, we supplement each of the estimating equation with either the strongest economic or the strongest cultural characteristic. That is, we regress the immigrant earnings penalty on each country characteristic and either the share of the labor force in agriculture (column 4) or the cultural and religious distance between residents of the country and the US

(column 5). We find that cultural and religious distances are the most robust predictors of labor market performance upon first arrival.

2.7 Second generation migrants in the US labor market

Occupational convergence between immigrants and natives may take more than one generation. On the one hand, second generation migrants were educated in the US and, therefore, were likely fluent in English and may have been exposed to US norms and culture. On the other hand, occupational differences could persist over generations if, for example, second generation migrants grew up in migrant enclaves or inherited occupational skills from their parents.

We compare the occupation-based earnings of US-born men whose parents were born abroad to US-born men whose parents were born in the US (hereafter referred to as US natives, even though second generation immigrants are also born in the US). Because Census records are not publicly available, we are unable to construct panels for this era. Instead, we use the 1% IPUMS samples of the US Census from 1900-1950 to compare the children of first generation immigrants from various sending countries to their parents' generation and to US natives.[59] In particular, we define two samples of non-Southern males aged 20 to 60. For first generation immigrants, we use the Censuses of 1900 to 1920 to compare foreign-born men with US natives. Second generation immigrants, defined as men with two parents from the same country of origin, are drawn from the Censuses of 1920 to 1950 and compared with US natives in those years.

We estimate the following age-earnings profile separately for each group and for each country of origin: immigrants (first generation), US natives in the same Censuses and ages as the

[59] Note that for second generation migrants and US natives, assimilation patterns will not be biased by return migration. In addition, the first generation migrants in this analysis must have stayed in the US for at least twenty years and thus are unlikely to include men who will later return to Europe.

immigrants, sons of immigrants (second generation), US natives in the same Censuses and ages as the second generation sample:

$$Y_{it} = \left\{ \begin{array}{l} \alpha_t + \beta_a Age_{it} + \beta_{a2} Age_{it}^{\ 2} + \beta_{a3} Age_{it}^{\ 3} + \beta_{a4} Age_{it}^{\ 4}{}_i + \\ Migrant_i * \sum_k \gamma_k YearsIUS_{itk} + \varepsilon_{it} \end{array} \right\} \quad (2)$$

As before, our outcome variable is occupation-based earnings converted to 2010 dollars. In Figure 7, we illustrate the results from equation 2 for a person who is 25 years old in either 1910 (first generation versus natives) or in 1930 (second generation versus natives). We assume the first generation migrant moved to the US in 1890.[60]

Figure 7 suggests strong evidence of persistence across generations. If the first generation immigrants out-performed natives (England, Scotland, Wales, France, Italy, Austria, Germany, Russia), so did the second generation and vice versa (Norway, Portugal). A notable exception is Finland, in which first generation migrants held lower-paid occupations but second generation migrants held higher-paid occupations. Consistent with Borjas (1994), there is evidence of convergence between natives and immigrants across the immigrant generations, although this convergence is slow for most countries.

2.8 Alternative specifications and earnings measures

We return in this section to our main analysis, which compared the age-earnings profile of immigrants and natives from 1900 to 1920. Table 7 assesses the sensitivity of our findings to a series of alternative specifications and substitute measures of occupation-based earnings.[61] The

[60] Results are robust to alternative specifications of the age effects.

[61] For brevity, we show only the main coefficients on the "years in US" indicators for the repeated cross-section and panel samples. Online Appendix D contains graphs for each of these specifications in the format of Figure 2.

first section of Table 7 considers alternative specifications for equation 1. In Panel A, we omit the country-of-origin fixed effects, thereby estimating the assimilation profile using variation that occurs both within and between sending countries. In this case, permanent immigrants fare even better, earning $600 more than natives upon first arrival. This modification does not alter the comparison of permanent and temporary immigrants nor the degree of convergence observed in each sample. Panel B includes indicators for a series of finer arrival cohorts (arrived between 1886-1890; 1891-1895; 1896-1900; arrival before 1885 is the omitted category). These controls eliminate the initial earnings gap between migrants and natives in the repeated cross-section; yet, permanent immigrants continue to earn $600 more than typical migrant upon first arrival. In Panel C, we interact the country-of-origin fixed effects with the arrival cohort dummy. The premium earned by permanent immigrants is preserved.

The next section of Table 7 introduces alternative dependent variables. Panel D uses the logarithm of our occupation-based earnings measure. In this case, immigrants in both the cross-section and the panel out-earn natives upon first arrival, by 5 percent and 9 percent respectively. Permanent migrants maintain their advantage relative to the total migrant pool, again consistent with negatively-selected return migration. Differences between the logarithm and levels specifications are driven by the concentration of natives at the top end of the occupation-based earnings distribution; these lucrative occupations are more heavily weighted in the levels specification. The next sub-section discusses the earnings distribution of migrants and natives in more detail.

Our main dependent variable is based on occupation-based earnings in 1950. In that year, farmers earned incomes below the median, whereas, in 1900, farming was a relatively high-paid occupation. To examine the sensitivity of the results to farmers' earnings, Panel E arbitrarily

raises farmers' income by 20 percent. Because natives were more likely than immigrants to be owner-occupier farmers, this modification increases the immigrant earnings penalty. However, the degree of convergence and the comparison between immigrants in the cross-section and panel are unchanged. Panel F replaces the 1950-based earnings measure with mean earnings by occupation from the 1900 Cost of Living survey.[62] Given the greater income inequality in 1900 and the concentration of immigrants in the lower half of the income distribution, the initial gap between immigrants and natives is substantially larger in this specification (between $2,700 and $3,200 in 2010 dollars). Immigrants in both the cross-section and matched panel samples experience more convergence relative to natives with time spent in the US (around $1000). Nevertheless, as before, there is a substantial gap between permanent immigrants and the total immigrant pool, suggesting that return migrants were negatively selected.

The final section of Table 7 presents estimates of equation 1 that address aspects of the migration decision. Panel G excludes the 20 percent of the migrant sample who arrived in the US before the age of 10.[63] Young immigrants may experience systematically different rates of assimilation due to heightened fluency in English or education in the US school system (Friedberg, 1993; Hatton, 1997; Bleakley and Chin, 2010). We find similar results to the full sample when we exclude child immigrants.

Panels H and I introduce state fixed effects and interactions between state fixed effects and an indicator for living in an urban area. The state to which a migrant moves is a choice, and so including state fixed effects raises concerns of endogeneity. However, these specifications

[62] We note that this measure has several disadvantages relative to our main measure based on the 1950 occupation score. First, the Cost of Living surveys were not nationally representative but instead focused on urban married households. Second, income in the surveys is missing for a number of occupations (including farmers, which we instead have to infer from the US Census of Agriculture).

[63] We choose the age of 10 because it is an age at which most people did not work, even in this historical period. Results are similar at cutoffs of age 12 or 14 as well.

may shed light on the mechanism underlying the earnings difference between immigrants and natives. First, immigrants may achieve earnings parity with natives by moving to locations with an industry mix conducive to high-paid occupations (Borjas, 2001). Second, immigrants may earn the same nominal wage as natives but face lower real wages if they settle in more expensive states or urban areas. Adding location fixed effects results in a larger initial gap between immigrants and natives (around $1500), suggesting that the earnings parity in the main specification is achieved largely through location choice. As before, we find a gap between permanent immigrants and the total migrant pool, consistent with negative selection of return migrants, as well as a similar amount of convergence between immigrants and natives.[64]

2.9 Conclusion

We construct a new panel dataset of native- and foreign-born men in the US labor market at the turn of the twentieth century, an era in which US borders were open to all European migrants. This Age of Mass Migration is not only of interest in itself, as one of the largest migration waves in modern history, but is also informative about the process of immigrant assimilation in a world without migration restrictions. Most of the previous research on this era relies on a single cross-section of data and finds that immigrants started with lower-paid occupations than natives but caught up with natives after spending some time in the US.

In our panel dataset, we instead find that immigrants who remained in the US did not hold lower-paid occupations than US natives, even upon first arrival, and moved up the occupational ladder at the same rate as natives. We conclude that the apparent convergence in a single cross-section reflects a substantial decline in the quality of migrant cohorts over this

[64] In online Appendix D, we also graph the implied effects of years spent in the US from a specification that allows immigrants to have their own age-earnings profile. Results are qualitatively similar to those in the main specification.

period as well as a change in composition of the migrant pool as negatively-selected return migrants leave the US over time. Our paper further demonstrates the importance of accounting for differences in migration patterns across sending countries. Permanent migrants from countries with cultural proximity to the US performed far better than natives upon first arrival, while permanent migrants from other countries performed far worse. Yet immigrants from all countries, regardless of their starting position, experienced little occupational convergence with natives. Moreover, we find that these differences in performance across sending countries are persistent across generations.

Contemporaries questioned the ability of European immigrants to assimilate in the US economy and called for strict migration restrictions that favored countries with highly-skilled residents. Our results indicate that these concerns were unfounded: the average permanent immigrant in this era arrived with skills similar to those of natives and experienced identical rates of occupational upgrading over their lifecycle. These successful outcomes suggest that migration restrictions are not necessary to ensure migrant assimilation. At the same time, we also note that migrants that arrived with low skill levels did not manage to close their skill gap with natives over time. This finding undercuts the commonly-held view that, unlike today's migrants, past waves of European immigrants, even those who arrived without the ability to read or to speak English, were able to quickly catch up with natives.

Table 2.1: Sample sizes and match rates by place of birth

Country	1900 # in universe	Number matched	Match rate, total	1900 # Unique	Match rate, unique
A. 1900 source: IPUMS					
Austria	4,722	397	0.084	--	--
England	7,296	916	0.126		
France	11,615	728	0.063		
Germany	19,855	2,891	0.146		
Ireland	9,737	1,115	0.115		
Italy	6,649	1,076	0.162		
Norway	3,541	575	0.162		
Russia	5,641	771	0.136		
Sweden	6,164	633	0.102		
US natives	10,000	1,891	0.190	--	--
B. 1900 source: Ancestry.com					
Belgium	6,060	545	0.090	5,962	0.091
Denmark	34,594	1,980	0.058	17,425	0.114
Finland	23,843	828	0.035	22,197	0.037
Portugal	12,585	584	0.046	8,362	0.070
Scotland	53,091	4,349	0.082	15,529	0.280
Switzerland	22,276	3,311	0.149	20,588	0.161
Wales	17,767	1,342	0.076	9,876	0.135

Notes: The sample universe includes men between the ages of 18 and 35 in 1900. Immigrants must have arrived in the US between 1880 and 1900. We exclude all blacks and native born men living in the South. For large sending countries and the native born, we start with the 1900 IPUMS sample (Panel A). For smaller sending countries, we begin with the complete population in 1900. The text describes our matching procedure. The number of matched cases refers to men who match to both the 1910 and 1920 Censuses. We report the number of unique cases by first name, last name, age and country-of-birth and the match rate for this group in columns 4 and 5 for the smaller countries, for which we have a complete population.

Table 2.2: Common occupations for natives and foreign-born in matched samples, 1920

	Natives				Foreign-born			
	Occupation	Freq.	%	Mean income	Occupation	Freq.	%	Mean income
1.	Farmer	352	24.82	$12,702	Farmer	3,301	18.09	$12,702
2.	Manager	129	9.10	36,991	Manager	1,999	10.95	36,991
3.	Laborer	117	8.25	18,070	Laborer	1,791	9.81	18,070
4.	Salesman	75	5.28	22,253	Operative	1,102	6.04	20,931
5.	Operative	71	5.00	20,931	Foreman	603	3.30	34,194
6.	Clerical	45	3.17	22,579	Mine operative	596	3.27	21,854
7.	Carpenter	45	3.17	21,728	Machinist	578	3.17	28,961
8.	Machinist	45	3.17	28,961	Carpenter	529	2.90	21,728
9.	Farm laborer	39	2.75	12,674	Salesman	495	2.71	22,253
10.	Foreman	27	1.90	34,194	Clerical	326	1.79	22,579
Total (top 10)		945	66.61	$20,460		11,320	62.03	$22,542
Outside top 10		481	33.39	$29,015		6,929	37.97	$25,258

Notes: See notes to Table 1 for sample restrictions. Occupations based on 'OCC1950' IPUMS variable. Mean incomes reported in 2010 dollars.

Table 2.3: Comparing matched panel sample with population, 1920
Occupation-based earnings in $2010 dollars

	Mean, Panel sample	Difference, Panel sample - population	
		Levels	Logs
Native born	$23,200	52.92	0.010
		(301.546)	(0.013)
Foreign born	$23,471	368.75	0.024
		(127.42)	(0.006)

Notes: Occupation-based earnings based on 'OCCSCORE' IPUMS variable, converted into 2010 dollars. Regressions in columns 2 and 3 pool the 1920 IPUMS cross-section with our matched sample and regress occupation-based earnings on a dummy variable for being in the matched sample. Standard errors are in parentheses.

Table 2.4: OLS estimates, Age-earnings profile for natives and foreign-born, 1900-1920, Occupation-based earnings in $2010 dollars

	(1) Cross-section	(2) Pooled cross-section and panel	
RHS variable		(a) Cross-section coefficients	(b) Panel coefficients
0-5 yrs in US	-1184.27	-314.14	324.16
	(223.14)	(185.97)	(225.66)
6-10 yrs US	-673.57	53.51	448.96
	(200.01)	(170.30)	(200.74)
11-20 yrs US	-378.28	126.81	295.88
	(171.53)	(131.57)	(143.65)
21-30 yrs US	-273.55	126.06	99.42
	(179.52)	(136.40)	(143.22)
30 yrs in US	-18.00	103.85	149.71
	(217.551)	(176.42)	(177.12)
Arrive 1891+	---	-742.61	-230.78
		(107.11)	(154.45)
Native born	---	---	-118.68
			(167.99)
N	205,458	262,248	

Notes: See Table 1 notes for sample restrictions. Columns report coefficients from estimation of equation 1. Column (1) pools three cross-sections (1900-20); the regression in column (2) adds the matched panel sample. The coefficients in sub-column (a) are interactions between the right-hand side variables listed and a dummy for being in the cross-section, while sub-column (b) reports interactions between the right-hand side variables and a dummy for being in the panel. The omitted category is native-born men in the cross-section. Coefficients on age, Census year dummies, and country-of-origin fixed effects not shown.

Table 2.5: Occupation-based earnings distribution, 1900-20 in 2010 dollars

	Cross-section		Panel	
	Immigrants	Natives	Immigrants	Natives
10th	$9,900	$8,100	$11,700	$8,100
25th	$18,000	$12,600	$18,000	$12,600
50th	$20,700	$20,700	$20,700	$20,700
75th	$22,500	$25,200	$22,500	$25,200
90th	$28,800	$34,200	$28,800	$34,200
99th	$37,800	$55,800	$37,800	$56,700

Notes: Occupation-based earnings for men aged 16-38. Immigrants restricted to men who have lived in the US for 10 years or less.

Table 2.6: Predicting cross-country differences in immigrants' initial occupation-based earnings relative to natives

Dependent variable = Initial difference in occupation-based earnings (immigrants versus natives)

Characteristic of sending country (RHS variable)	Mean/standard deviation of RHS variable	Univariate regression*	Multivariate regression: Add economic variable**	Multivariate regression: Add cultural variables***
Share in agriculture	0.466 (0.172)	-6526.86 (3113.67)	-7476.85 (3619.31)	3546.71 (4309.10)
Real wage	57.726 (25.636)	43.93 (24.67)	23.70 (23.77)	12.79 (17.28)
Natural increase	10.406 (3.635)	-7.62 (169.49)	-85.76 (156.14)	-206.82 (105.02)
Infant mortality rate	174.933 (54.934)	10.02 (10.15)	16.09 (9.48)	7.35 (8.19)
Linguistic distance	0.526 (0.344)	-3419.61 (1534.52)	-2229.88 (2540.56)	1090.03 (1860.67)
Cultural distance	1.053 (0.588)	-2999.37 (677.38)	-2610.20 (961.15)	-1848.62 (920.47)
Religious distance	0.148 (0.045)	-39,433.23 (8484.51)	-39,140.04 (11,222.16)	-22,244.94 (12,943.87)

Notes: Initial earnings difference between permanent immigrants and US natives is measured using the coefficient on "0-5 years in US" (relative to natives) in panel sample. N = 16 except for the following RHS variables: infant death (missing for Portugal); cultural distance (missing for Russia) and real wage (missing for Finland, Russia and Switzerland). Economic characteristics are taken from Mitchell (2007) and real wage from Williamson (1995). We accessed the cultural distance measures from Sin (2011). These measures were originally collected by Alesina, et al. (2003); Fearon (2003) and Hofstede (1980). Real wage is indexed to US = 100. Natural increase is crude birth rate minus crude death rate. Infant mortality rate measured per 1000 live births. Linguistic distance theoretically varies between 0 (close) and 1 (far). Religious distance measures the chance that two randomly selected residents, one from the country of origin and one from the US, are of different religious backgrounds. Cultural distance varies between 0.3 (close) and 2.1 (far) in this sample. Standard errors are in parentheses.

* Column 3 regresses the immigrant earnings penalty on each of the RHS variables listed in the first column in turn.

** Column 4 adds "share in agriculture" to each regression (except in the "share in agriculture" row, in which case the real wage is added).

*** Column 5 adds the cultural distance and religious similarity measures to each regression. In "cultural distance" row, we only add religious similarity and in "religious similarity" row we only add cultural distance.

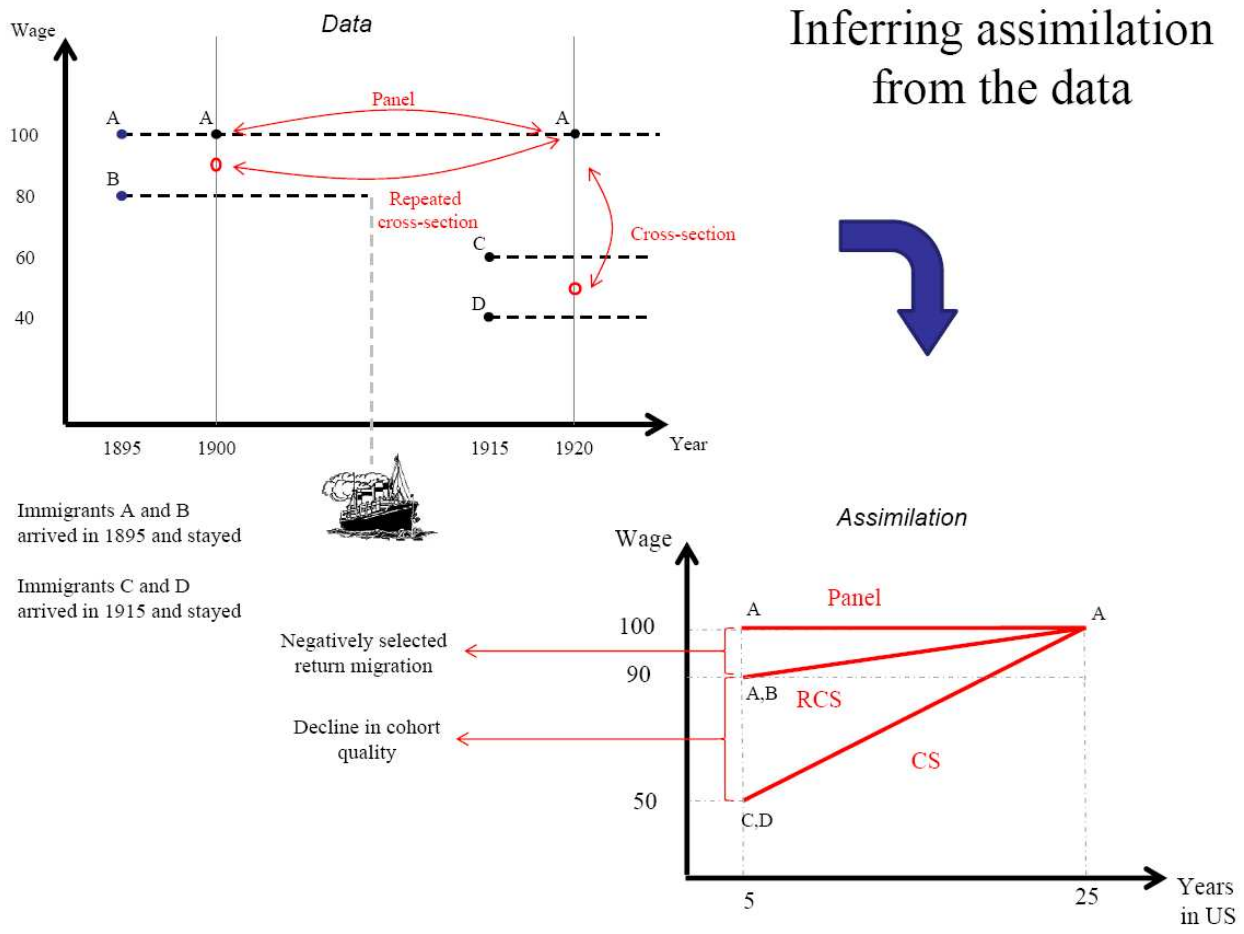
Table 2.7: Robustness for age-earnings profile in panel sample, 1900-1920

	A. Without country FE		*B. 4 arrival cohorts*		*C. Country x cohort FE*	
	RCS	***Panel***	***RCS***	***Panel***	***RCS***	***Panel***
0-5 years in US	-132.53	632.18	44.52	588.60	-6.806	640.95
	(169.89)	(214.37)	(219.02)	(262.28)	(247.55)	(277.98)
6-10 yrs in US	362.45	736.20	86.587	399.58	332.24	702.95
	(145.07)	(180.92)	(196.75)	(226.47)	(228.31)	(250.96)
11-20 yrs in US	474.73	569.06	226.58	347.85	436.95	586.86
	(92.613)	(110.09)	(155.93)	(167.73)	(203.68)	(211.28)
21-30 yrs in US	436.39	359.30	196.66	136.77	417.65	387.28
	(100.04)	(108.00)	(157.90)	(165.04)	(206.22)	(211.53)
30+ yrs in US	300.30	398.08	184.00	198.59	414.38	457.35
	(148.23)	(147.04)	(187.25)	(189.01)	(236.21)	(239.64)
	D. ln(occupation score)		***E. Raise farmer income***		***F. 1900 income***	
	RCS	***Panel***	***RCS***	***Panel***	***RCS***	***Panel***
0-5 years in US	0.051	0.087	-670.49	26.20	-3241.10	-2717.00
	(0.010)	(0.011)	(182.15)	(221.63)	(148.88)	(186.45)
6-10 yrs in US	0.066	0.083	-340.30	120.39	-2694.03	-2033.67
	(0.008)	(0.009)	(166.45)	(196.15)	(142.32)	(164.90)
11-20 yrs in US	0.063	0.069	-248.11	-36.01	-2257.03	-1972.49
	(0.006)	(0.007)	(128.91)	(140.41)	(113.37)	(119.66)
21-30 yrs in US	0.053	0.060	-227.43	-206.59	-2055.00	-1945.88
	(0.006)	(0.006)	(133.38)	(139.12)	(115.48)	(119.79)
30+ yrs in US	0.044	0.057	-221.87	-129.70	-1828.41	-1755.42
	(0.008)	(0.008)	(170.33)	(169.90)	(137.24)	(138.07)
	G. Drop child migrants		***H. State FE***		***I. State * urban FE***	
	RCS	***Panel***	***RCS***	***Panel***	***RCS***	***Panel***
0-5 years in US	-419.28	212.13	-1654.14	-1138.32	-2218.65	-1430.21
	(191.28)	(232.69)	(198.71)	(301.55)	(197.41)	(301.41)
6-10 yrs in US	-95.30	259.10	-1285.28	-759.39	-2009.73	-1132.61
	(177.25)	(210.17)	(190.91)	(257.23)	(189.70)	(257.31)
11-20 yrs in US	-50.80	67.75	-1193.63	-960.18	-1860.04	-1032.53
	(144.09)	(160.15)	(157.29)	(201.16)	(156.93)	(199.08)
21-30 yrs in US	120.75	-81.71	-1073.68	-1139.49	-1659.34	-1111.17

	(148.59)	(161.72)	(163.78)	(209.51)	(163.46)	(210.92)
30+ yrs in US	118.75	62.60	-1008.40	-577.14	-1539.02	-478.06
	(201.37)	(209.79)	(196.08)	(229.82)	(193.29)	(230.63)

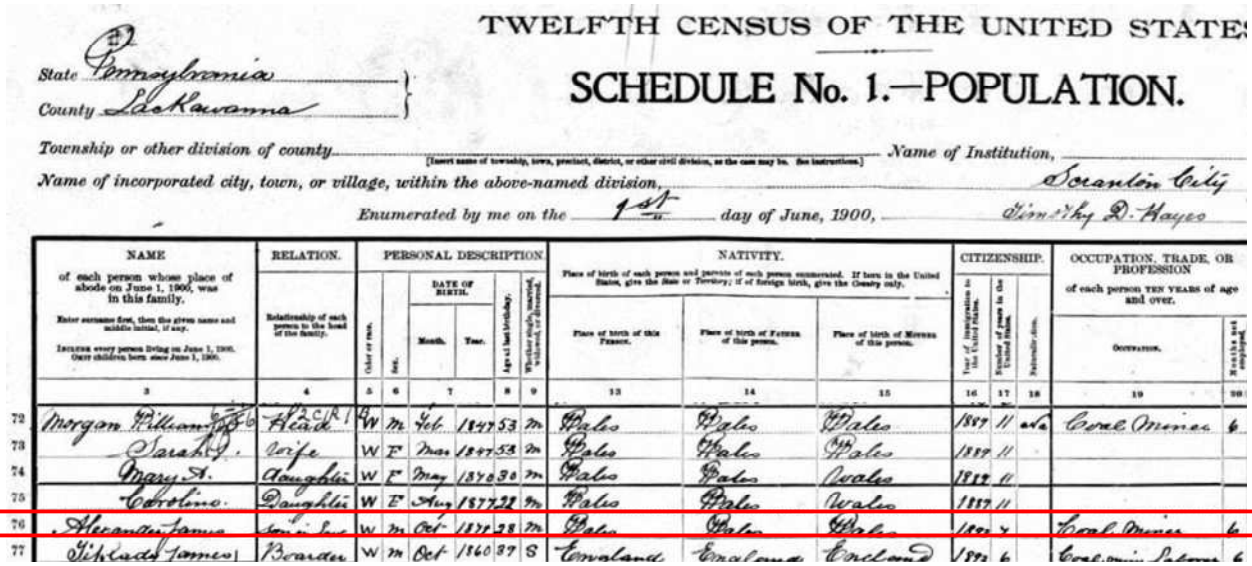
Notes to Table 7: See notes to Table 4 for sample restrictions. All regressions follow the specification in Table 4 with the exception of the modification listed in panel titles. In Panel B, the four arrival cohorts are 1880-85; 1886-1890; 1891-95; and 1896-1900. Panel C interacts the single cohort indicator (1891-1900) with country fixed effects. In Panel E, we raise farmers' income by 20 percent. Panel F replaces the 1950 occupation score measure with occupation-based income from the 1900 Cost of Living Survey. Panel G drops immigrants who arrived in the US before age 10 or after age 40. Standard errors are in parentheses. Sample size for Panels A-E is 262,248. Panel F has 264,338 observations; Panel G has 246,365; Panel H has 228,793 and Panel I has 227,930.

Figure 2.1: Inferences about immigrant assimilation in cross-sectional and panel data



Notes: The graph in the northwest corner depicts earnings for four hypothetical migrants. For illustrative purposes, we assume that natives earn 100 in every year. Migrants A and B arrived in 1895 and earn 100 and 80 respectively. Migrant B returns to his home country in 1909. Migrants C and D arrived in 1915 and earn 60 and 40 respectively. The graph in the southeast corner illustrates inferred assimilation profiles from a series of hypothetical datasets. The line labeled 'CS' refers to the 1920 cross section, wherein a researcher would compare migrants C and D to migrant A. The line labeled 'RCS' refers to two repeated cross sections (1900 and 1920), with which a researcher could compare migrants A and B in 1900 to migrant A in 1920. The line labeled 'Panel' refers to a panel data that follows migrant A over time.

Figure 2.2: Sample Census manuscripts illustrating matching procedure, 1900-1910-1920



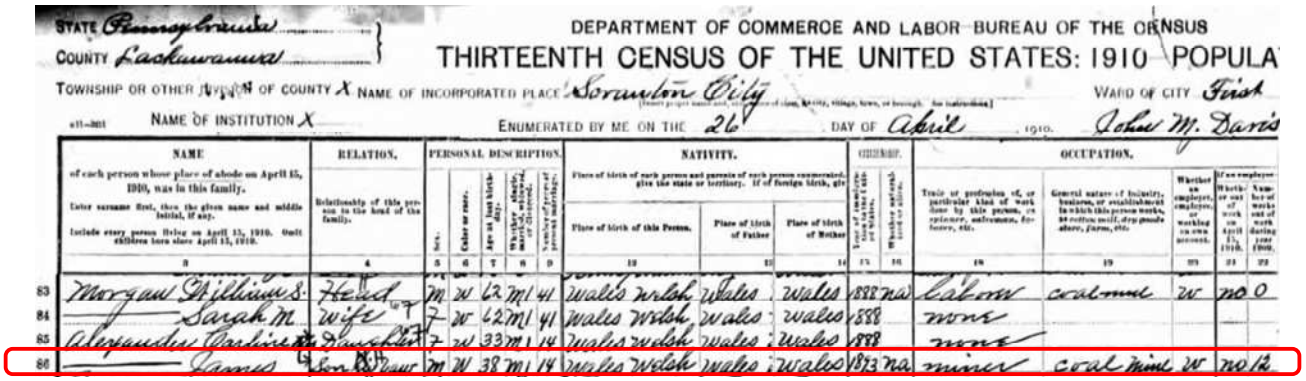
TWELFTH CENSUS OF THE UNITED STATES

SCHEDULE No. 1.—POPULATION.

State Pennsylvania
County Lackawanna

Township or other division of county ____ Name of Institution, ____
Name of incorporated city, town, or village, within the above-named division, Scranton City
Enumerated by me on the 1st day of June, 1900, Wm. Ray D. Hayes

	NAME of each person whose place of abode on June 1, 1900, was in this family.	RELATION.	PERSONAL DESCRIPTION. Color or race.	Sex.	DATE OF BIRTH. Month.	Year.	Age at last birthday.	Whether single, married, widowed, or divorced.	NATIVITY. Place of birth of this Person.	Place of birth of Father of this person.	Place of birth of Mother of this person.	CITIZENSHIP. Year of immigration to the United States.	Number of years in the United States.	Naturalized.	OCCUPATION, TRADE, OR PROFESSION of each person TEN YEARS of age and over. Occupation.	Months not employed.
72	Morgan Williams	Head	W	M	Feb	1847	53	M	Wales	Wales	Wales	1869	11	alien	Coal Miner	6
73	Sarah J.	wife	W	F	Mar	1847	53	M	Wales	Wales	Wales	1869	11			
74	Mary A.	daughter	W	F	May	1870	30	M	Wales	Wales	Wales	1888	11			
75	Caroline	Daughter	W	F	Aug	1877	22	M	Wales	Wales	Wales	1887	11			
76	Alexander James	Son in law	W	M	Oct	1879	20	M	Wales	Wales	Wales	1891	4		Coal Miner	6
77	Pritchard James	Boarder	W	M	Oct	1860	39	S	England	England	England	1891	6		Coal mine Laborer	6

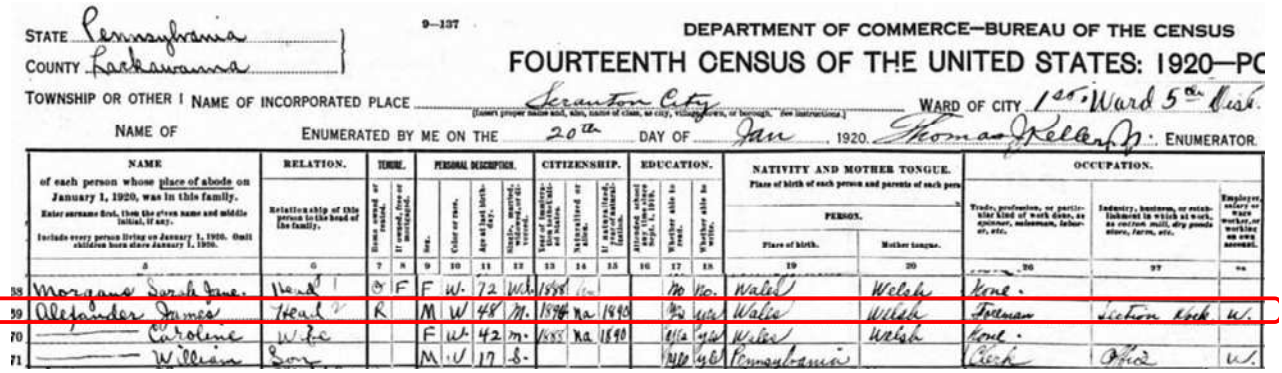


STATE Pennsylvania
COUNTY Lackawanna

DEPARTMENT OF COMMERCE AND LABOR—BUREAU OF THE CENSUS
THIRTEENTH CENSUS OF THE UNITED STATES: 1910 POPULA

TOWNSHIP OR OTHER DIVISION OF COUNTY X NAME OF INCORPORATED PLACE Scranton City WARD OF CITY First
NAME OF INSTITUTION X ENUMERATED BY ME ON THE 26 DAY OF April 1910. John M. Davis

	NAME of each person whose place of abode on April 15, 1910, was in this family.	RELATION.	PERSONAL DESCRIPTION. Sex.	Color or race.	Age at last birthday.	Whether married, widowed, or divorced.	Number of years of present marriage.	NATIVITY. Place of birth of this Person.	Place of birth of Father	Place of birth of Mother	CITIZENSHIP. Year of immigration to the United States.	Whether naturalized or alien.	OCCUPATION. Trade or profession of, or particular kind of work done by this person.	General nature of industry, business, or establishment in which this person works.	Whether an employer, employee, or working on own account.	If an employee, Number of weeks out of work during year 1909.
83	Morgan Williams S.	Head	M	W	63	M	41	Wales	Welsh Wales	Wales	1888	Na	Laborer	coal mine	W	no 0
84	Sarah M	wife	F	W	62	M	41	Wales	Welsh Wales	Wales	1888		none			
85	Alexander Caroline	Daughter	F	W	33	M	14	Wales	Welsh Wales	Wales	1888		none			
86	James	Son in law	M	W	38	M	14	Wales	Welsh Wales	Wales	1873	Na	miner	coal mine	W	no 12



STATE Pennsylvania
COUNTY Lackawanna

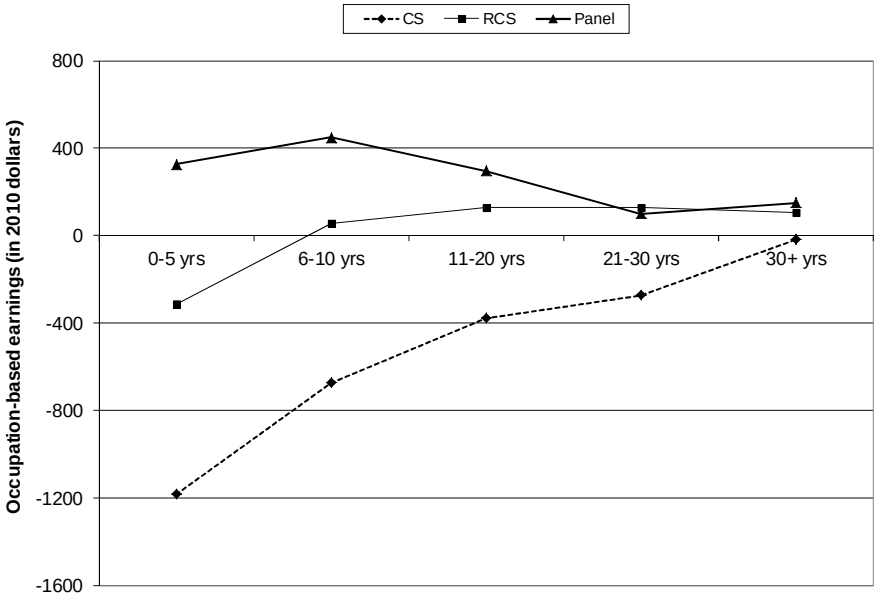
9—137

DEPARTMENT OF COMMERCE—BUREAU OF THE CENSUS
FOURTEENTH CENSUS OF THE UNITED STATES: 1920—PO

TOWNSHIP OR OTHER DIVISION OF COUNTY NAME OF INCORPORATED PLACE Scranton City WARD OF CITY 1st Ward 5th Dist.
NAME OF INSTITUTION ENUMERATED BY ME ON THE 20th DAY OF Jan 1920. Thomas J Kelly Jr. ENUMERATOR.

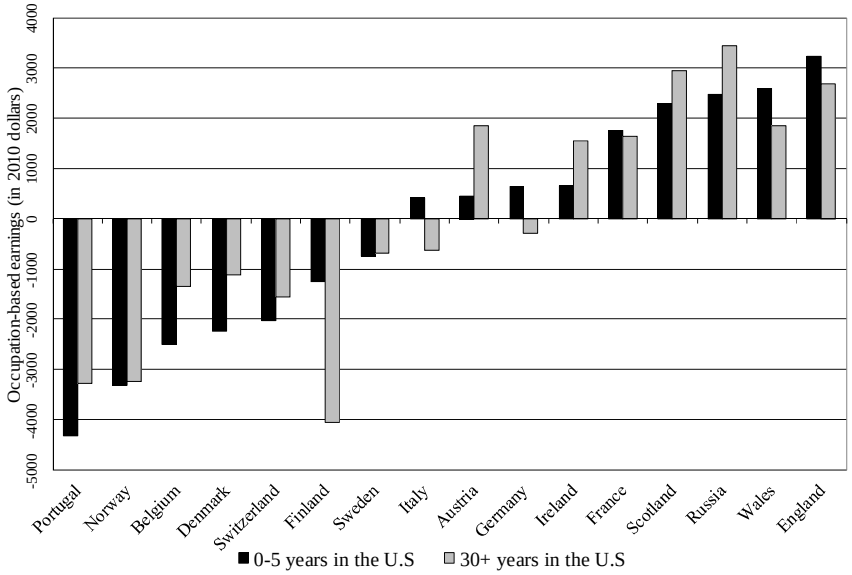
	NAME of each person whose place of abode on January 1, 1920, was in this family.	RELATION.	HOME. Home owned or rented.	If owned, free or mortgaged.	PERSONAL DESCRIPTION. Sex.	Color or race.	Age at last birth.	Single, married, widowed, or divorced.	CITIZENSHIP. Year of immigration to the U.S.	Naturalized or alien.	If naturalized, year of naturalization.	EDUCATION. Attended school since Sept. 1, 1919.	Whether able to read.	Whether able to write.	NATIVITY AND MOTHER TONGUE. Place of birth of each person and parents of each person. PERSON. Place of birth.	Mother tongue.	OCCUPATION. Trade, profession, or particular kind of work done, as spinner, salesman, laborer, etc.	Industry, business, or establishment in which at work, as cotton mill, dry goods store, farm, etc.	Employer, salary or wage worker, or worker on own account.
68	Morgan Sarah Jane	Head	O	F	F	W	72	Wid	1888				No	No	Wales	Welsh	None		
69	Alexander James	Head	R		M	W	48	M	1896	Na	1898		Yes	Yes	Wales	Welsh	Foreman	Section Work	W
70	Caroline	Wife			F	W	42	M	1888	Na	1898		Yes	Yes	Wales	Welsh	None		
71	William	Son			M	W	17	S					Yes	Yes	Pennsylvania		Clerk	Office	W

Figure 2.3: Convergence in occupation score between immigrants and native-born workers by time spent in the US, cross-sectional and panel data, 1900-1920



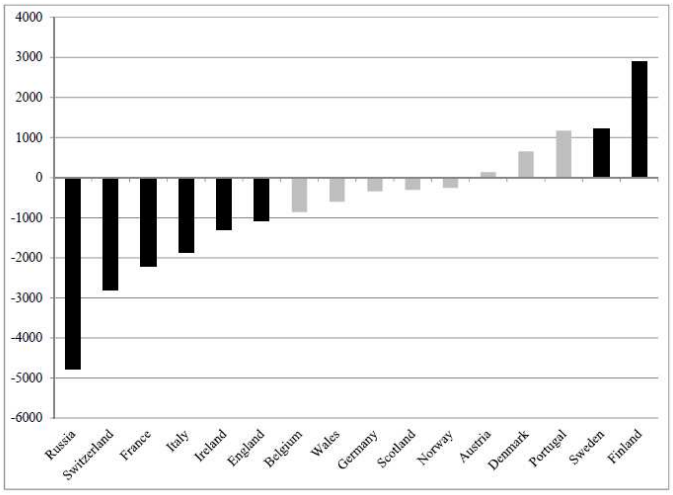
Notes: Plot of coefficients for "years spent in the US" indicators in equation 1. See Table 4 for coefficients and standard errors.

Figure 2.4: Initial earnings gap between native- and foreign-born in panel sample. Natives versus immigrants upon first arrival (0-5 years in US) and after time in the US (30+ years in US), by country of origin



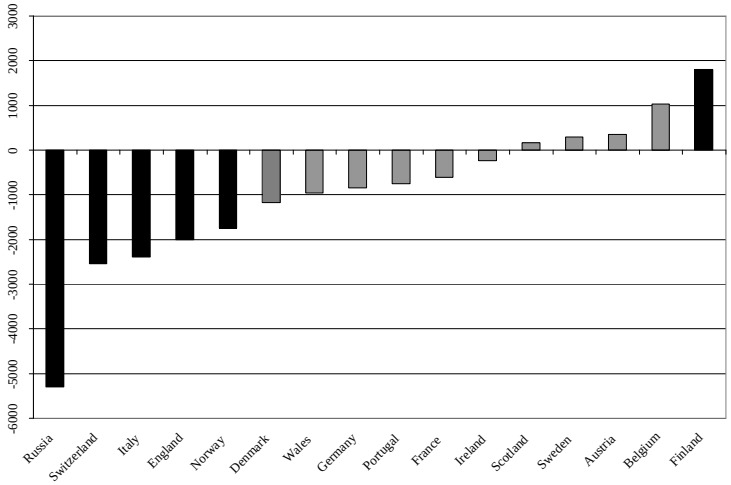
Notes: Graph reports coefficients on interaction between country-of-origin fixed effect and dummy variables for being in the US for 0-5 years or for 30+ years from regression of equation 1 in the panel sample. All coefficients for the 0-5 year interaction are significant except those for Austria, Germany, Ireland, Italy and Sweden. No differences between the 0-5 year and 30+ year coefficients are significant except that for Finland.

Figure 2.5: Changing quality of arrival cohorts, difference between immigrant penalty for early and late arrivals in panel sample, by country of origin



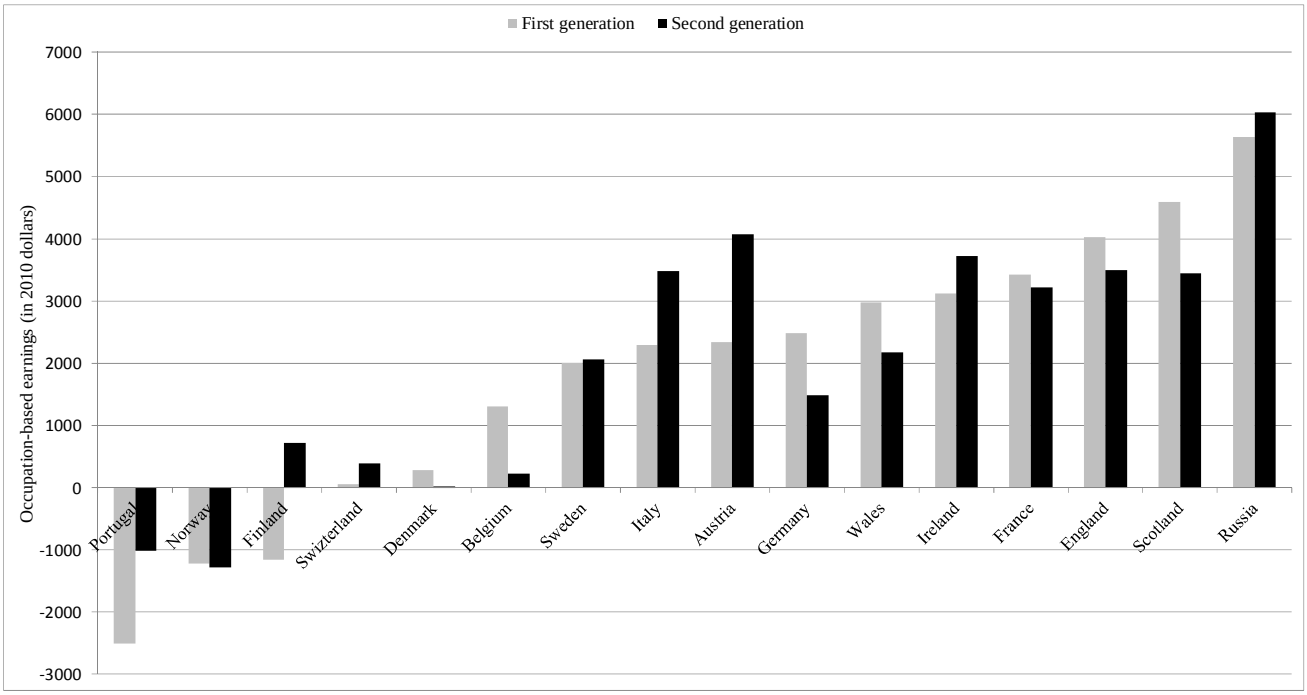
Notes: Estimates based on a version of equation 1 with four dummy variables for arrival cohorts in the panel sample (see Table 5, Panel B). The graph reports the difference between two coefficients: one interacts a country-of-origin fixed effect with the dummy variable for arriving in the US between 1880 and 1884 and the other interacts the country fixed effect with a dummy variable for arriving in the US between 1895 and 1900. Differences that are significantly different from zero are in black.

Figure 2.6: Implied selection of return migrants, Difference between estimated convergence in panel and repeated cross-section data, by country of origin



Notes: Figure reports the difference between immigrants' occupational upgrading relative to natives (defined as the difference between occupation-based earnings after 30+ years and after 0-5 years) in the cross-section versus the panel sample, by sending country. Results are from regression of equation 1 pooling the panel and cross-section samples. Coefficients that are significantly different from zero are in black.

Figure 2.7: Convergence in occupation-based earnings across immigrant generations. First-generation and second-generation migrants versus natives, By country of origin



Notes: We estimate the regression in equation 2 separately for each group and for each country – immigrants (1st generation), US natives in the same Censuses and ages as the immigrants, sons of immigrants (2nd generation), US natives in the same Censuses and ages as the 2nd generation sample. The bars for the first generation represent the difference in the predicted occupation-based earnings of an immigrant who came in 1890 and is 35 years old in 1910, relative to a 35-year old native. The bars for the second generation represent the difference in the predicted occupation-based earnings of a man born in the US to immigrant parents relative to a man born in the US to native parents, both of whom were 35 years old in 1930. First generation immigrants are taken from the panel sample. Natives and second generation immigrants come from IPUMS data in the respective Census year.

References

R. Abramitzky, L. Boustan and K. Eriksson, Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review*, forthcoming.

R. Alba and V. Nee, *Remaking the American Mainstream: Assimilation and Contemporary Immigration*. (Cambridge, MA: Harvard University Press, 2003).

A. Alesina, A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg, Fractionalization. *Journal of Economic Growth* **8** (2003), pp. 155-194.

J. Atack and F. Bateman, 'Matchmaker, Matchmaker, Make Me a Match': A General Personal Computer-Based Matching Program for Historical Research *Historical Methods* **25** (1992), pp. 53-65.

O. Bandiera, I. Rasul and M. Viarengo, The Making of Modern America: Migratory Flows in the Age of Mass Migration. Manuscript, 2010.

K. Benton-Cohen, The Rude Birth of Immigration Reform. *Wilson Quarterly* 2010.

F. Blau, Immigration and Labor Earnings in Early Twentieth Century America. *Research in Population Economics* **2** (1980), pp. 21–41.

H. Bleakley and A. Chin, Age at Arrival, English Proficiency, and Social Assimilation Among U.S. Immigrants. *American Economic Journal: Applied Economics* **2** (2010), pp. 165-192.

G. Borjas, Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants. *Journal of Labor Economics* **3** (1985), pp. 463–489.

G. Borjas, Immigrant and Emigrant Earnings: A Longitudinal Study. *Economic Inquiry* **27** (1989), pp. 21–37.

G. Borjas, Long-Run Convergence of Ethnic Skill Differentials: The Children and Grandchildren of the Great Migration. *Industrial and Labor Relations Review* **47** (1994), pp. 553-573.

G. Borjas, Does Immigration Grease the Wheels of the Labor Market? *Brookings Papers on Economic Activity* (2001), pp. 69-133.

L. Carter, A Hard Day's Night: Evening Schools and Child Labor in the United States, 1870-1910. Manuscript, 2008.

B. Chiswick, The Effect of Americanization on the Earnings of Foreign-born Men. *Journal of Political Economy* **86** (1978), pp. 897–921.

A. Constant and D. S. Massey, Self-selection, Earnings, and Out-Migration: A Longitudinal Study of Immigrants to Germany. *Journal of Population Economics* **16** (2003), pp. 631-653.

P. Douglas, Is the New Immigration More Unskilled than the Old? *Publications of the American Statistical Association* **16** (1919), pp. 393-403.

H. Duleep and D. Dowhan, Insights from Longitudinal Data on the Earnings Growth of U.S. Foreign-Born Men. *Demography* **39** (2002), pp. 485–506.

C. Dustmann, Earnings Adjustment of Temporary Migrants. *Journal of Population Economics* **6** (1993), pp. 153-168.

Z. Eckstein and Y. Weiss, On the Wage Growth of Immigrants: Israel, 1990-2000. *Journal of the European Economic Association* **2** (2004), pp. 665-695.

P-A. Edin, R. LaLonde and O. Aslund, Emigration of Immigrants and Measures of Immigrant Assimilation: Evidence from Sweden. *Swedish Economic Policy Review* **7** (2000), pp. 163-204.

B. Eichengreen and H. Gemery, The Earnings of Skilled and Unskilled Immigrants at the End of the Nineteenth Century. *Journal of Economic History* **46** (1986), pp. 441–454.

J. Fearon, Ethnic and Cultural Diversity by Country. *Journal of Economic Growth* **8** (2003), pp. 195-222.

J. Ferrie, A New Sample of Males Linked from the Public Use Micro Sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census Manuscript Schedules. *Historical Methods* **29** (1996), pp. 141-156.

----------, The Entry into the U.S. Labor Market of Antebellum European Immigrants, 1840–1860. *Explorations in Economic History* **34** (1997), pp. 295–330.

----------, *Yankees Now: Immigrants in the Antebellum United States, 1840-1860*. (New York: Oxford University Press, 1999).

J. Ferrie and J. Long, Long-Distance Migration and Intergenerational Mobility: British Emigration to the U.S. and Canada, 1851-1881. Manuscript (2011).

J. W. Frank and J. F. Mustard, The Determinants of Health from a Historical Perspective. *Daedalus* **123** (1994), pp. 1-19.

R. Fryer and S. Levitt, The Causes and Consequences of Distinctively Black Names. *Quarterly Journal of Economics* **119** (2004), pp. 767-805.

R. Friedberg, The Labor Market Assimilation of Immigrants in the United States: The Role of Age at Arrival. Manuscript (1993), Brown Univ.

O. Galor and O. Stark, Migrants' Savings, The Probability of Return Migration and Migrants' Performance. *International Economic Review* **31** (1990), pp. 463-467.

C. Goldin and R. A. Margo, The Great Compression: The Wage Structure in the United States at Mid-Century. *Quarterly Journal of Economics* **107** (1992), pp. 1-34.

C. Goldin, The Political Economy of Immigration Restriction in the United States, 1890 to 1921. In *The Regulated Economy: A Historical Approach to Political Economy*, eds. C. Goldin and G. D. Libecap. (Chicago: University of Chicago Press, 1994), pp. 223–57.

J. D. Gould, European Inter-Continental Emigration. The Road Home: Return Migration from the U.S.A. *Journal of European Economic History* **9** (1980), pp. 41-112.

C. Hanes, Immigrants' Relative Rate of Wage Growth in the Late 19th Century. *Explorations in Economic History* **33** (1996), pp. 35–64.

J. Hannon, Ethnic Discrimination in a 19th-Century Mining District: Michigan Copper Mines, 1988. *Explorations in Economic History* **19** (1982), pp. 28–50.

T. Hatton, The Immigrant Assimilation Puzzle in Late Nineteenth-Century America. *Journal of Economic History* **57** (1997), pp. 34–62.

T. Hatton and J. Williamson, *The Age of Mass Migration: Causes and Economic Impact*. (Oxford: Oxford University Press, 1998).

R. Higgs, Race, Skills, and Earnings: American Immigrants in 1909. *Journal of Economic History* **32** (1971), pp. 420–428.

J. Higham, *Strangers in the Land: Patterns of American Nativism, 1860–1925*. 2nd ed. (New Brunswick, NJ: Rutgers University Press, 1988, Orig. pub. 1955.)

G. Hofstede, *Cultures Consequences: International Differences in Work-related Values*. (Beverly Hills, CA: Sage, 1980).

W-Y. Hu, Immigrant Earnings Assimilation: Estimates from Longitudinal Data. *A.E.R. Papers and Proceedings* **90** (2000), pp. 368–72.

R. A. Hummer and J. T. Lariscy, Educational Attainment and Adult Mortality. In *International Handbook of Adult Mortality. Vol. 2*, eds. R. Rogers and E. M. Crimmins. (New York: Springer), pp. 241-261.

M. F. Jacobson, *Whiteness of a Different Color: European Immigrants and the Alchemy of Race*. (Cambridge, MA: Harvard University Press, 1999).

G. Jasso and M. Rosenzweig, How Well Do US Immigrants Do? Vintage Effects, Emigration Selectivity, and Occupational Mobility. *Research in Population Economics* **6** (1988), pp. 229–253.

S. Kim, Economic Assimilation of Foreign-Born Workers in the United States: An Overlapping Rotating Panel Analysis, Manuscript (2011). http://faculty.washington.edu/seikkim/seikkim_immorpm.pdf

A. Lleras-Muney, Were Compulsory Attendance and Child Labor Laws Effective? An Analysis from 1915 to 1939. *Journal of Law and Economics,* **45.2** (2002), pp. 401-435.

A. Lleras-Muney and A. Shertzer, The Americanization Movement, 1910-1930. Manuscript 2011.

S. Lieberson, *A Piece of the Pie: Blacks and White Immigrants Since 1880*. (Berkeley, CA: University of California Press, 1980).

D. Lubotsky, Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings. *Journal of Political Economy* **115** (2007), pp. 820-867.

W. Merriam, *Twelfth Census of the United States, Taken in the Year 1900: Statistical Atlas*, United States Census Office, Washington, 1903.

P. McGouldrick and M. Tannen, Did American Manufacturers Discriminate Against Immigrants Before 1914?. *Journal of Economic History* **37** (1977), pp. 723–746.

C. Minns, 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** (2000), pp. 326-350.

B. Mitchell, *International Historical Statistics 1750-2005: Europe, 6th edition.* (New York: Palgrave Macmillan, 2007).

C. M. Moehling, State Child Labor Laws and the Decline of Child Labor. *Explorations in Economic History* **36** (1999), pp. 72-106.

T. Muller, *Immigrants and the American City*. (New York: New York University Press, 1993).

J. Perlmann, Selective Migration as a Basis for Upward Mobility? The Occupation of the Jewish Immigrants to the United States, ca. 1900, working paper 1999.

M. Piore, *Birds of Passage: Migrant Labor and Industrial Societies*. (Cambridge University Press, 1980).

S. Ruggles, J. T. Alexander, K. Genadek, R. Goeken, M. B. Schroeder, and M. Sobek, *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

I. Sin, *Insights from Book Translations on the International Diffusion of Knowledge*, Stanford Dissertation, May 2011.

J. Williamson, The Evolution of Global Labor Markets since 1830: Background Evidence and Hypotheses. *Explorations in Economic History* **32** (1995), pp. 141-196.

M. Wyman. Round Trip to America: The Immigrants Return to Europe, 1880-1930. (Ithaca, NY: Cornell University Press, 1996).

R. Zakharenko, Return Migration: An Empirical Investigation. MPRA Paper No. 13755 (2009). http://mpra.ub.uni~muenchen.de/13755/

Chapter 3: Does the language of instruction affect labor market outcomes? Evidence from South Africa

3.1 Introduction

The language of instruction in primary school and beyond is a subject of debate, both in the United States in the context of Hispanic students and in post-apartheid South Africa where schools must choose between the mother-tongue and English or Afrikaans for Black students. The language of instruction could directly influence language skill and literacy, but could also directly or indirectly affect other important outcomes, such as wages and educational attainment. To the extent that the language of instruction affects school quality it could influence health outcomes, fertility, and social factors such as crime.[65]

Instruction in a student's native language could have both positive and negative effects. Educating students in their mother tongue for the first years of school will raise the chances of building non-language cognitive skills such as literacy and numeracy. In addition, school might be more accessible for students. Students could be more likely to enter school because they can understand the language. If they were going to drop out when the language of instruction changes to English, then increasing years of mother tongue instruction will prevent dropout.

There could also be positive school quality effects. For example, in South Africa, most teachers of Black children are Black so by teaching in English they are teaching in their second language. Quality of instruction might go up if they teach in their first language. There is evidence that teachers who speak the same language as their students are also better able to impose discipline, increasing the probability that students learn well (Taylor and Vinjevold,

[65] In this paper I look at literacy, educational attainment, wages, and fertility. The South African census does not ask about other health outcomes.

1999). Finally, parents who speak only the mother tongue are more likely to be able to help with school work if their student's work is in a language they know.

On the other hand, teaching in English could increase English language skills, the language of most market transactions. In contemporary South Africa, using one language of instruction could enhance diversity in classrooms rather than separating students across language and therefore race. English skills are important for obtaining higher paid jobs, particularly in post-colonial countries where government jobs require the colonial language. There is a wealth of literature about returns to English speaking ability which is primarily based on the United States.[66]

Estimating the effect of the language of instruction is difficult because it is often an endogenous choice that parents can make on behalf of their children. In contemporary South Africa, parents can choose the school that their child attends and can also influence the language of instruction through voting for School Governing Bodies. Richer parents are more likely to send their children to schools with instruction in English.

This paper uses a difference-in-difference framework in the context of an exogenous change in the language of instruction to identify the effect of increasing a student's potential years of mother tongue instruction from four years to six years. In 1953, the new *apartheid* government in South Africa passed the Bantu Education Act which mandated eight years of mother tongue instruction for all Black students. Prior to the law, instruction was in the student's mother tongue for less than eight years. Before the policy change, mother tongue instruction lasted for four years in the Transvaal, Cape, and Free State provinces and for six years in the

[66] See, for example, Bleakley and Chin (2007), McManus, Gould and Welch (1983), Chiswick (1991), and Dustmann (1994).

Natal province. I use the 1980 South African census to examine labor market and other outcomes of men and women aged 28 to 48 who were affected by the reform.

My results indicate that increasing exposure to mother tongue instruction had small positive effects on male wages of 1-4%. I find that literacy and education levels increase in response to the policy change and that the wage increase is through these channels. I also examine results for women and find smaller effects on wages despite increases in education and literacy.

This paper is most closely related to Angrist and Lavy (1997) and Angrist, Chin, and Godoy (2008). Angrist and Lavy (1997) use a natural experiment in Morocco to infer the effect of French language skills on test scores and the returns to school. Using a change from French to Arabic instruction after grade six, their difference-in-difference specification uses cohorts affected and unaffected by the change as well as students who obtained more than grade 6 and those who did not. They find strong negative effects of "Arabization" on both test scores and returns to school.[67]

Angrist, Chin, and Godoy (2008) use a similar identification strategy to determine the effects of Spanish-only schooling on language skills in Puerto Rico. They use the change in 1949 from some years of instruction in English to all instruction in Spanish. Controlling for education-cohort trends, they find that the language of instruction does not matter for English skills. Their paper highlights the importance of accounting for selection of skill groups into higher grades as well as controlling for trends which can bias results.

My paper makes three contributions to the literature on language of instruction. First, I use cross-province variation and over-time variation which allows me to construct difference-in-

[67] An important difference between my paper and Angrist and Lavy is that I focus on instruction in mother tongue. In contrast, a majority of the population of Morocco is Berber and does not speak Arabic or French as a first language so this policy probably did not affect more than the urban minority

difference estimates without using educational attainment as a source of variation. Educational attainment is a potentially problematic source of variation because it is likely endogenous to the policy reform. If there was selection into higher grades induced by the policy change and these were less able students, then the estimates of the effect will be biased. I have cross-province variation which allows me to still construct difference in difference estimates without using educational attainment. Second, as argued above, the policy change in South Africa was exogenous to the population it affected. It was imposed by the white government and was not in response to popular demands. The policy changes in Morocco and Puerto Rico were the result of national debates about the best language of instruction. Finally, by using the 1980 census, I look at long run effects of this policy change.

The paper is organized as follows. Section 2 provides background on South Africa's population and history, and outlines the language reform in detail. Section 3 presents the difference-in-difference identification strategy. Section 4 explores the results and Section 5 concludes.

3.2 Background

South Africa's population consists of three major racial groups. The Black, or African, population made up approximately 80% of the population in 1980. Whites constituted 9% which can be broken into Afrikaans- (60%) and English-speaking (40%) groups. The Coloured population constituted an additional 9% of the population and was found predominantly in the Cape Province.[68] Before apartheid was introduced in 1948, the English speaking white population was substantially more educated than the Afrikaans speakers. Until 1994, the Black

[68] In South Africa, "Coloured" means mixed race. Most Coloured people are of Afrikaans and Black descent and speak Afrikaans as a first language. This race was not subject to all of the *apartheid* restrictions but did have a specific education department and set of schools.

and Coloured populations were not eligible to vote or influence policy; even though the formal system of *apartheid* was not introduced until 1948, a less formal system restricted the vote and government to whites only.

Mariotti (2009) argues that the differences in educational opportunities between Afrikaans and English speakers and the resulting income differences influenced the direction of policy in South Africa throughout the 20th century. The Afrikaans speaking population, much larger than the English population, dominated South African politics for most of the 20th century. The National Party, supported by the Afrikaans population, won the 1948 election and instituted *apartheid*, a system of formal separation of the races. The *apartheid* government's goal was to separate the lives of the races, with the Black population providing a labor source for unskilled and semi-skilled jobs. To this end, major policy changes included the restriction of movement of the Black population, the creation of Black "homelands", formal job reservations for whites, and a reform of the education system in the early 1950's. [69]

The *apartheid* government worried about competition for semi-skilled jobs between Blacks and white Afrikaans speakers, so formal job reservation was legalized through the prohibition of Black union membership. Job skill classifications were used to determine which races could work in these jobs. Mariotti (2009) shows that this system had partially broken down by 1980 due to increased skills of whites and increasing demand for semi-skilled labor. Blacks were more likely to be working in semi-skilled jobs in 1980 than a decade previously, but there were still substantial restrictions on the jobs they could obtain. To the extent that these

[69] Formal homelands were created in areas with land unsuitable for major agricultural production. The apartheid government envisioned that the Black population would live solely in these areas, which would eventually become nominally independent, and that Black workers would only serve as migrant labor. The homelands gradually obtained "independence", with the Transkei, Bophutatswana, and Venda homelands being independent by 1980, with an additional seven still considered a part of South Africa. Despite the nominal independence of the homelands, they were never recognized by the international community.

restrictions meant that skill was not rewarded in the labor market, any positive effects of the language reform on literacy would not necessarily have translated into higher wages.

Before *apartheid*, the education system was a mix of publicly-provided education and mission or other private schools; provinces enjoyed autonomy over education policy. In 1951, the new government established the Eiselen Commission whose task was to examine the Black education system and make suggestions for a national education policy. In addition to a revision of the curriculum, the commission recommended that all Black students be taught in their mother tongue for the first eight years of school.

In 1953, the *apartheid* government passed the Bantu Education Act which established a separate department of education for Black education as well as national control over all education policy. The unstated objective of the policy was to create a population skilled enough to meet the growing demand for industrial and mining labor but not literate enough to join the intellectual class. Black education was meant to be training to work unskilled and semi-skilled jobs.

The Act also implemented the Eiselen Commission's recommendation that all students be taught in their mother tongue for the first eight years of school (Taylor and Vinjevold, 1999). Students were then taught after year eight in a mix of English and Afrikaans. The predominant white language in the area was required as a subject from the second year of school and the other "white" language was introduced in the fourth year (Hartshorne, 1992: 197).

Language of instruction policy before the 1950's was largely a relic of the practices of mission schools, which were established in the 19th century. Natal schools taught in the student's mother tongue for six years and in the three other provinces—the Transvaal, Cape, and Free

State—instruction was in the student's mother tongue for only four years. All students were taught in either English or Afrikaans thereafter.

After the reform, the new language policy was implemented from the fifth year of school (grade five) to the eighth year of school (grade eight) gradually, beginning in 1955. In 1955, all fifth graders were taught in their mother tongue. In 1956, mother tongue instruction was extended to grade six. By 1958, all students through grade 8 were taught in their mother tongue. Outside of Natal the implementation of the policy meant that students in grade five or below in 1955 would have had eight years of mother tongue instruction but anyone who had passed grade 5 by 1955 would have had only four years. For example, a student who was in sixth grade or beyond in 1955 was already taught in English and so did not switch back to her mother tongue.

In Natal this meant that the policy did not take effect until 1957 when grade seven switched from instruction in English to mother tongue. Here the first cohort affected was those who were in fifth grade in 1955 and consequently in seventh grade in 1957. Therefore, anyone who had passed fifth grade by 1955 was not affected by the change in either the treatment or control provinces. Hartshorne (1992: 198) argues that the policy was strictly enforced.

3.3 Identification Strategy

My difference-in-difference strategy exploits the different exposure to mother tongue instruction across cohorts and provinces. My approach is similar to that of Angrist and Lavy (1997) and Angrist, Chin, and Godoy (2008) in that it uses a policy change as a natural experiment to estimate a difference-in-difference equation, but differs because it uses cross-province instead of cross-education group variation.

My difference-in-difference strategy is described in Table 2. I compare cohorts affected and unaffected by the policy change and also compare provinces which were affected in different ways. I define a variable TCF to include being born outside of Natal, or in the Transvaal, Cape, or Free State provinces. The "old" cohort includes individuals born before 1942 while the "young" cohort includes those born in or after 1942. The first difference is the difference across cohorts within each province. In TCF, the "old" cohort was taught in their mother tongue for four years while the "young" cohort had eight years of mother tongue instruction. Similarly, in Natal, the "old" cohort received six years of mother tongue instruction, while the "young" cohort experienced eight years. Therefore, the policy change induced an increase of four years of mother tongue instruction outside of Natal, in TCF, and only two years in Natal.

The second difference is across provinces. Subtracting the change of two years from grade six to eight in Natal from the change of four years from grade four to eight in the Transvaal, my difference-in-difference estimator measures the effect of increasing mother tongue instruction by two additional years.

To interpret my difference-in-difference estimates as causal, I must make certain assumptions. First, I assume that children started school at age eight and progressed through one year of school each year. The school starting age of eight is well documented (South African Bureau of Racial Affairs, 1955). The assumption that students progressed through one year of school each year is strong for a developing country, but is a necessary assumption for my identification strategy. Second, I assume that children attend school in their province of birth which is reasonable given the low internal migration rates and the migration restrictions in this time period. Finally, I must assume that Natal and the other provinces had similar pre-trends in the variables of interest.

My first specification compares individuals in the two cohorts of interest who were born in Natal to those born outside of Natal:

$$y_{ijmp} = \alpha + \beta young_j * TCF_m + \theta_j + \gamma_m + \rho_p + \varepsilon_{ijp} \quad (1)$$

for individual *i* born in year *j* in municipality *m*, and who lives in province *p*. Dummy variables for *TCF* and *young* are absorbed by fixed effects for birth municipality and for each age. Fixed effects are also included for province of residence in 1980. The parameter β is the difference-in-difference estimator for the effect of increasing years of mother tongue instruction by two additional years.

The second specification allows for more flexibility between provinces and estimates a separate β for each province:

$$y_{ijmp} = \alpha + \beta_1 young_j * Transvaal_m + \beta_2 young_j * Cape_m + \beta_3 young_j * Freestate_m + ... + \varepsilon_{ijp} \quad (2)$$

where the same fixed effects are included as above. The parameters β_1, β_2, and β_3 represent the effect of increasing years of mother tongue instruction by two years in the Transvaal, Cape, and Free State provinces, respectively.

Note that these difference-in-difference estimates are intent to treat (ITT) effects, not average treatment effects (ATE). I do not require that any student actually attain grade 5 or grade 7 in order to be considered in the treatment group. However, if no one passed grade 4 or grade 6, we would not expect to find significant results.

The primary concern with this identification strategy is that there are differential trends in the outcome variables between Natal and the other three provinces. If income, for example, is growing faster in the Cape Province than in Natal, then we might see that the "young" cohort has higher wages in the Cape after the policy change just because of these differential trends.[70]

[70] I'm planning to test this using older cohorts who were not affected by the policy change. I will create a placebo test assigning "treatment" to a cohort who was unaffected and check that we do not see similar results.

3.4 Census Data and Descriptive Statistics

The analysis uses individual-level census data from a 100% sample of the 1980 South African census.[71] This census collects information about age, municipality of birth[72], province of residence, education level, yearly income, employment status, and the self-reported ability to speak, read or write English and Afrikaans. Additionally, for women, the questionnaire asks the number of children ever born to the individual.

Multiple measurement problems in the Census could potentially affect the results. First, the data exhibit significant age heaping at multiples of five as well as at age 32. Individuals who report one of these ages have lower average educational attainment and lower reported income. This age heaping does not fall at age 38, the cutoff between young and old cohorts, so it will only affect the results if it is differential across provinces. Table 3 provides evidence that the age heaping is relatively consistent across provinces.

Second, twenty-two percent of blacks were not enumerated in the census (South African Data Archive, 1980). If this under-enumeration was in areas with high returns to language skills (e.g. urban areas), then this will potentially affect my results if the extent of migration into these areas differed across provinces and cohorts. For example, if those who were more affected by the policy change were more likely to move to urban areas, then undercounting these individuals would lead to understating my results.

[71] The full count of the 1980 census is available on Statistics South Africa's website. Additional censuses are available from 1970, 1985, 1991, 1996, 2001, and 2007 but 1980 is the only census which includes detailed information about place of birth as well as income for the Black sample. The 2001 census reports province of birth as defined in 2001, but the provincial boundaries changed after the end of apartheid so it is difficult to create a 1-1 mapping between these reports and where the individual would have been born and educated. The 2001 census also only reports income in intervals.

[72] In 1980, there were 324 reported municipalities of birth. Including fixed effects for birth municipality allows me to control for any differences in school quality at this level.

Finally, the Black homelands which had already attained "independence" in 1980 were not enumerated even though in 1955 they were affected by the policy change. Migrants out of these homelands were enumerated in the census, so if these migrants are different than non-migrants then the difference-in-difference coefficient could be biased. I check my results using only blacks born outside of independent homelands.

A full list of the dependent and the independent variables used in the analysis are defined in Table 4. I restrict my attention to individuals who were between ages 28 to 48 in 1980 because younger cohorts were potentially affected by various disruptive incidents in schools in the 1970's, and older cohorts do not provide a good comparison group due to trends in educational attainment.[73] This age group is divided into young and old cohorts based on the age at the time of the language reform. As discussed above, the young cohort is defined as anyone who was born in or after 1942, or who is 38 or less in 1980.

Individuals are assigned to their province of birth based on the reported municipality of birth. I create dummy variables for having been born in Natal, being born outside of Natal ("TCF"), and individual dummies for being born in the Transvaal, Cape, or Free State provinces. As noted above, I assume that individuals are educated in their province of birth. I create dummy variables for province of residence in 1980 which are also included as controls in my primary specifications.

Log wages are defined as the natural log of reported yearly income for those who were employed at the time of the census. I explored potential selection into employment, but the language reform appears to have had no significant effect on the probability of being employed.

[73] Results are robust to changes in the age cutoff on either side.

Variables for the number of children ever born for women and years of education are obtained directly from the census.[74]

Finally, the census asks whether individuals can speak or can speak, read, and write, Afrikaans and English. I use these separate variables to construct an indicator variable for having some skill in one of the two languages, defined as being able to speak or speak, read, and write. A variable for being literate in a "labor market" language is defined as being able speak, read, and write either English or Afrikaans. I consider both languages together rather than individually due to the predominance of Afrikaans in the Free State and Transvaal provinces and of English in the Natal and Cape provinces.

Descriptive statistics for each cohort of men and women, respectively, are presented in Table 5 and Table 6. For both groups, education levels show an upward trend in both provinces, which appears to be partly driven by a larger percentage of individuals having attended school. Both literacy and the ability to speak English or Afrikaans are lower in Natal but increase between cohorts in both the treatment and control provinces.

Before analyzing the effects of the language reform, I first verify that the policy change did induce a greater number of years of instruction in mother tongue. In particular, I expect that *actual* years of exposure increased more outside of Natal, where the years of *potential* exposure increased by four years, than in Natal where years of potential exposure increased by only two years. Indeed, I find that, for men, years of exposure increases by just over a year in Natal. Similar patterns hold for women. The reason why the increase in actual years of exposure does not equal the potential years of exposure is because not every student stays in school through grade six, seven, or eight. In fact, around half of the population attained at least grade 5 so a

[74] Unfortunately, the 1980 census data lumps together grades 1 through 3 so anyone who attends school but does not pass third grade is assigned three years.

significant part of my sample was affected by the reform. This is important given that my difference-in-difference coefficients are intent to treat estimates.

3.5 Results

A. Men

I begin by examining the effect of the language reform on wages and other adult outcomes for cohorts of men affected by the policy change in different ways across provinces. I first consider the comparison between young (affected) and old cohorts and the difference between Natal and the other three provinces. Difference in difference estimates of β for men in specification (1) are presented in Panel A of Table 7. Second, I compare each province separately to Natal in Panel B. This panel presents the estimates for each β_i in specification (2). Separate regressions are presented for seven different outcome variables. Recall that the coefficients represent the effect of increasing years of instruction in a student's mother tongue from by two additional years.

Column (1) suggests that the effect of increasing mother tongue instruction for two years was to raise wages later in life by 1.4%. This relationship is marginally significant in the whole country, but is larger (3.4%) and highly significant in the Free State province. The estimates are between 20 to 50 percent of the return to a year of education of 7%. Wages in the Free State province are about 20% lower on average than the other three provinces which might explain why I see larger effects on wages in this province

The remainder of the table explores other outcomes which may help explain the increase in wages from this policy change. First, I consider whether the policy increased levels of education. I find that there was a response to the policy change of 0.089 years more of education

in the whole country. The effect was larger at 0.117 years in the Transvaal and was 0.086 years in the Free State. There were no significant effects on education in the Cape Province. It appears that slightly less than half of the effect of the program on wages can be explained by observed increases in years of education. That is, given a return to schooling of 7%, an increase of 0.089 years of education would increase wages by 0.64% (=0.089*7). This is almost half of the 1.4% increase in wages in the country as a whole.

Beyond years of schooling, two additional outcomes that may be affected directly by mother tongue education are ability to speak English or Afrikaans and the ability to read and write, either in one of these languages or in one's mother tongue. In column (3), I find negative effects of mother tongue instruction on the ability to speak English or Afrikaans. In the Transvaal, students are 1.5 percentage points less likely to speak English or Afrikaans, while in the Free State they are 4.1 percentage points less likely. These estimates are quantitatively large, representing a decrease of 2.8% and 6.6%, respectively, on the national mean ability of 60%.

Column (4) presents results for literacy in English or Afrikaans, measured as having the ability to speak, read, and write one of these languages. The effect seems to be homogenous across provinces. The estimated effect of increasing mother tongue instruction by two years is a 1.5 percentage point increase in the probability of being literate in English or Afrikaans. Again, these effects are large, showing a 3.5% increase on the national mean of 41%.[75] In column (5), I show some weak evidence that ability to read and write in any language went up, suggesting positive human capital effects of the policy change. The policy seems to reduce spoken English fluency, which can be absorbed as a by-product of classroom instruction in English, but to have

[75] Ideally, I would like to use this increase in literacy induced by the policy change for an IV estimate of the return to literacy in the 1980 South African labor market. However, since the policy change also induced a response in terms of education, it is not a valid instrument for literacy.

increased written literacy in English, which is a formal subject, the learning of which appears to improve when taught in one's native tongue.

Finally, I consider whether higher wages might be a product of a more mobile population moving to cities or high wage areas. In column (6), I find that those affected by the policy change are slightly more likely to have changed municipalities between 1975 and 1980, but that the effect is not large.

In Table 10, I consider to what extent we can explain the increase in wages from the increase in mother tongue instruction by including education, English/Afrikaans speaking ability, literacy, and migration in the regression. The original increase in wages of about 1.4% decreases by almost half to 0.08% when including education as an additional control in the regression, suggesting that the effect on education accounts for half of the effect on wages. When I add ability to speak English or Afrikaans, the estimate goes down slightly to 0.06%. The second variable that seems to matter after education is ability to read or write English or Afrikaans. Including literacy makes the coefficient decrease to 0.03%. The results in this table suggest that most of the effect on wages goes through increases in education and literacy.

B. Women

Similar results are found for women. Table 8 presents estimates for women of specification (1) and (2) in Panel A and Panel B, respectively. An additional outcome variable, number of children ever born, is considered for women.

Column (1) shows significant effects on wages only in the Transvaal where wages increase by 2.3%.[76] The weak effect of the program on women's wages occurs despite a significant effect on educational attainment for women in column (2). The effect was an increase

[76] Women are much less likely to be employed than men and are found in a smaller range of different jobs.

of 0.155 years in the whole country, 0.164 in the Transvaal, and 0.199 in the Free State. These results lend support to the hypothesis that mother tongue instruction increases accessibility to school and retention in school for girls. This also illustrates the fact that using education attainment in the difference-in-difference specification would be incorrect since education levels seem to have endogenously responded to the change.

Estimates of the effect on English or Afrikaans speaking ability and literacy in columns (3) and (4) are similar to those for men. Increasing years of mother tongue instruction by two years decreased the ability to speak English or Afrikaans by 1.7 percentage points in the Free State. The effect on literacy is an increase of 2.2 percentage points in the whole country, 1.8 percentage points in the Transvaal, and 3.5 percentage points in the Free State. Women are also more likely to be able to read and write any language; column (5) shows that the probability goes up by between 1.3 and 3.2 percentage points.

In column (6), I find little evidence of those who were affected by the policy are more mobile. Finally, for women only, I consider whether this policy change resulted in a change in fertility patterns. We might expect that fertility decreases with education levels and therefore that since this policy induced an increase in education, we will also see a decrease in fertility. In fact, this pattern is borne out in Column (7). Total fertility of women affected by the policy goes down by almost 0.1 children in the Cape Province and by almost 0.2 children in the Free State Province.

C. Robustness

In Table 10, I consider a variety of robustness checks. As noted above, we might worry about non-enumeration of independent homelands in 1980. I therefore drop those born in the

four homelands which were independent by 1980.[77] Column (2) restricts to those born outside of independent homelands. If anything, results are stronger when excluding these homelands. This is most likely because of selective migration out of the homelands; those from the Transkei were most likely to be working in low wage jobs as mine labor or truck drivers.

In Column (3), I show that the results are similar when using age and age-squared as controls instead of a series of dummy variables for each age. Columns (4) and (5) assume that students start school at age 9 and age 10, respectively; this is equivalent to starting school at age 8 and repeating one or two grades. Results are robust to varying the time at which students start school.

3.6 Conclusion

The Bantu Education Act provides a way to establish the effects of mother tongue instruction on a range of outcomes in 1980. The results taken together are consistent with the hypothesis that mother tongue instruction increased literacy and educational attainment. Mother tongue instruction appears to have had positive effects on school retention for girls, which was one of the hypothesized effects.

This paper illustrates that the effects of mother tongue instruction can be different depending on context. In Morocco, Angrist and Lavy (1998) found that Arabic instruction reduced French reading and writing skills and that this was reflected in lower wages. Angrist et. al (2008) find no effects of Spanish instruction on English skills. My results are opposite in that mother tongue instruction increased literacy, this was reflected in slightly higher labor market outcomes, and I find significant negative effects on spoken language skills.

[77] These four homelands were called TVBC, short for Transkei, Venda, Bophutatswana, and Ciskei. The Transkei, Ciskei, and Venda were located entirely within the Cape Province. Bophutatswana was primarily in the Transvaal Province.

Overall, it seems that mother tongue instruction might have positive effects on school quality as evidenced through the increases in literacy of men and women. This cannot be disentangled from the relative benefits of teaching in English versus teaching English as a subject, as was done after the reform. The estimates presented in this paper most likely overstate the effects of simply changing the language of instruction, since they include effects on instructional quality and the ability of parents to help their children with homework.

In post-*apartheid* South Africa, most children are taught in English from grade one because their parents view this as the best way to learn. Results from this paper suggest that this is perhaps too early for students to begin in English.

Tables

Table 3.1: Years of Potential Mother Tongue Instruction by Birth Province and Birth Year

Cohort	Birth Year	Years of Mother Tongue Instruction	
		Natal	TCF
Old	1932	6	4
Old	1933	6	4
Old	1934	6	4
Old	1935	6	4
Old	1936	6	4
Old	1937	6	4
Old	1938	6	4
Old	1939	6	4
Old	1940	6	4
Old	1941	6	4
Old	1942	6	4
Young	1943	8	8
Young	1944	8	8
Young	1945	8	8
Young	1946	8	8
Young	1947	8	8
Young	1948	8	8
Young	1949	8	8
Young	1950	8	8
Young	1951	8	8
Young	1952	8	8

Table 3.2: Years of Potential Mother Tongue Instruction by Cohort and Birth Province

	Born before 1943 ("Old")	Born in or after 1943 ("Young")	Difference:
TCF	4	8	8 – 4
Natal	6	8	8 – 6
		Difference in Difference:	6 – 4

Note: TCF refers to the Transvaal, Cape, Free State provinces

Table 3.3: Percentage of Men Reporting Each Age by Birth Province

Age	Transvaal	Natal	Free State	Cape
28	7.84	7.87	8.03	7.85
29	6.16	6.25	6.09	6.36
30	9.03	8.89	8.48	8.56
31	4.76	4.75	4.62	4.69
32	6.54	6.3	6.37	6.36
33	4.38	3.76	4.19	4.65
34	4.66	4.38	4.28	4.68
35	5.74	5.64	5.82	5.4
36	5.15	5.34	5	4.79
37	3.37	3.33	3.37	3.77
38	5.66	5.75	5.47	5.37
39	4.11	4.33	3.93	4.67
40	7.21	7.33	7.26	7.64
41	2.88	2.53	2.77	3.09
42	4.12	4.4	4.08	4.14
43	2.72	2.68	2.66	2.8
44	2.59	2.5	2.34	2.6
45	4.6	5.07	4.83	4.16
46	2.6	2.64	2.97	2.64
47	2.2	2.3	3.71	2.36
48	3.67	3.97	3.73	3.44

Table 3.4: Variable Definitions

Independent Variables	
TCF	Born in Transvaal, Cape or Free State Provinces
Young	Born in or after 1943
Transvaal	Born in the Transvaal Province
Cape	Born in the Cape Province
Freestate	Born in the Free State Province
Dependent Variables	
Years of education	Number of years of education completed
Log wage	Natural log of reported income
Speaks White	Speaks English or Afrikaans
Literate White	Reads and Writes English or Afrikaans
Literate Any	Reads and Writes in any language
Migrant	Changed municipalities in past five years
Employed	Reports being an Employer or Employee
Children	Number of children ever born (women)

Table 3.5: Summary Statistics for Men, 1980

	All	Natal "Old"	TCF "Old"	Natal "Young"	TCF "Young"
Age	36.38	43.02	42.95	32.47	32.46
	(5.96)	(2.96)	(2.95)	(3.24)	(3.21)
Income	1517.28	1542.11	1505.34	1520.07	1521.3
	(1232.26)	(1341.43)	(1339.05)	(1202.79)	(1154.06)
Years of school	4.24	3.5	3.75	4.29	4.61
	(3.64)	(3.61)	(3.61)	(3.66)	(3.62)
Percentage with no education	0.35	0.44	0.41	0.34	0.31
	(0.47)	(0.49)	(0.49)	(0.47)	(0.46)
Percentage with more than 4 years of education	0.49	0.39	0.43	0.49	0.54
	(0.49)	(0.48)	(0.50)	(0.50)	(0.50)
Average years of mother tongue education	3.57	2.85	2.27	4	4.34
	(3.01)	(2.68)	(1.91)	(3.25)	(3.24)
Speaks English or Afrikaans	0.60	0.37	0.61	0.43	0.66
	(0.49)	(0.48)	(0.49)	(0.50)	(0.47)
Literate in English or Afrikaans	0.41	0.30	0.37	0.37	0.45
	(0.49)	(0.46)	(0.48)	(0.48)	(0.50)
Potential years of mother tongue exposure		6.00	4.00	8.00	8.00

Note: Standard deviations in brackets

Table 3.6: Summary Statistics for Women, 1980

	All	Natal "Old"	TCF "Old"	Natal "Young"	TCF "Young"
Age	35.53	44.29	44.2	30.8	30.72
	(7.55)	(3.76)	(3.70)	(4.06)	(4.04)
Income	773.67	811.57	731.76	812.25	786.08
	(1055.95)	(1438.93)	(1152.75)	(1008.72)	(961.61)
Years of school	4.11	3.06	3.44	4.07	4.64
	(3.61)	(3.39)	(3.54)	(3.56)	(3.60)
Percentage with no education	0.37	0.49	0.45	0.36	0.31
	(0.49)	(0.49)	(0.50)	(0.47)	(0.46)
Percentage with more than 4 years of education	0.56	0.43	0.47	0.56	0.62
	(0.50)	(0.50)	(0.50)	(0.50)	(0.48)
Average years of mother tongue education	3.55	2.60	2.10	3.85	4.40
	(3.08)	(2.67)	(1.94)	(3.21)	(3.28)
Speaks English or Afrikaans	0.57	0.31	0.57	0.39	0.66
	(0.49)	(0.46)	(0.49)	(0.49)	(0.47)
Literate in English or Afrikaans	0.41	0.26	0.35	0.35	0.48
	(0.49)	(0.44)	(0.48)	(0.48)	(0.50)
Children ever born	3.07	4.24	4.24	2.51	2.40
	(2.82)	(3.24)	(3.27)	(2.36)	(2.27)
Potential years of mother tongue exposure		6.00	4.00	8.00	8.00

Note: Standard deviations in parentheses.

Table 3.7: Difference in Difference Estimates of the Effect of Increasing Mother Tongue Instruction by Two Years: Men

	(1) Log Wages	(2) Years Education	(3) Speak White	(4) Literate White	(5) Literate Any	(6) Migrant	(7) Employed
Panel A: All Provinces							
TCF	0.014*	0.089*	-0.018***	0.015***	0.009*	0.002***	-0.002
	(0.008)	(0.048)	(0.006)	(0.005)	(0.006)	(0.001)	(0.005)
R-Squared	0.1182	0.1351	0.1915	0.1141	0.0939	0.0763	0.1046
Panel B: Provinces Entered Separately							
Cape	0.009	0.037	-0.014**	0.012	0.010	0.011***	-0.010
	(0.011)	(0.058)	(0.006)	(0.009)	(0.006)	(0.003)	(0.011)
Transvaal	0.012	0.117**	-0.015**	0.016**	0.008	-0.001	0.001
	(0.009)	(0.060)	(0.005)	(0.006)	(0.007)	(0.001)	(0.005)
Free State	0.034***	0.086*	-0.041***	0.019***	0.016**	0.0001	-0.001
	(0.009)	(0.046)	(0.005)	(0.005)	(0.007)	(0.001)	(0.040)
p-value	0.086	0.066	0.000	0.004	0.094	0.014	0.602
R-Squared	0.1183	0.1352	0.1915	0.1141	0.0939	0.0763	0.1046

Notes: N = 1,415,005 for column (1) and N=1,816,640 for the remaining columns. Standard errors are clustered at the level of birth municipality. Coefficients presented are the difference in difference coefficients (β) or the coefficient on the interaction between the young cohort and the respective province. All regressions include fixed effects for municipality of birth, province of residence, and dummies for each age. P-values are reported for the null hypothesis that all three young*province coefficients are zero. The regressions restrict to ages 28-48.

Table 3.8: Mechanisms, Outcome = Log(Income)

	(1)	(2)	(3)	(4)	(5)
TCF	0.014*	0.008	0.006	0.0031	0.0030
	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
Extra Controls		E	E, S	E, S, L	E, S, L, M
R-squared	0.1183	0.2102	0.2108	0.2156	0.2156

Notes: N = 1,415,005. Coefficients are presented for the young*TCF variable. The letters in "Extra Controls" refer to other variables included in the regression. "E" = Education level; "S" = Speaks English or Afrikaans; "L"= Literate in English or Afrikaans; "M" = migrated in last five years. Standard errors are clustered at the level of birth municipality. All regressions include fixed effects for municipality of birth, province of residence, a dummies for each age. Regressions restrict to ages 28-48.

Table 3.9: Difference in difference Estimates of the Effect of Increasing Mother Tongue Instruction by Two Years: Women

	(1) Log Wages	(2) Years Education	(3) Speak White	(4) Literate White	(5) Literate Any	(6) Migrate in past 5 years	(7) Children Ever Born
Panel A: All Provinces							
TCF	0.019	0.155***	0.001	0.022***	0.023***	0.001	-0.064
	(0.011)	(0.050)	(0.005)	(0.006)	(0.008)	(0.001)	(0.039)
R-Squared	0.2417	0.2325	0.2844	0.1817	0.1932	0.0373	0.1186
Panel B: Provinces Entered Separately							
Cape	0.013	0.094	-0.001	0.023	0.013	0.006***	-0.092*
	(0.018)	(0.113)	(0.015)	(0.017)	(0.015)	(0.002)	(0.051)
Transvaal	0.023*	0.164***	0.007	0.018***	0.024***	-0.006	0.016
	(0.012)	(0.050)	(0.006)	(0.005)	(0.008)	(0.005)	(0.031)
Free State	0.017	0.199***	-0.017***	0.035***	0.032***	0.001	-0.194***
	(0.017)	(0.049)	(0.004)	(0.006)	(0.008)	(0.006)	(0.041)
p-value	0.085	0.002	0.808	0.000	0.003	0.128	0.132
R-Squared	0.2417	0.2326	0.2844	0.1818	0.1932	0.0374	0.1186

Notes: N = 643,236 in column (1) and N=1,695,944 for the remaining columns. Standard errors are clustered at the level of birth municipality. Coefficients presented are the difference in difference coefficients (β) or the coefficient on the interaction between the young cohort and the respective province. All regressions include fixed effects for municipality of birth, province of residence, and dummies for each age. P-values are reported for the null hypothesis that all three young*province coefficients are zero. The regressions restrict to ages 28-48.

Table 3.10: Robustness

	(1) No Restriction	(2) Drop Homelands	(3) Age controls	(4) Start school 1 year later	(5) Start school 2 years later
Panel A: All Provinces					
TCF	0.014*	0.019**	0.013*	0.017*	0.013*
	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)
Panel B: Provinces Entered Separately					
Cape	0.009	0.027**	0.008	0.005	0.014
	(0.011)	(0.009)	(0.011)	(0.012)	(0.009)
Transvaal	0.012	0.012	0.011	0.017*	0.009
	(0.009)	(0.009)	(0.009)	(0.010)	(0.009)
Free State	0.034***	0.034***	0.032***	0.040***	0.030
	(0.009)	(0.009)	(0.009)	(0.101)	(0.010)
p-value	0.0015	0.0008	0.0029	0.0001	0.0123
N	1,415,005	1,195,642	1,415,005	1,415,005	1,415,005

Notes: Column (2) restricts to those born outside of independent homelands. Column (3) includes age and age-squared instead of a series of dummy variables for each age. Columns (4) and (5) assume that students start school at age 9 and age 10, respectively; this is equivalent to starting school at age 8 and repeating one or two grades. Standard errors are clustered at the level of birth municipality. Coefficients presented are the difference in difference coefficients (β) or the coefficient on the interaction between the young cohort and the respective province. All regressions include fixed effects for municipality of birth, province of residence, and dummies for each age. P-values are reported for the null hypothesis that all three young*province coefficients are zero. The regressions restrict to ages 28-48.

References

Angrist, Joshua and Victor Lavy. (1997) "The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco," *Journal of Labor Economics*, 15(1).

Angrist, Joshua, Aimee Chin, and Richard Godoy (2008) "Is Spanish-only Schooling Responsible for the Puerto Rican Language Gap?" *Journal of Development Economics.*

Case, Anne and Angus Deaton. (1999) "School Inputs and Educational Outcomes in South Africa," *Quarterly Journal of Economics,* 114(3).

Case, Anne and Motohiro Yogo. (1999) "Does School Quality Matter? Returns to Education and the Characteristics of Schools in South Africa," NBER Working Paper 7399.

Chiswick, Barry R. (1991) "Speaking, Reading, and Earnings among Low-Skilled Immigrants," *Journal of Labor Economics.* 9: pp 140-170.

Dustmann, Christian. (1994). "Speaking Fluency, Writing Fluency, and the Earnings of Migrants". Discussion Paper no. 905. Center for Economic Policy Research.

Harsthorne, Kenneth Brown. (1992) *Crisis and Challenge: black education 1910-1990.* Oxford University Press, Cape Town.

Mariotti, Martine (2009). "Labor Markets during Apartheid in South Africa". Australian National University Working Papers in Economics and Econometrics No. 503.

McManus, Walter, William Gould, and Finis Welch. (1983). "Earnings of Hispanic Men: The Role of English Language Proficiency." *Journal of Labor Economics* 1: pp101-103.

South African Bureau of Racial Affairs (1955). *Bantu Education: Oppression or Opportunity*. Stellenbosch.

South African Data Archive (1980). Codebook to the 1980 Population Census. http://sada.nrf.ac.za/CodebookPDF/S0071.pdf.

Taylor, Nick and Penny Vinjevold (1999). *Getting Learning Right*. Report of the President Education Initiative Research Project, Joint Education Trust, Johannesburg.