Multiple Measures of Historical Intergenerational Mobility: Iowa 1915 to 1940*

James J. Feigenbaum†

February 22, 2017

Abstract

Was intergenerational economic mobility high in the early twentieth century in the United States? Comparisons of mobility across time are complicated by the constraints of the data available. I match fathers from the Iowa State Census of 1915 to their sons in the 1940 Federal Census, the first state and federal censuses with data on income and years of education. With this linked sample, I can estimate intergenerational mobility between 1915 and 1940 based on earnings, education, occupation, and names. Across all these measures, I document broad consensus that rates of persistence were low in Iowa in the early twentieth century. Within my sample, rural sons from Iowa had more intergenerational mobility than their urban peers and the grandchildren of the foreign-born were more mobile than the grandchildren of the native-born.

*This article was previously circulated as *A New Old Measure of Intergenerational Mobility: Iowa 1915 to 1940*. I thank Claudia Goldin, Lawrence Katz, Rick Hornbeck, Nathan Nunn, Kjell Salvanes, Gary Solon, and Sandy Jencks for detailed comments and suggestions, as well as Raj Chetty, Ed Glaeser, Simon Jaeger, Akos Lada, Robert Margo, Christopher Muller, Martin Rotemberg, Bryce Millett Steinberg, Patrick Turley, two anonymous referees, and seminar participants at the Cliometrics World Congress, the Harvard Labor Lunch, the Harvard Economic History Tea, and in the Harvard Multidisciplinary Program in Inequality & Social Policy. Viroopa Volla, Andrew Creamer, Justin Meretab, and especially Alex Velez-Green all provided excellent research assistance. This research has been supported by the NSF-IGERT Multidisciplinary Program in Inequality & Social Policy at Harvard University (Grant No. 0333403). Replication data and programs for this paper are available online.

†Industrial Relations Section, Department of Economics, Louis A. Simpson International Building, Princeton University, Princeton, NJ 08544, email: `jjf2@princeton.edu`, site: `http://jamesfeigenbaum.github.io`.

How strong is the link between a child's outcomes in adulthood and the accident of his or her birth? And how does economic mobility in the early twentieth century compare to mobility today? How much more common were Horatio Alger's rags-to-riches heroes in the early twentieth century than in the early twenty-first? Comparing mobility over time is complicated by different measures of mobility across studies; often the chosen measures are influenced and constrained by the available data. In historical work on intergenerational mobility, income or earnings data is rare and so occupations and (more recently) names are the proxies for status used.[1] In contemporary data, scholars often calculate mobility with earnings and education.[2] But recent studies on trends in intergenerational mobility are unable to trace income mobility before the 1980s (Lee and Solon, 2009; Chetty et al., 2014b).

I take advantage of two historical data sources that enable me to measure mobility in many different ways—including intergenerational elasticity of earnings, rank-rank persistence, occupational score elasticity, occupational transitions, education persistence, imputed status based on both given names and rare surnames. Because I can estimate these various mobility measures within the same source, I shed light on how well these various measures agree with one another, at least for the early twentieth century. I match fathers from the Iowa State Census of 1915 to their sons in the 1940 Federal Census, the first state and federal censuses with data on income and years of education.[3]

I find that these measures of intergenerational mobility are quite consistent. I find generally high rates of mobility across all measures. These measures are also internally consistent: I find more mobility for the sons of urban Iowa than for the sons of rural Iowa, as well as more mobility for the grandsons of the foreign born than native-born grandsons.[4]

Was there more mobility historically than today? I find a lower intergenerational income elas-

[1]While the United States federal census began collecting information on respondents' occupations in 1850, the census did not include data on either years of educational attainment or annual income until 1940.

[2]Even with earnings data, sharp disagreements remain over whether to estimate intergenerational elasticities or rank-rank mobility and the effects of the observation of multiple years of earnings, possibly at the household or individual level, enabled by administrative data.

[3]My study builds on the earlier mobility work of Parman (2011), who also draws on the 1915 Iowa State Census to measure intergenerational mobility, linking men in the 1915 Iowa census back to their households in the 1900 census and then finding their fathers and sons in Iowa in 1915.

[4]I use grandparent nativity rather than son or father nativity because the subsamples are more even: nearly one-third of my sample have four foreign-born grandparents, another one-third have four native-born grandparents, and the rest 1, 2, or 3. A much smaller share have foreign-born parents and only a handful of the sons are foreign born themselves. I am unable to estimate mobility across race in my sample because Iowa's population in 1915 was nearly all white. Hertz (2009) documents strong disparities in mobility between blacks and whites, arguing that American immobility in the recent period is driven by extremely high persistence of outcomes among African Americans.

ticity during the first half of the twentieth century in the US than studies find in the second half of the century. I also measure intergenerational mobility using the rank-rank parameter (Chetty et al., 2014a) and similarly find more mobility historically. The results for education, occupational mobility—measured either using transition matrices or occupation score—and name-based mobility all point in the same direction. But such differences between contemporary and historical mobility could be spurious. For one, my sample is not a random draw of the American population. By population density, urbanisation, and the share foreign born, Iowa is nearly the median state. On other dimensions, Iowa in 1915 is an outlier: it was also almost universally white, invested more and earlier in education than other states, and had relatively low levels of inequality. In addition, measurement error, either due to the historical nature of my data, the difficulty of creating longitudinal linked samples, or the single year I am able to observe fathers and sons in the census may push me to find excess mobility. I work to mollify these concerns in a few ways. In the earnings data, I show that the estimated differences between contemporary and historical mobility remain after adjusting the contemporary sample to mirror the historical sample in measurement noise and demographic and geographic composition. The agreement across many measures of mobility, some based on outcomes like education and occupation that should be less noisy than a single year of earnings, also sharpens the comparative result. Ultimately, my results do not prove there is less mobility today than in the past, but, taken along with other evidence comparing mobility over the twentieth century (Parman, 2011; Long and Ferrie, 2008), strengthens the belief that there was more mobility in the early twentieth century.

The paper proceeds as follows. In the first section, I discuss the historical data that I draw on and my data collection and census linking procedures. I describe what, if any, bias the census linking procedure may induce and compare Iowa in 1915 to the rest of the nation. In section two, I review the various methods of measuring intergenerational mobility that I am able to apply to my sample of linked fathers and sons. In the third section, I present my estimates of intergenerational mobility in the early twentieth century for income, education, occupation, and name-based status. Section four concludes the paper.

1 Data and Record Linkage

1.1 Linking the Iowa 1915 and Federal 1940 Censuses

I draw my primary data for measuring intergenerational mobility in the United States early in the twentieth century from the 1915 Iowa State Census and the 1940 US Federal Census, both of which include measures of the earnings, education, and occupations of the respondents. I describe both data sources in this section, as well as the method used to link fathers in 1915 to their sons in 1940.

The 1915 Iowa Census enumerated the 2.3 million Iowa residents in 1914. It was the first American census of any kind to include data on both annual income and years of education in addition to more traditional census measures, and it also includes respondent name, age, place of residence, birthplace, marital status, race, and occupation. I use the Iowa State Census sample digitised by Claudia Goldin and Lawrence Katz for their work on the historical returns to education (Goldin and Katz, 2000, 2008). The Goldin-Katz sample includes 26,768 urban residents (5.5% of the total urban population of Iowa in 1915) and 33,305 rural residents (1.8% of the total rural population). Figure 1 presents a map of the counties and cities included in the Goldin-Katz sample. The three large Iowa cities sampled are Des Moines, Davenport, and Dubuque.[5] In 1915, the population of Des Moines was approximately 97,000 people, making it the 64th largest city in the country. Davenport and Dubuque were smaller, with approximately 46,000 and 39,000 people, respectively. The rural counties in the sample were selected by Goldin and Katz on the basis of both image and archive quality, as well as to provide a diverse geographic sample within the state, as shown in Figure 1.[6]

[Figure 1 about here.]

To construct my sample for census matching, I limit the Goldin-Katz sample to families with boys aged between 3 and 17 in 1915. These sons will be between 28 and 42 when I observe them again in 1940, which should reduce measurement issues due to life cycle variability in annual income. I restrict my analysis to sons in 1915, because name changes make it impossible to locate most daughters in the 1940 Census. This leaves me with a sample of 7,580 boys in Iowa in 1915, 6,071 of

[5] The census manuscripts for Sioux City, one of the other large cities in Iowa, were unreadable and not collected by Goldin and Katz (Goldin and Katz, 2000).

[6] For more details on the construction of the Goldin-Katz sample, see Goldin and Katz (2000).

whom have fathers in their households, and the requisite data on both the father's education and income.

To locate these sons in 1940, I utilise the 100% 1940 census sample deposited by Ancestry.com with the NBER. I collect the set of possible matches, using the son's first and last name, middle initial (when available), state of birth, and year of birth. Then, I train a record-linking algorithm and use the scores generated by that algorithm to identify the correct matches for each son from 1915 in the 1940 data.[7]

Once the matched sons are identified, I record the pertinent data from the 1940 census, enumerated in the shadow of the Great Depression and on the eve of WWII.[8] I use the 1940 census because it is the only census suitable for tracking the sons of Iowa. The 1940 Census was the first federal census to collect data on incomes, weeks of work, and years of education of the entire population.[9]

[7] The machine learning approach, which I detail in Feigenbaum (2016a), teaches an algorithm how to replicate the careful hand-linking work of a researcher, but at scale and with extreme consistency. I generate the training data for the algorithm by manually linking a 30% random sample of sons from Iowa in 1915 into the 1940 census. To be considered as a possible link, records must first match exactly on state of birth, be within 3 years' difference in year of birth, and within 0.3 Jaro-Winkler string distance in both first and last names. Then, within that filtered list of possible matches, the records are double-entry matched to the 1940 census manually by trained research assistants. Records determined to match are marked as such, records without a clear and certain match in the 1940 census are marked as unmatched. With this corpus of links, I then train a match algorithm. The match algorithm is used to reduce between-researcher variability in match quality, to speed up the matching process, and ensure data replication. The method improves on previous efforts based on phonetic matching because typos and transcription errors will not cripple the matching. The matching algorithm uses Jaro-Winkler string distances in first and last names, exact matches on state of birth, absolute difference in year of birth, Soundex matches for first and last names, middle initial matches, matching first and last letters of first and last names, and other record-based variables to predict whether a record is a true or false match. The algorithm also factors in the match quality of other possible matches for the given record searched for, only making matches when a record is a significantly better match than other possibilities. Based on cross-validated out of sample predictions within my training data, the match algorithm has a true positive rate of nearly 90% and a positive prediction rate of 86%. In Table A.3 of the appendix, I detail the exact weights on the match algorithm used.

[8] How might the war and Depression affect my analysis? I expect limited effects of WWII. While the war in Europe and Asia were well underway in 1940, Pearl Harbor was still nearly two years away during the April 1940 enumeration. There was no war mobilisation in the United States in 1939 or 1940. The US spent around 2% of GDP every year on defence from 1931 to 1940, compared to 5.6% in 1941 and 16% in 1942. Spending peaked at 41% in 1945. Beyond direct defence spending, US production for the war effort is also non-existent in 1939. Cash-Carry, for example, did not begin until September 1939 (Lend-Lease in 1941) and production did not ramp up until well after the 1940 census was taken. American shipyards produced as many ships in 1941 as they had from 1938 to 1940 (Tassava, 2008). The price and wage controls (so-called 'General Max') were instituted during 1942, targeting March 1942 levels and suggesting that the war (or wartime policy) effects on wages or the wage distribution as of 1939 or 1940 were limited. The Great Depression's effect is more difficult to estimate with only 10 counties and 3 cities in Iowa in my sample. Feigenbaum (2016b) shows cities with more severe Depression downturns had lower mobility, but those effects are all relative. The overall effect of the Depression on mobility is an open question.

[9] The earnings in the 1940 census are top-coded at $5000. However, only 44 of the sons in my sample report such high earnings; I code them as earning $5000 for my analyses. Past federal censuses record contemporary school enrollment for each person (child), but not years of schooling completed for adults no longer in school. Earnings data was collected in 1940 only for wage and salary workers. The data collected are the 'total amount of money wages or salary' but enumerators were instructed: 'Do not include the earning of businessmen, farmers, or professional persons derived from business profits, sale of corps, or fees'. For more, see `https://usa.ipums.org/usa/voliii/inst1940.shtml#584`. The importance of this missing data will vary with the fraction of farmers and other business owners in

While such data was also collected in 1950 and 1960, federal censuses are privacy-restricted for 72 year after enumeration. The data on names required for linking the 1950 census will not be accessible until 2022, 2032 for the 1960 census.[10] Because it is a national sample, I do not have to worry about losing many sons to out-migration, which might otherwise bias my estimates.

My match rate is roughly 59%, which surpasses the rates of previous literature linking between censuses.[11] The match rates for the rural and urban samples are comparable: I match 57.3% of sons growing up in urban areas in 1915 and 60.4% of sons from rural areas. Ultimately, I have a sample of 4,478 father-son pairs; this size is also comparable to many other projects measuring intergenerational mobility, both historically (Long and Ferrie, 2013) and recently (Lee and Solon, 2009).

1.2 What Predicts Matches?

The complexity of linking individuals between historical datasets could introduce bias to any downstream analyses. However, I argue that my final sample does not suffer any crucial construction defects because simple transcription errors are the most likely obstacle to linking between a son observed as a child in 1915 and as an adult in 1940. To test this, I calculate a number of string- and character-based statistics using the first and last names of the sons in my sample and compare the magnitude of their effects on match rates as compared to more economically important variables.

First, I determine the name commonness of both the first and last name, relative to all names in the pooled IPUMS sample of the 1910 and 1920 censuses.[12] A more common name is less likely to have a unique match in the 1940 Census, even after limiting the possible targets by state of birth and year of birth.

Second, I calculate the length of each son's first and last name. Longer names are more likely

my sample. It does not, however, affect farm labourers, whose earnings are reported the same as other occupations. Of my matched sample, 13.7% of the sons in 1940 are farm owners or operators without income. Initially, I drop these observations with missing earnings data in analyses on income data. However, in Appendix A.3, I impute earnings for farmers using the 1950 census, which did collect data on capital income and non-wage and salary earnings. Using these imputed earnings, I estimate even higher levels of mobility than in my main results.

[10] I do make use of the 1% anonymised IPUMS sample of 1950 to impute earnings for farmers in 1940, as well as the whole sample as a check on the earnings data in 1940.

[11] Parman (2011) reports match rates of nearly 50% using hand matching. Guest et al. (1989) match at 39.4%. Other attempts at census linking using phonetic codes such as Soundex have lower match rates.

[12] The commonness statistic is measured as the share of 100 people in the pooled 1910 and 1920 sample with the same first (last) name. It ranges from 0.00118 (or roughly 1 person in 100,000 with the same name—these names are unique in my sample) to 1.72 for first names (John) or 1.02 for last names (Miller). Abramitzky et al. (2012) use relative commonness as a predictor of census match success as well.

to be incorrectly transcribed, but they are also more likely to be distinctive.

Third, I attempt to predict typographical errors using character similarity scores. Cognitive scientists and typographers have studied how likely certain letters are to be mistaken for one another or how similar two letters are visually. For example, readers are much more likely to mix up lower case p and q than they would be p for k. Further, some letters are more likely to be mis-transcribed than others: s is quite visually unique while l and n are both visually similar to other letters.[13] A name with a number of l's or n's in it is more likely to be mis-transcribed and thus not matched when I search in the 1940 Census.[14] I use a matrix of letter visual similarity from Simpson et al. (2013) to compute, for first and last names, a similarity score.[15]

Finally, I calculate a name's Scrabble score as an alternative measure of both name commonness and name simplicity.[16] Names with low Scrabble scores are likely to be made up of relatively common characters and are less likely to be changed or Americanised over time (Biavaschi et al., 2013).

Table 1 presents the results from a series of linear probability models, predicting whether or not a son in 1915 is uniquely matched ahead to the 1940 Federal Census. Sons with more common last names are less likely to be matched, while first name commonness has a smaller, positive effect. Sons with longer first names or first names with higher similarity scores are more likely to be found, but both of these effects are quite small.[17] The overall explanatory power of these variables in Table 1 is quite low; however, I argue that this underscores the randomness with which some sons are linked and others are not. Mismatches are driven by transcription errors (both by census enumerators and in contemporary data entry) which are largely random, even if some predictors (common names, letter similarity) have some power. I include controls for all of these name string

[13] l is likely to be confused with f and i for example, while n is similar to both h and m.

[14] Recall matches are made using census indices transcribed by Ancestry.com.

[15] Specifically, Simpson et al. (2013) conduct surveys of college students and other native and non-native English readers to assess the similarity of letters on a 7 point scale, where 7 indicates exactly the same and 1 extremely different. For example, i and l have a similarity score of 6.13, while w and t have a similarity score of exactly 1. I take the highest (non-self) similarity score for each letter as a measure of a letter's likelihood of being mis-transcribed. Figure A.1 in the appendix graphs these scores for each letter. Then, I calculate the average of these scores for all letters in a given string (name). The scores from Simpson *et al* are based on both lower case and upper case letters in block type. As many of the census files are in script, a visual similarity matrix for cursive letters would be ideal, but such a measure does not exist in the typography literature. As a robustness check, I also use a letter matrix of confusion probability from McGraw et al. (1994) and find a high correlation between each letter's similarity score.

[16] Biavaschi et al. (2013) introduce the use of Scrabble scores into the economic literature. They use this measure to predict name changes by immigrants to the United States during the early twentieth century. Scrabble point values were based, originally, on the frequency of letters on US newspaper front pages.

[17] With controls for commonness and length, the Scrabble scores do not seem to relate to match rates.

properties in all subsequent analysis.

[Table 1 about here.]

More serious issues could be generated by differential matching rates according to father, son, or family characteristics in 1915. In my sample I find little evidence that such characteristics strongly affect the probability of matching. In Table 2, I present the estimated effects of a set of variables observed for fathers and sons in 1915 on the probability of positively locating the son in the 1940 Census.[18] Each row in the table is a separate linear probability regression, reporting the coefficient of the listed X variable while controlling for first and last name commonness, length, letter similarity, and Scrabble score. I am slightly more likely to match sons who had higher income or more educated fathers (or mothers) in 1915, but these effects are both economically and statistically insignificant. For example, the probability of matching a son with a father at the 25th percentile of income is only 1 percentage point lower than matching a son with a father at the 75th percentile of income. Similarly small effects of both father's and mother's education can be seen as well. I am also less likely to match sons in the urban sample which follows from the slightly higher match rate among rural sons. I am also more likely to link sons born in Iowa, even after conditioning on name string characteristics.[19] All analysis undertaken in this paper will include controls for son's place of birth, place of residence in 1915, and, where appropriate, father's place of birth.

[Table 2 about here.]

1.3 How Does Iowa 1915 Compare to the US in 1915?

Iowa undertook an enumeration of its population in 1915 unlike any other in American history. In addition to gathering reliable 'characteristics of population [and] agricultural and other industries', the census also collected data on the earnings of its citizens and the extent of education in population. The census report announced that the finding 'happily confirms the claim of very high rank

[18] Results in this matching exercise are robust to alternative regression models, including logit and probit models. I use a simple linear probability model for ease of interpretation.

[19] 86% of the sons in my sample were born in Iowa so there is no difference between the 25th to 75th percentile for that covariate.

for Iowa in educational standing'.[20] That Iowa collected such earnings and education data makes the present study possible, but its rarity among contemporary censuses also raises a question: how unique was Iowa and how might that affect my estimated mobility rates?

How does my sample compare to the rest of the US in 1915? To be certain, the father and son pairs in my linked sample are not a random of the nation in 1915 and Iowa is not the median state. In Table 3, I compare Iowa to the rest of the nation and the Midwest in both 1910 and 1920 on a variety of dimensions.[21] In some ways, Iowa is a representative state: it is ranked in the middle of the United States in population density, urbanisation, and rural share, as well as sex ratio. Iowa was a more agricultural economy than the nation, but still one-third of states had more farms per capita.

However, the racial make-up of Iowa, its education, and inequality are outliers. The state was more than 99% white in both 1910 and 1920, even a few points whiter than the rest of the Midwest. Only 12% of the Iowa population was foreign born in 1910 (falling to 9% in 1920), lower than the national rate, but only slightly behind the median state. That contrasts with the prevalence of second-generation immigrants in Iowa: 28% of Iowa were the native born but to foreign parents in 1910, 8 points higher than the national rate.[22] Iowa's commitment to education—it was a leading state in the high school movement (Goldin and Katz, 2011)—is also apparent in Table 3. Iowa led the nation in 1910 and 1920 in literacy and was highly ranked in school enrollment as well. Goldin and Katz (2000) document the high returns to education in Iowa in 1915, both among white-collar and blue-collar workers Iowa was also above the median in newspaper circulation and number of newspapers (Gentzkow et al., 2011). Finally, when I calculate inequality using the distribution of farm sizes in the census, Iowa is the second most equal state in the continental US in both 1910 and 1920.

[Table 3 about here.]

How might the differences between Iowa in 1915 and the rest of the nation affect my analysis?[23]

[20]Preface to the Iowa 1915 Census Report, page iii. Given that Iowa was the first and only state to collect data on educational attainment, it is unclear how such a rank was calculated, but the fact that such data was even collected points to Iowa's standing.

[21]I include the states of Illinois, Indiana, Michigan, Ohio, Wisconsin, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota and South Dakota in the Midwest.

[22]I aggregate the census totals for native born with two foreign parents and native born with one foreign parent.

[23]Could any uniqueness about the year 1915 affect my analysis? I find that unlikely. The United States did not

Iowa's racial homogeneity may create more mobility than in the rest of the country. If the Gatsby Curve—as Krueger (2012) dubbed the correlation in OECD countries in the last decade between intergenerational mobility and inequality—held in the early twentieth century, again, perhaps Iowa will demonstrate more mobility than other states in this era because of its low inequality levels. Alternatively, the relatively low share of the foreign born may reduce mobility, if the children of recent immigrants are likely to experience fast upward mobility as they assimilate. In the analysis that follows, I take Iowa's uniqueness into account when comparing mobility over time and attempt to contrast my findings not with mobility overall but with contemporary estimates of mobility in an Iowa-like sample.

In the appendix, I present summary statistics on the men in my linked sample, exploring how the fathers and the sons in the sample differ from the matched fathers and sons. The first two columns of Table A.1 present summary statistics for the fathers of children between 3 and 17 in the Goldin-Katz Iowa State Census sample. Fathers found are the fathers for whom sons were located in the 1940 Census through Ancestry.com. Average yearly earnings for the fathers are approximately \$1000 in 1915 dollars. The average father had a half year more than a common school education (eight years) and was approximately 42 years old in 1915. Of the fathers in my sample, those fathers for whom I matched a son into the 1940 Census earned very slightly more, though not significantly so, measured either in levels, logs, and weekly earnings. The final two columns of Table A.1 present summary statistics for the Iowa sons in my sample. Only summary data for sons with complete information in the 1940 Federal Census is reported in the table, which lowers the number of observations in the final column to 3,971. The located sons earned more than \$1400 in 1940, which is roughly the same in real terms as the average earnings for their fathers in 1915; the lingering effects of the Great Depression may have reduced any real income gains overall.[24] Also notable is the fact that the sons had on average two more years of schooling than their fathers.

enter WWI until April 1917 and the draft did not begin until May 1917. Fighting in Europe began in July 1914, but little if any economic effects of the first six months of the war would have been felt by the residents of Iowa. Rockoff (2008) documents no rise in the industrial production or the stock market in response to the beginning months of European fighting. Changes in taxing, spending, and borrowing, particularly the famous liberty bond programme, were not introduced until October 1917. None of the fathers in my sample had been mobilised and any fathers working in industries that might later be subject to government price or wage controls were not subject to them in 1914. In the nearly 800 page report on the 1915 Iowa Census, published in 1916 by the enumerators, the ongoing European war was not mentioned.

[24] I measure all dollar amounts in this paper in nominal terms. Because I use logged earnings in my regressions and income for all sons is measured in 1940 and for all fathers in 1915, any nominal to real conversions drop out into the unreported constant term.

This is a striking example of the effect of the high school movement and the expansion of public education in Iowa, previously documented by Goldin and Katz (2008) and Parman (2011).

2 Measures of Intergenerational Mobility

The measure of intergenerational mobility is often constrained and influenced by the information available in linked intergenerational samples, both historically and in contemporary data. In this paper, because I observe earnings, occupations, education, as well as names, of the fathers and sons from Iowa 1915, I calculate a variety of different mobility measurements on a consistent dataset and am able to establish whether or not the measures produce similar results. In this section, I describe the various measures I am able to consider.[25]

2.1 Intergenerational Income Mobility

The most common measure of intergenerational economic mobility is the intergenerational elasticity of income (IGE), estimated by regressing the son's log income against his father's log income. Corak (2006), Solon (1999), and Black and Devereux (2011) present thorough reviews of the IGE literature.[26] These reviews all indicate a lack of historical data on intergenerational mobility: Corak (2006) documents 41 studies of the US IGE, none of which presents data before 1980. One aim of my project is to establish a correct measure of IGE well before the period previously studied in this literature.[27]

Corak's preferred measure of IGE in the US is 0.47, in line with the reviews presented by both Solon and Black and Devereaux.[28] Hertz (2009), who is able to split the sample by race, estimates

[25] One major limitation imposed by my data is that I am unable to estimate wealth mobility. Often drawn from probate records (Clark and Cummins, 2015), American federal censuses recorded data on wealth only three times (1850, 1860, and 1870). Ager et al. (2016), looking in the postbellum South, find high rates of mobility of wealth, in agreement with the findings I present later in this paper.

[26] The estimated elasticity of income between one generation and the next is commonly referred to as an IGE and I will use that abbreviation here.

[27] Aaronson and Mazumder (2008) take a different tack when measuring intergenerational income mobility. They use successive waves of the US federal census, from 1940 to 2000, and construct synthetic parents for observed individuals. They find low levels of mobility in 1940, but more mobility each decade until 1980. Mobility falls again in 1990 and 2000. However, the parents are constructed only using state of birth, age, and race; thus rather than regressing the son's income on the father's, they regress the son's income on the average income of men of the same race in the son's state of birth. While that is a possible proxy for the father's income, it does not seem sufficiently detailed or granular to detect small shifts in the intergenerational transmission of income.

[28] Corak also gives lower and upper bounds of between 0.40 and 0.52. Mazumder (2015) argues for a parameter of 0.6

an IGE of between 0.39 and 0.44 for whites, the relevant comparison here given the high white share in my Iowa sample. Corak (2006) also documents large variations between US studies measuring the intergenerational elasticity of income.[29]

However, the IGE imposes a very particular functional form relationship between the son's and father's earnings.[30] Chetty et al. (2014a) present evidence from recent US tax data that suggests this assumption is false; Corak and Heisz (1999) show the same with Canadian tax data. At both tails of the income distribution, the linear relationship between the father's log income and the son's log income breaks down. Following Dahl and DeLeire (2008), Chetty et al. (2014a) and Chetty et al. (2014b) estimate a rank-rank parameter of intergenerational mobility, regressing the son's income percentile (within his cohort) against the father's income percentile (within his cohort). The graphical evidence presented in Chetty et al. (2014a) suggests that the implied linearity in the rank-rank specification is a more accurate fit of the data. For their sample of the US, Chetty et al. (2014a) estimate a rank-rank parameter of 0.341 overall and 0.336 for sons. However, similar to the IGE literature, there are few historical estimates of the rank-rank parameter: Chetty et al. (2014b) plot trends in mobility for sons born between 1971 and 1993, but they cannot extend their sample further back in time.

I am able to estimate both IGE and rank-rank mobility in my data because I observe earnings for fathers in Iowa in 1915 and earnings for their sons in the 1940 Federal Census.

My study is not the first to use the earnings data in the Iowa 1915 census to estimate intergenerational mobility in the early twentieth century. Parman (2011) uses a clever multiple match technique to create father-son links within the Iowa 1915 census, even though the adult men in

[29]See, for example, the first table in Corak's appendix. IGE estimates in the literature range from 0.09 to 0.61. Because of this variation scholars have focused on measuring changes over time in IGE within one consistent dataset. However, the results in this literature have also been rather inconsistent. Mayer and Lopoo (2005) use the PSID and collect a sample of 30 year olds, regressing the son's income at age 30 on a three year average of the father's income. They find a large and statistically significant downward trend in the IGE, suggesting that mobility has increased significantly in the last several decades. Levine and Mazumder (2002) present more mixed results in work using the NLS, GSS, and PSID. Levine and Mazumder observe sons between the ages of 28 and 36 in 1980 and again in 1990. They find increasing mobility in the PSID, but decreasing mobility in the NLS and GSS. Lee and Solon (2009) argue that past work has been plagued by non-classical measurement error. To correct this, they argue that rather than observing the son's outcomes once or twice and throwing away the rest of the data, researchers should make use of the full sample. Drawing on PSID data for cohorts of sons and daughters born between 1952 and 1975, Lee and Solon do just that. I focus on the Lee and Solon results for fathers and sons to keep in line with the analysis I am able to perform in my data. Controlling for a quartic in the ages of both parents and children, they only limit the sample to sons between 25 and 48. They find a simple average IGE of 0.44 over the period and no statistically significant trends in IGE between 1976 and 2000.

[30]Specifically, a linear relationship between log father's income and log son's income.

the Iowa sample are no longer in their parent's households. Parman matches the men backwards in time to the 1900 Federal Census to construct childhood households. Though most are heads of household in 1915, these are Parman's 'sons' in the analysis. The reconstructed households yield the name, state of birth, and other demographic characteristics of the 'fathers' in 1900. Parman then matches these fathers forward in time to the 1915 sample, thus observing both fathers and sons in Iowa in 1915 and estimating IGEs based on income reported in the Iowa State Census. Parman finds an IGE of approximately 0.11 for all father-son pairs or 0.17 for non-farmer father-son pairs. These low estimates paint a picture of high levels of mobility in the early twentieth century.[31]

Parman finds very low rates of persistence in status. However, his analysis is limited in two key ways by the data available to him when he undertook his study.[32]

First, his sample was made up of very old fathers and very young sons. The average age of fathers in Parman's sample is between 57 and 65, depending on the particular specification. This age range is on the far right tail of the IGE studies in the literature and thus is very likely to present a very low IGE, due to life-cycle-induced measurement errors. The average age of the sons in Parman's sample is between 25 and 30, and this may also bias his results towards a very low IGE. As Grawe (2006) and Haider and Solon (2006) describe, trends in life cycle earnings—particularly, the fact that people with higher permanent income experience more earnings growth earlier in their career—can bias the estimate of β (Mazumder, 2015). Empirically, the sample age bias can be seen in Figure 2, based on the American IGE literature surveyed by Corak (2006). As the first figure shows, the older the average age of the fathers in the study samples, the lower the estimated intergenerational elasticity. The second figure shows a similar but weaker relationship holding in the opposite direction between estimated IGEs and the average age of sons in the sample.[33] Because Parman measured both fathers' and sons' earnings in the 1915 census, his sample was necessarily made up of fathers later in the life cycle and sons early in theirs. With the recent availability of the 1940 Federal Census, I am able to draw my sample from a set of fathers and sons with average ages of 40 and 35.

[31] These estimates are similar but much lower than the income IGE estimates I will present later in this paper.

[32] Full access to the 1940 Federal Census, including all citizens' names, was not available until April 2012, well after Parman completed his research.

[33] Both of these best fit lines are statistically significant in the univariate regression, but the relationship between father's age and estimated IGE is much stronger. The points graphed in Figure 2b suggest instead that with sons ages ranging from approximately 30 to 35, the estimated IGE should not be a function of the data sample.

[Figure 2 about here.]

The second data-driven limitation is that the Parman (2011) sample is restricted to fathers and sons both living as adults in Iowa in 1915 because the 1940 Federal Census was not available until 2012. How large is the bias of this restriction, and in what direction does it push the intergenerational income mobility results? In the results section that follows, I will be able to use my sample to better understand the magnitude and direction of the problem by limiting my sample to only those sons who still live in Iowa in 1940. To preview, I find that the bias is large and negative: selecting the sample using only sons remaining in Iowa in adulthood reduces both the measured IGE and rank-rank parameters.[34] Because the state of residence in adulthood is, in part, jointly determined with the outcome of interest (earnings), controlling for it or splitting the sample based on it is problematic. In order to estimate an accurate IGE parameter for the early twentieth century, one needs to be able to observe sons that remain in their father's state of residence and sons that move elsewhere.[35]

Thus, while the results in Parman (2011) suggest that income IGE was very low and that income mobility was very high in Iowa in 1915, data constraints complicate the comparison of the estimated IGE to other time periods and places.

2.2 Intergenerational Education Mobility

Due to data constraints, there has been little work on educational mobility in the US historically.[36] Hertz et al. (2007) present the most comprehensive measures of intergenerational elasticity of education across many different countries and regions. They find an IGE of education for the US of 0.46, suggesting more mobility of education in the US than in South America (0.60) but less than in Western Europe (0.40).[37] With data collected on years of education in 1915 in Iowa and in 1940 nationally, I am able to construct the earliest educational mobility estimates for the United States.

[34] These findings are in the final rows of both panels A and B in Table 4.

[35] My sample is technically restricted to those sons still living in the US and enumerated in the Federal Census. However, the number of sons moving abroad in this period is likely very low and thus the bias is likely to be insignificant.

[36] Parman (2011) measures the effects of public education on income mobility, but does not estimate father to son educational mobility directly. Outside of the US, Checchi et al. (2013) study Italian cohorts born between 1910 and 1970 and find very high IGEs and low mobility of education in their early samples, relative to the recent period.

[37] The use of the IGE term and elasticity more generally is a bit of an abuse of notation. The IGE literature on education estimates these parameters using levels on levels, rather than log on log.

In addition, because education is significantly less noisy as a proxy for status than a single year of earnings, the educational mobility results can serve as a check on mobility based on earnings, which may be biased down due to noise.[38]

2.3 Intergenerational Occupational Mobility

The study of historical intergenerational mobility has focused on the study of occupational mobility because occupational data has historically been much more available. Early work on this topic was undertaken by Thernstrom (1964, 1973), studying the occupations of successive generations in Boston and Newburyport, MA. Thernstrom tends to find quite high upward mobility, but a lot of white collar stability as well. Duncan (1965) finds more upward and less downward mobility in 1962 relative to the occupational transition matrices of 1952, 1942, or 1932, relying on data gathered from Occupational Changes in a Generation (OCG). However, neither Duncan nor any of the subsequent work based entirely on the OCG data is able to measure occupational mobility for earlier periods.

Guest et al. (1989) compare a nineteenth-century sample, built by matching fathers and sons in the 1880 to 1900 censuses, to the OCG. They find less upward mobility and more occupational inheritance in the nineteenth century. However, for fathers and sons who are not farmers, the association is both economically smaller and statistically weaker. The results depend a great deal on where Guest et al. put farmers in the occupational distribution.

To avoid the fraught issue of how to rank occupations—especially without available average income, education, or wealth data by occupation—the economics literature has turned to occupational transition matrices, which are agnostic about movements up or down the occupational ladder and instead focus only on movements by the son out of the father's occupational category. In particular, Altham and Ferrie (2007) present the Altham statistic, which has become the standard measure of intergenerational occupational mobility in economics. To compute these measures of occupational mobility, fathers and sons are each grouped by occupation into one of four broad categories—farmer, white collar, skilled and semi-skilled labour, and unskilled labour—within an occupation transition matrix. The Altham statistic measures the strength of association between

[38]In my sample, it is very unlikely that any of the fathers or sons continued education beyond when I observe them in their 30s or 40s.

both the rows and columns of a transition matrix and between any two matrices. Altham statistics can be defined for any two matrices. Specifically, let both P and Q be $r \times s$ matrices with elements p_{ij} and q_{ij}. Then the Altham statistic is:

$$d(P,Q) = \left[\sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{l=1}^{r} \sum_{m=1}^{s} \left| \log \left(\frac{p_{ij} p_{lm} q_{im} q_{lj}}{p_{im} p_{lj} q_{ij} q_{lm}} \right) \right|^2 \right]^{1/2} \tag{1}$$

Altham and Ferrie (2007) use the $d(P,Q)$ notation to convey the sense in which the Altham statistics are distance measures.[39] $d(P,I)$, where I is the occupation transition matrix of perfect mobility (that is, a matrix with ones in all rows and columns), can be used as a measure of distance from independence.

One of the strongest criticisms of using occupations to study long-term trends in intergenerational mobility is the difficulty in classifying farmers (Xie and Killewald, 2013).[40] Comparison of mobility measures across time is complicated—perhaps even driven—by the movement out of agriculture in the US. For example, Guest et al. (1989) conclude that there was more social mobility in the post-WWII period in the US than there had been in the nineteenth century, but they suggest that this reflects the high-heritability of farming and the declining shares of farmers since the late nineteenth century. In this paper, I attempt to control for that by comparing relatively homogeneous samples over time, particularly by constructing a sample in the PSID or other contemporary data that is as rural, white, and agricultural as my Iowa sample. I also focus more analysis on the urban Iowa sons, almost none of whom had farmers for fathers or became farmers themselves.

The second problem posed by farmers is their extreme distribution of earnings. In the standard census sources, including both the 1915 Iowa State Census and the 1940 Federal Census, an individual classified as a farmer may be a small-scale tenant farmer, renting his land and equipment and working a small plot. However, owners of very large farms are also classified simply as farmers. It is quite possible for a father and son who are both farmers to have very different incomes. Similarly, a shift between a father and son from farming to another occupational category may represent an increase or a decrease in income. Further, to what extent is the wide variation in

[39] As a distance measure, Altham statistics satisfy the triangle inequality. For any three $r \times s$ matrices A, B, C, it is true that $d(A,C) \leq d(A,B) + d(B,C)$.

[40] Income measures for farmers, when available, are not a panacea either. To the extent farmers are engaged in subsistence farming, income will be a poor measure between generations.

income among farmers driven by measurement error or transitory income shocks (annual weather shocks, for example)?

When considering an intergenerational sample drawn from a population with a large share of farmers, measures of income mobility and occupational mobility may diverge. In this paper, I measure both income and occupational mobility, as well as educational mobility, for such a population.

2.4 Name-Based Intergenerational Mobility

Both given names and surnames contain socio-economic data. Recent research has utilised this to estimate intergenerational mobility, either as a complement to exact parent-to-child links or a substitute when such links are not possible.

Olivetti and Paserman (2015) create pseudo-links using the content of first names. If families of different means or status have different naming patterns, then even after a child leaves his or her parent's household, the child's name carries with it a proxy for his or her parent's outcomes. For any given name in the population in one census, they calculate the average occupational status (using occupation scores, as described above) of fathers with children of that name. Then, in a later census containing the adult selves of the children in the previous step, they relate the outcomes of the second generation to the implied status of their (unobserved) parents based on their names. Olivetti and Paserman (2015) document decreases from 1870 to 1940 in mobility, consistent with the trends in Long and Ferrie (2007). Because I am working with historical data that includes names of both generations, I am able to implement these name-based mobility measures as a complement to estimates based on my linked sample.

Surnames also convey information about the previous generation. In recent research, Greg Clark and coauthors use rare last names to measure mobility, both between parents and children (Clark and Cummins, 2015), as well as in the very long run (Clark, 2014). Guell et al. (2015) formalise such a method of measuring mobility based on surnames and show declining mobility in Catalonia over the century. One major advance with the rare surname method is its flexibility: whether based on admissions lists from elite academic institutions or probate records or voting records, any data containing both last names and a measure of status can be used to estimate mobility. Clark (2014) argues that these measures tend to suggest far less mobility than one generation income or occupational status mobility measures. Studying mobility based on the over- or under-representation among doctors or lawyers of historically prestigious surnames in the United States, Clark et al. (2012) find advantages persisting for at least 5 generations. While my two generation sample does not permit such long run estimates of mobility, I am able to apply Clark's method to the sons of Iowa to estimate the changing shares of elite surnames in certain occupations, focusing on doctors and lawyers as in Clark et al. (2012).[41]

3 Intergenerational Mobility Estimates for Iowa from 1915 to 1940

How much intergenerational mobility was there in Iowa in the early twentieth century? How do the measures of mobility reviewed in the previous section compare? When data constrains researchers from studying one or more of these measures, are the other measures plausible substitutes? In this section, I estimate the intergenerational mobility of income, earnings rank, education, and occupation, as well as mobility measures based on the socio-economic content of names. I document that these measures are all in general concordance with one another in my sample in two ways. First, all measures suggest that early twentieth-century Iowa was a place of high intergenerational mobility. Second, the relative rates of mobility between urban and rural sons, as well as between sons with native-born versus foreign grandparents, are all roughly consistent across measures. I also argue that, based on a variety of measures, there was likely more mobility in Iowa historically than there is today.

I begin by presenting raw intergenerational correlations of fathers' and sons' outcomes in Figure 3: log annual income (3a), years of education (3b), and occupation score based on 1950 scores (3c) and on 1915 scores (3d).[42] The intergenerational correlation is one possible measure of mobility and holds trends in the marginal distributions—the dispersion of earnings among fathers and among sons—fixed. In log earnings, the correlation is 0.14; the education correlation is a bit higher (0.24).[43] Clearly, there is a strong positive relationship between outcomes for fathers in 1915 and

[41] Guell et al. (2015) develop an alternative metric for deriving the informational content of surnames.

[42] Naturally, there are many father-son pairs with the same outcome levels as other pairs. In an attempt to display this density at certain points on the graphs, I have used both hollow scatterplot markers and jittered the data. The best fit lines are, of course, drawn based on the full sample before jittering. The $5000 top coding in the 1940 census is apparent on the right side of Figure 3a.

[43] The occupation-based measures differ, depending on the year used to construct occupation scores. The 0.14 correlation coefficient based on 1915 scores suggests a lot of mobility, while the 0.36 estimate based on 1950 scores does not.

sons in 1940 but the exact measure of the respective slopes of these lines—and how those slopes compare across the various measures of mobility, as well as with the estimates of mobility in the recent period—is the key question of this paper.

[Figure 3 about here.]

3.1 Intergenerational Mobility of Earnings

I measure mobility of earnings in two primary ways. First, following the intergenerational mobility literature, I use intergenerational elasticities (IGE) (Corak, 2006; Solon, 1999; Black and Devereux, 2011). The canonical formulation regresses the son's adult outcome, in my case as measured in the 1940 Census, on the father's adult outcome, as measured in the 1915 Census. Let Y_i be the outcome of interest, either (log) income or education. The model I estimate can be summarised as:

$$Y^s_{i,1940} = \alpha + \beta \cdot Y^f_{i,1915} + Q^s(age^s_{i,1915}) + Q^f(age^f_{i,1915}) + Q^s(age^s_{i,1915}) \times Y^f_{i,1915} + \epsilon_i \tag{2}$$

β can be thought of as a persistence parameter: larger estimates mean a tighter link between father and son and thus less mobility.

Second, following Dahl and DeLeire (2008) and Chetty et al. (2014b,a), I also use rank-rank estimates. Again, I regress the son's outcomes on the father's outcomes, but where outcomes are the relative positions or percentiles in the income distribution. For sons observed in 1940 I use the full 1940 IPUMS census sample to calculate the full income distribution of white men, aged 28-42, matching the demographics of my sample. For fathers, income data are not available for a nationally representative sample. I instead calculate the full income distribution of white men in the Goldin-Katz Iowa 1915 census sample with the same age range as the fathers in my sample. Ranks are scaled as percentiles between 0 and 1; a rank of 0.5 indicates that the father or son is at the median for annual income.

To reduce any measurement error induced by life cycle income effects (Grawe, 2006), I follow Lee and Solon (2009) and include quartic age controls for both the father and the son, defined as Q^s and Q^f above, as well as an interaction between the son's age and the father's outcome. In the interaction term, I normalise son's age in 1940 relative to age 40 (Haider and Solon, 2006).[44]

[44] With this normalisation, the estimated β represents the relationship between son's and father's outcomes when

The fact that I define my sample to observe sons between ages 28 and 42 in the 1940 Census also reduces life cycle driven measurement error. As some of my observed sons are brothers, I cluster standard errors at the family level. I also include an Iowa 1915 county fixed effect, subsuming an urban or rural control. The results are robust to the inclusion of controls for family-size effects, county fixed effects, and the name string control variables described previously.[45]

Panel A of Table 4 presents my estimates of the IGE of income across a variety of samples. Both the father's and son's incomes are measured as annual log earnings.[46] The first specification is a simple univariate regression of the son's log earnings on the father's log earnings. In specification two, I include controls for name string properties that might affect matching, 1915 county of residence fixed effects, and quartic controls in father and son age. In the third specification, I also include an interaction between son's normalised age and father's log earnings to control for life cycle measurement error (Grawe, 2006; Haider and Solon, 2006).

My baseline estimates for the IGE parameter for the full sample of Iowa fathers and sons range from 0.199 to 0.258, as shown in the first row of Table 4. The literature suggests an IGE of 0.47 for income in the United States today (Corak, 2006). Lee and Solon (2009) argue that the IGE of income has been roughly stable for cohorts observed between the late 1970s and the early 2000s. My results suggest that this recent stability does not extend historically, and that there was much more intergenerational mobility of income in the early twentieth-century US than there is today.[47]

---

the son is age 40. I follow Lee and Solon (2009) in normalising to 40.

[45]The county fixed effects indicate the county of residence when the son is observed in Iowa in 1915. The name string controls include first and last name commonness, length, letter similarity, and Scrabble scores, all attempts to control for differential matching rates between the 1915 and 1940 censuses. While I include these various controls to reduce measurement error, both Chetty et al. (2014a) and Nybom and Stuhler (2014a) present extensive results that suggest the rank-rank measures of intergenerational mobility are much less susceptible to biases. Working with the universe of US tax records, Chetty et al. (2014a) argue that estimates are stable even with just one year of income observed for both fathers and sons, though this is still an unresolved issue in the literature. Further, they document that the exact age when fathers or sons are observed has very little effect on the measurement of mobility, so long as the fathers are observed between the ages of 30 and 55 and the sons are observed after age 30. Nybom and Stuhler (2014a) replicate these lessons for the estimation of rank-rank mobility using Swedish data. The stability of my estimates of rank-rank mobility with and without various controls suggests that the rank-rank parameter is quite robust in my historical sample as well.

[46]To ensure comparability with contemporary estimates, I use annual earnings, not weekly earnings. Results using weekly earnings are similar and in fact lower than those presented in Table 4, suggesting even more mobility in the early twentieth century.

[47]One concern with the results presented thus far is the reliance on the log transformations of the income data. By logging income, the assumption made is that small changes in income for very poor fathers have much higher returns (to the son's income) than smaller changes further up the income distribution. In the appendix, I show that these results are robust to alternative transformations of the father's and son's income variables, including both levels and square roots.

[Table 4 about here.]

Similar to my IGE results, I find much more income mobility historically than today. The rank-rank parameter ranges from 0.167 to 0.217 in the first row of Panel B of Table 4, the main sample with all linked father-son pairs. Chetty et al. (2014a) measure a rank-rank parameter of 0.341; among just male children, they find a rank-rank estimate between 0.307 and 0.317.

However, any measurement error will tend to bias down estimates of intergenerational mobility (Solon, 1999). Further, though Iowa is in some ways broadly representative of the US in 1915—with respect to urban and rural shares—the differences in my estimated mobility may reflect differences between Iowa and the rest of the country, not differences between time periods. In fact, according to contemporary data, children born in Iowa are among the most economically mobile in the entire country, across many measures (Chetty et al., 2013).

I attempt to standardise my comparisons in two ways. First, I construct a sample of recent intergenerational data that is demographically comparable to my Iowa sample, drawing on data from the PSID. To do this, I limit the PSID to include only white father-sons pairs (99% of my linked Iowa sample is white). I also limit the PSID to sons who grew up in the Midwest.[48] The results are presented in the second rows of each panel in Table 4 (Panel A for IGE, Panel B for rank-rank).

To calculate a comparable contemporary IGE, rather than follow Lee and Solon (2009) and measure the father's income as the average of his income when the matched son is between 15 and 17 years old, I use the father's income when his son is 10.[49] In doing so, I attempt to replicate the noise in my historical data from only observing income once. The son's income is observed in each year that the son is in the PSID and is between the ages of 28 and 42, to match my 1940 census data. Both income variables are measured in 2000$.[50] Limiting the PSID to sons born in the Midwest, I estimate an IGE between 0.33 and 0.50, depending on the use of state fixed effects

[48] The Midwest region is defined in the PSID as Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, and South Dakota. I do not limit the PSID sample just to sons raised in Iowa as there are only 385 father-son pairs with the requisite data.

[49] Mazumder (2009) underlines the downward bias on mobility estimates when using only a single-year of earnings data. I cannot create additional years in my historical data, but I can ensure the contemporary comparison is afflicted by the same potential single-year bias as my Iowa data. I use 10 because this is the midpoint of my age range for sons in the 1915 sample. If I do not observe a father in the year when his son is 10, I use the year when the son is closest to 10 in the PSID sample.

[50] While I attempt to match my age and county fixed effects from my Iowa sample results with age quartics and 'grew up' fixed effects, I do not observe either family size or name strings in the PSID.

and age controls.[51] Though the Iowa-like samples in the PSID are small and contain repeated observations of the same set of father-son pairs over many years, the results suggest that the lower mobility I find historically is driven neither by the demographic composition of my data nor by the single-year measurements of income. I come to a similar conclusion when comparing my estimates of rank-rank mobility with contemporary measures. The first comparison, shown in the second row of Panel B of Table 4, is the rank-rank mobility using the PSID sample. For the Midwest sample, I measure a lower parameter than is found nationally, suggesting the Midwest is more mobile; however, these estimates are far larger than what I find historically in the full sample.

As a second alternative construction of comparable recent mobility parameters, I use the county level results reported by Chetty et al. (2014a). Unfortunately, such a comparison can only be made for the rank-rank parameter, as Chetty et al. (2014a) do not calculate IGE at any disaggregated geographies.[52] When I calculate the weighted average of rank-rank mobility, weighing by the shares of my sample living in each county in 1915, I find a rank-rank parameter of 0.303, similar to the result from the contemporary PSID data and, more importantly, far larger than the rank-rank parameter of 0.169 to 0.219 that I find in my historical sample.[53]

In the appendix, I show that neither false matches in my census linking procedure nor higher levels of measurement error in historical data could account for my estimates of lower IGE parameters (and thus higher mobility) in the 1915 to 1940 sample relative in the contemporary sample. Via simulation, I introduce both mismatches and measurement error into my Iowa-like PSID sample considered above. The share of false matches would have to approach 50% for mismatching to account for the estimated differences in IGE parameters, which seems highly unlikely.[54] As detailed in the data section, the matches were carefully constructed based on first and last names, year of birth, state of birth, and gender. In addition, the measurement error simulations suggest that earnings measures from the 1915 and 1940 censuses would have to be considerably noisier than

[51]Specification 3 includes interactions with father's outcomes and son's ages. However, the PSID includes a small number of father-son pairs observed repeatedly which may complicate the interpretation of this estimate. Specification 1 and 2 are preferred and more conservative as comparisons.

[52]Hertz (2009) does calculate mobility rates by race; given the very high white share in Iowa in 1915, his estimates of 0.39 or 0.44, depending on the adult equivalent adjustments in each specification, are similarly indicative of reduced mobility in the present period relative to 1915 Iowa.

[53]I can also split the sample between the urban and rural counties in my analysis. The weighted average of rank-rank mobility is 0.355 in the three urban counties and 0.268 in the 10 rural counties.

[54]That is, 50% of sons that I find in the 1940 census and link back to 1915 on the basis of the son's first and last names, state of birth, and year of birth would have to be the wrong person. For the rank-rank parameter, the mismatch error required to shrink the difference between the estimates is roughly 30%.

earnings measured in the recent period to generate the large difference in IGEs.

Are the high rates of mobility that I find driven by the movement off the farm in the early twentieth century?[55] To answer this question, I compare differential mobility for both sons of rural and urban Iowa, splitting the sample according to where the sons were living when their fathers were first sampled in the 1915 Iowa State Census.[56] The results for these subsample analyses are presented in the third and fourth rows of Table 4. Only 18 of the urban sons have a father farmer in 1915 and only 60 are farmers in 1940. Sons observed in rural Iowa in 1915 are more mobile than their urban peers as measured both by IGE and rank-rank parameters, though the differences are not statistically significant for the rank-rank mobility estimate. All measures still show more mobility historically than is estimated in today. The much higher levels of mobility for rural sons may be driven by the large increases in access to public education even in remote, rural regions of Iowa (Parman, 2011). Alternatively, the high levels of mobility may be caused in part by movement off the farm; this finding is consistent with the model of human capital transmission presented by Nybom and Stuhler (2014b), which suggests that periods of structural transformation in the economy weaken the links between parents' and children's outcomes.

Further isolating the effects on mobility of the shift away from agriculture, I limit the samples in the fifth rows of both panels A and B to only sons with fathers who were not farmers.[57] Again, mobility is lower than the contemporary estimates, though much closer to the urban sample than the rural sample. Overall, these urban and non-farmer-father subsamples suggest that the lower levels of mobility found historically are not artifacts of poor measurement of farmer income, whether that mismeasurement is driven by classical measurement error, by the difficulty of farmers to distinguish between net and gross income in census responses, or by transitory income shocks (such as adverse weather or crop-destroying pests).

I also find that the grandchildren on the foreign-born have more mobility, as shown in the sixth and seventh rows in Table 4. Drawing on data on each son's grandparent's place of birth in the 1915 census, I partition the sample into sons with four native-born grandparents and with four

[55] Or are the results driven by the difficulty of accurately measuring income for farmers?

[56] As presented in Figure 1, the rural counties included in the Goldin-Katz sample are Adair, Buchanan, Carroll, Clay, Johnson, Lyon, Marshall, Mitchell, Montgomery, and Wayne and the urban cities are Davenport, Des Moines, and Dubuque.

[57] This sample is made up of the urban sample and nearly half of the rural sample.

foreign-born grandparents.[58] This corresponds to the high rate of upward mobility Perez (2016b) finds in Argentina in the nineteenth century.

Could peculiarities about the 1940 census be driving the low persistence parameters I estimate? Based on the eighth and ninth rows of Table 4, I argue that neither the lack of capital income data in 1940, nor the census enumeration in the shadow of the Great Depression and on the eve of World War II explain my findings.[59] Missing capital income, I am forced to exclude the 13.7% of the sons in 1940 who were farm owners or operators without income or business owners in the Table 4 measures of mobility previously discussed. As Bjorklund et al. (2012) highlight in Sweden, rates of persistence are extremely high at the top of the distribution and this is likely driven by wealth. Excluding the earnings of such sons could lead me to find a lower persistence parameter and more mobility. However, in row 8, I impute earnings for farmers using the 1950 census, which did collect data on capital income and non-wage and salary earnings. Earnings are imputed using years of education, age, state of residence, and state of birth. Using these imputed earnings, I estimate even more mobility than in my main results.[60] In row 9, I expand this imputation to the entire sample; the wage distribution may have still reflected the Depression in 1940, but such shocks may have dissipated by 1950.[61] While I am unable to link the sons in my sample ahead to the 1950 census—because access to the names in the 1950 census is restricted until 2022—I can use it to assess whether something about the 1940 census is driving my results. Rather than only impute capital income for a portion of the sons in my sample, as in row 8, in row 9 I impute total '1950' earnings for every son.[62] This exercise suffers slightly with a difficult imputation as the variables I

[58] Each of these groups makes up roughly one-third of the sample. The other third of sons had 1, 2, or 3 foreign grandparents.

[59] As noted previously, only wage and salary earnings were recorded in the census. The data collected is the 'total amount of money wages or salary', and enumerators were instructed: 'Do not include the earning of businessmen, farmers, or professional persons derived from business profits, sale of corps, or fees'. For more, see `https://usa.ipums.org/usa/voliii/inst1940.shtml#584`.

[60] To impute total earnings, I regress the log of total income in 1950 on a full set of years of education indicators, a quartic in age, state of residence fixed effects, and state of birth fixed effects from data on farmers in the IPUMS 1% sample of the 1950 census. I use the results from that regression to predict income for the farmers in 1940, normalising income from 1950$ to 1940$. For more details on the imputation of capital income in 1940 and graphs of the relationships between income and education and age for farmers in 1950, see Appendix A.3.

[61] Of course, the enumeration of the 1950 census falling during the Great Compression (Goldin and Margo, 1992) and after a world war may complicate the distribution of earnings then as well.

[62] As in the previous exercise, I use the 1950 1% IPUMS sample. I regress total earnings in 1950 on a full set of education indicators, a quartic in age, state of residence fixed effects, and state of birth fixed effects. To improve the precision of my imputation, I add occupation and industry indicators, as well as urban and rural status fixed effects (variables that did not vary when I was only considering farmers). Finally, I also interact state of residence with an occupation score variable, education level, and age, allowing the effects of these covariates to vary by state.

observe in the 1950 1% sample explain just under half of the variation in log total earnings in 1950. Nevertheless, I estimate a very low rate of persistence in both panel A and panel B.

Overall, my estimates suggest more mobility historically than today, measured both by IGE and with the rank-rank parameter.[63] Further, it appears that in the early twentieth century, the IGE and rank-rank measures of mobility generally agree, both in comparison to contemporary data and comparing across the urban and rural subsamples and sons with native or foreign grandparents in my data.

What is the probability that a son born in a given quintile will be in the same or another quintile in 1940? Panel A of Table 5, an income quintile transition matrix, can be used to answer such questions. Each cell represents the probability that a son with a father in a given income quintile (identified by the column) will be in a given income quintile in 1940 (given by the row). A son whose father is in the bottom quintile in 1915 has only a 13.8% chance of being in the top quintile in 1940, while the odds that a son born in the top quintile remains there in 1940 are 36.2%. What is the likelihood a son falls into the bottom quintile? Not surprisingly, those odds fall with father's rank: a son born in the top quintile has only a 7.2% chance of being in the bottom quintile in 1940, while the odds are more than twice as high (15.0%) that a son born in the bottom quintile remains there. While the table clearly presents a degree of intergenerational immobility, when these percentages are compared to a transition matrix for the recent period (sons born in the 1980-82 cohorts) from Chetty et al. (2014a), there is in fact more mobility between 1915 and 1940. In Panel B of Table 5, I calculate the quintile transition matrix for Iowa today, weighting the commuting zone level results from Chetty et al. (2014a) by the share of 1915 sons in each county. Sons born in the bottom quintile are ten points more likely to remain there as adults in the contemporary data than historically: 26.1% to 15.0%. Historically, sons born in the second quintile have a 70% chance of being in a higher quintile in adulthood; that probability is only 65% in contemporary data. At the top end, however, transition probabilities are similar between the two periods: sons born in the fourth or fifth quintiles have 24.0% and 36.2% probability of being in those quintiles as adults in

[63] I do not find as little persistence in status as in Parman (2011), which measured mobility by linking fathers and sons in Iowa in 1915. Because the 1940 Federal Census was not available until 2012, Parman was forced to restrict his sample to fathers and sons both living as adults in Iowa in 1915. Is this a bias? In the final rows of both panels A and B in Table 4, I calculate IGE and rank-rank parameters for this subset of sons still living in Iowa in 1940. The results suggest that the bias is both large and negative: selecting the sample using only sons remaining in Iowa in adulthood reduces both the measured IGE and rank-rank parameters.

the Iowa sample, compared with 27.8% and 41.3% probabilities in the Iowa 1915 weighted Chetty et al. (2014a) sample.

[Table 5 about here.]

3.2 Alternative Measures of Intergenerational Mobility

The richness of my historical linked sample enables me to estimate several measures of intergenerational mobility beyond earnings. In this subsection I show that my findings based on the IGE and rank-rank parameters are generally in line with estimates of mobility based on education, occupations—both occupation scores and occupational status transition tables—and names.

In addition to serving as a (potentially) more accurately measured check on my income results, the education persistence estimates are a valuable and important historical parameter. I estimate the intergenerational mobility of education and present these results in Panel A of Table 6.[64] My fathers and sons are both observed at a pivotal moment of change in public education. The growth of mass public schooling in the United States, first in common schools during the later half of the nineteenth century and then through the high school movement in the early twentieth century, made education widely available and free (Goldin and Katz, 2008, 2011). Goldin and Katz (2008) also argue that this increase in human capital helped spur national growth and prosperity in the following century. Whether this massive public investment in education also reduced the link between a son's and his father's educational outcomes can help scholars understand the role of public programmes in shaping or changing inequality.

The literature suggests an IGE of education of 0.46 (Hertz et al., 2007). As Table 6 shows, I find a much lower IGE parameter for schooling, between 0.206 and 0.264. This suggests that, like mobility measures based on income, educational mobility in the US was higher in the early twentieth century than it is today.[65] Similar to the results presented on income mobility, there is more mobility of education among rural sons than urban sons as well, though these differences are not always statistically significant. Finally, the last two rows of Panel A in Table 6 are also in

[64] The education persistence regressions are estimated in levels rather than logs and so they are not true intergenerational elasticities.

[65] Concerns about noise driving down the estimated IGE parameters are less important for education, as any given annual measurement of years of schooling completed (for an adult) is a very accurate measure of lifetime years of education completed.

agreement with the earnings mobility results: the persistence of educational attainment is much lower among sons with foreign grandparents than sons with native grandparents.

[Table 6 about here.]

I also estimate intergenerational mobility using occupation scores.[66] These scores measure the median earnings in a given occupation and may contain less measurement error than annual income observations. The occupation scores are not a panacea, even with income or earnings often unavailable in historical data. The occupation score commonly used is calculated by IPUMS from a 1950 census report. However, occupations in 1950 are difficult to link to occupations in earlier years, given the changing nature of tasks within an occupation and development or death of other occupations. Further, given the large changes in the returns to human capital and specific skills throughout the last two centuries, the median earnings for even the same exact occupation in two periods may be poorly correlated (Goldin and Katz, 2008). Nevertheless, given the widespread use of these measures in the historical literature on both intergenerational mobility and more broadly as a substitute or proxy for income, I can replicate my analysis with the occupation scores. I do so using both the standard 1950-based occupation scores from IPUMS as well as a 1915-based score that I construct from the full Iowa sample.[67]

When measured with occupation score, mobility was quite high in the early twentieth century, corroborating my findings with income and education.[68] I present the occupation mobility results in Panels B and C of Table 6. I estimate an IGE parameter between 0.184 and 0.265 for the occupational score measure based on income data from 1915 Iowa. This is higher than my IGE estimate for income, but lower than my IGE estimate for education. However, when I use the 1950-based occupation score measure, I find significantly larger IGE estimates, indicating less mobility. These measures are higher than any previous IGEs estimated in this paper, suggesting only slightly

[66] For research on economic outcomes in periods without earnings data, many economists have turned to such occupation score measures, including Olivetti and Paserman (2015) and Abramitzky et al. (2012, 2013). I estimate mobility rates based on occupation scores during the Great Depression and find that, like earnings-based measures, occupational mobility was reduced for sons growing up in cities with more severe Depression downturns (Feigenbaum, 2016b).

[67] In the appendix, I detail the construction of these measures and compare them. They are highly correlated; however, some occupation groups are clear outliers, suggesting that they moved up or down the income scale between generations.

[68] Occupation score is not available in the contemporary PSID sample and so I cannot compare these estimates to a parallel recent estimate. Further, given the availability of income data in the recent period, I am not aware of any studies that attempt to estimate intergenerational mobility of occupation score.

more mobility in the early twentieth century than today, if any. These results are driven by farming fathers: when I subset the analysis to the urban sons or exclude sons of farmers, the results are more consistent across both measures of occupation score. There are a large number of farmers in my sample, and the relative positions of farmers and their median incomes changed quite a lot between 1915 and 1950. Based on the incomes reported in the 1915 Iowa State Census, farmers were at the median of the occupation distribution in 1915 (in Iowa), but in 1950, farmers ranked around the bottom 10th percentile of occupation groups by median income (nationally). This instability reflects the complication of using a measure like occupation score to determine intergenerational mobility, especially at a time of large structural change in the economy. It also reflects the difficulty of crudely classifying all famers with the same simple median income score, given the huge income differences in reality between wealthy owners of large farms and poorer tenant farmers.

Finally, I turn to names as a source of data on intergenerational mobility, following the recent work by Olivetti and Paserman (2015), Clark and Cummins (2015) Clark (2014), Clark et al. (2012), and Guell et al. (2015).

Using a son's first name to proxy for his father's economic status, I find high rates of mobility, as presented in Table 7. To create these estimates of mobility, I start with the Olivetti and Paserman (2015) data, drawing the average occupation score among fathers in 1910 for children with a given name. I then assign predicted father's occupation scores (based on the sons' first names) to the sons I locate in 1940, creating a pseudo-link. The variation in these results across specification is more dramatic than in the previous tables, perhaps underscoring the noise in measuring status using the first names and occupation scores. However, specification 1 most closely resembles the specification in Olivetti and Paserman (2015), and I focus on that. These results are remarkably consistent with the results reported in Olivetti and Paserman (2015) when they restrict their sample to the sons of the Midwest. In their Table 8, Olivetti and Paserman (2015) estimate a mobility from 1910 to 1930 of 0.27; from 1920 to 1940 the mobility is 0.35. These results bound my findings for my full sample, mobility of 0.300 with log earnings as outcome or 0.353 with log 1950 occupation score as the outcome.[69]

[Table 7 about here.]

[69] Olivetti and Paserman (2015) note this concordance as well in footnote 29, page 2719.

The results in the subsamples are more ambiguous. When 1940 outcomes for the sons are measured with log earnings, I do find more mobility among rural sons than urban sons, as before. However, using occupation scores as the son's outcome as in Panel B suggests no statistically significant difference in mobility between rural and urban sons, at least in specifications 1 and 2. Similarly, sons with foreign grandparents are slightly more mobile when outcomes are measured in log earnings, but substantially less mobile according to occupation score outcomes; the statistical precision is also much weaker for many of these coefficients. These differences are difficult to parse, but may suggest that the first-name imputations of the fathers' status is complicated to interpret among families a generation or two removed from immigration.

I find similarly high rates of mobility when using surname-based measures, following the work Greg Clark and others.[70] Rare surnames provide another form of pseudo-linking between generations, because men with a very rare surname in one generation are likely to be the fathers of men with that same rare surname in the next generation.

I start by locating rare, elite surnames in Iowa in the complete count 1920 Federal Census.[71] For surnames held by 10 to 30 adult men, I find the names overrepresented among doctors and lawyers, the two high status groups studied by Clark in the US. Following Clark, I define the relative representation of a surname k among doctors as

$$RR_k^{doctors} = \frac{doctors_k \div doctors_{total}}{population_k \div population_{total}} \tag{3}$$

all measured in 1920. The relative representation of lawyers is defined analogously. Among the set of 115 surnames with an $RR_k^{doctors}$ greater than 10, how often does this name feature among doctors in 1940?[72] The relative representation among doctors is 2.76, clearly still overrepresented, and among the 140 surnames with $RR_k^{lawyers}$ greater than 10, the relative representation in 1940 among lawyers is 2.99.

[70] My application of the Clark method diverges in slightly because I am unable to identify high status Iowa names *before* my sample and instead use occupational scores and high status occupations in 1920.

[71] I use men in Iowa in 1920 rather than the men in the Iowa census of 1915 because the Goldin-Katz sample of Iowa 1915 is only a 2.6% sample of the state. Identifying rare names in the complete census is much more accurate than identifying them from a small sample of the census, but such a complete count of Iowa in 1915 has not been digitised.

[72] Names include 'Alt', 'Appel', 'Willett', and 'Woodhouse'. None of the names in the list are especially noteworthy. This contrasts to the elite names Clark et al draw from the high income tax payer lists in 1923 and 1924 which includes Vanderbilt, Roosevelt, and Winthrop, but accords with the names Clark et al draw from the Ivy League graduate database from 1650 to 1850 (Clark p. 4: 'only a very few have any resonance').

How does this persistence of high occupational status compare to the findings from Clark in the US? Ashkenazi Jewish names had a similarly high prevalence in the first generation from 1920 to 1949 among doctors (between 6.73 to 13.64 depending on assumptions about their share of the population), but still had relative representations of more than 5 three generations later (from 1980 to 2009). Two other elite groups studied by Clark et al were the 1920s Rich (high earners whose tax bills were listed in the 1923 and 1924 newspapers) and Ivy League graduates form 1650 to 1850. Neither are as prevalent among doctors in the early twentieth century as the Ashkenazi Jews—relative representation ranging form 2.97 to 5.86—and both regressed to rates in the same range as the sons of elite Iowa a generation later. My results suggest that there was likely more mobility in Iowa than among the samples Clark examines in the US, but an exact comparison is complicated by the differences in sources.

3.3 Mobility According to Occupation Transitions

Similar to income, educational, and occupation score mobility, there appears to be more broad-category occupational mobility during my period of study than there is today. I measure occupational mobility using Altham statistics (Altham and Ferrie, 2007). My occupational transition table from 1915 to 1940 is presented in Table 8. Occupations are categorised by linking occupational strings (exactly as entered by the census enumerators) to the 1940 occupational code charts for both the 1915 father and 1940 son samples.[73] Sample sizes are different from previous portions of the analysis because not all occupation strings could be matched to occupation groups.

[Table 8 about here.]

The distance between the occupation transition table and the identity table, I, can be thought of as a measure of occupational immobility—the larger the distance, the more likely it is that sons enter the same occupational class as their fathers. Calculating the Altham statistic for Table 8 yields $d(IA1940, I) = 16.14$. Long and Ferrie (2007) report $d(US1880, I) = 12.09$,$d(US1900, I) = 14.58$, and $d(US1973, I) = 20.76$. The Altham statistic generated by my linked sample of fathers and

[73] In the margins of the original 1940 census manuscripts, exact occupation codes are included. Using both these occupation codes (when they are recorded) and the exact occupation strings, I have attempted to carefully match occupations from my data to the 1940 occupation master list. There may be measurement error in the exact matching. However, given that occupations are then collapsed to the four broad categories used in the Altham statistic, errors in occupation matching will bias the final results only if occupations are coded into the wrong broad category.

sons from 1915 to 1940 is larger—thus indicating less mobility—than the measures presented by Long and Ferrie (2007) for the nineteenth century, and statistically significantly different from these historical measures as well. These results are summarised in Figure 4 and in the first column of Table 9. In addition, echoing the results presented previously suggesting more mobility historically than today, there appears to be more mobility between 1915 and 1940 than between 1950 and 1976—as compared to the more contemporary estimates reported by Long and Ferrie.

[Figure 4 about here.]

[Table 9 about here.]

Relative to the full sample, I find more mobility among both the urban sample and the rural sample, as shown in the second and third columns of Table 9.[74] I also find more mobility in the urban sample than in the rural sample.[75] However, none of these differences is statistically significant, and I cannot reject that mobility was the same in Iowa overall as in the urban and rural subsamples.

In this section, I have argued that the various measures of mobility that I am able to calculate in my dataset are all roughly in agreement: the early twentieth century was a period of high mobility, at least for the sons of Iowa. Further, mobility appears to have been higher among the sons of rural Iowa than their urban peers and among the sons of foreign-born grandparents relative to sons with native-born grandparents. The results also suggest that there was more intergenerational mobility—less persistence—in Iowa in the early twentieth century than today across these many measures of mobility.

4 Conclusion

This paper presents estimates of intergenerational mobility for men born in Iowa between 1900 and 1910 across a variety of measures. Mobility based on earnings, education, occupation, and the socio-economic content of names all roughly align, suggesting that this was a period of high mobility

[74] The fact that the Altham statistics of two disjoint sets can each be smaller than the Altham statistic of their union is algebraically allowed, but seems to an undesirable property of Altham statistics. In theory, an alternative statistic might possess a form of continuity and the intermediate value theorem.

[75] This difference is the reverse of what I found with respect to both income and educational mobility in the previous section.

and that there was less persistence in status for rural sons than urban sons and for the grandsons of the foreign-born than grandsons of the native-born. The results also suggest that the mobility rates for this generation were higher than those for men born since 1960. Because many of the challenges of estimating mobility historically such as imperfect census linking and observing only a single year of earnings all serve to reduce the estimated persistence parameter and inflate mobility, any comparison between eras is not conclusive, but my results are another piece of evidence in favor of the theory that mobility was higher in the early twentieth century than today, echoing the findings of Parman (2011), Long and Ferrie (2008), and others.

Both Lee and Solon (2009) and Chetty et al. (2014b) find relative stability in intergenerational mobility over the past two to three decades. If mobility was higher among sons born between 1900 and 1910, then this recent stability could not be a permanent feature of intergenerational mobility in the United States. At what point in the twentieth century did economic mobility decline? Was there a sharp transition from one stable level of mobility to another, or was the shift a gradual decrease in mobility over several decades? Or, as Olivetti and Paserman (2015) document, did decade-to-decade fluctuations in mobility not always align with the trend over the century? Was there variation across the US in this change? And what caused this shift? Did the Great Compression induce a new era of lower mobility (Goldin and Margo, 1992)? Or did the change come later and affect sons born during mid-century and entering the labour force in the 1970s or 1980s? The data to answer this question exists but much of it is not yet accessible. Eventually, with the 2022 release of the full non-anonymous 1950 census—and the 2032 release for the 1960 census—it will be possible to track intergenerational mobility through the middle of the twentieth century.

If there was indeed more mobility in Iowa in the early twentieth century than there is today, an important question is why. In addition to providing the first national micro-records of income and years of education, the 1940 Federal Census was conducted on the heels of two major economic events of the twentieth century: the Great Depression and the Dust Bowl. Feigenbaum (2016b) estimates less intergenerational mobility among sons growing up in cities more severely hit by the Depression. Hornbeck (2012) uses variation in dust bowl severity to measure the effects of environmental catastrophes on economic outcomes, finding that the Dust Bowl led to immediate and persistent reductions in agricultural land values and production and to large population out-migrations. These Dust Bowl disruptions could have either broken links between generations and promoted mobility or disproportionately reduced the prospects of children from poorer families. The high school movement and the huge expansion of access to public education could have also been a driver of mobility.[76] Sons had on average two more years of schooling than their fathers. I find high levels of educational mobility: a son's completed years of schooling are only weakly related to his father's education. The wide availability of free schooling, it appears, could thus sever the link between a father's and son's educational attainment. The general transition away from an agriculture-based economy may have also played a role. I estimate higher levels of mobility among rural sons, many of whom were the sons of farmers. However, the high levels of mobility persist in samples restricted to the sons of non-farmers in both urban and rural Iowa. Variation in the access to transportation or differences in inequality across Iowa could also play an important role.[77]

Unfortunately, the nature of my linked sample from Iowa 1915 prevents me from answering these questions. I observe father and son pairs in only 10 rural counties and 3 cities. Though there is variation in many of these potential drivers of mobility across the state, there are far too many explanations to reliably estimate the effects. Parsing out what drove the mobility rates in Iowa in the early twentieth century—and what explains the variation across and within urban and rural areas—will have to be addressed in future research with data better suited to the task with more cross sectional variation.[78]

[76] Parman (2011) shows how education in Iowa might have affected mobility in the previous generation.

[77] Perez (2016a) documents the mobility effects of infrastructure access in nineteenth century Argentina and the density of railroads varied across Iowa in 1915. The Gatsby Curve and the classic Becker and Tomes (1979) theory of mobility both highlight the relationship between mobility and inequality.

[78] One virtue of comparing relative mobility across an historical sample is that any biases induced by the matching procedure or the sample or the historical data should be differenced out.

References

Aaronson, Daniel and Bhashkar Mazumder. 2008. "Intergenerational Economic Mobility in the United States, 1940 to 2000." *Journal of Human Resources* 43:139–172.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2012. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review* 102:1832–1856.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2013. "Have the poor always been less likely to migrate? Evidence from inheritance practices during the age of mass migration." *Journal of Development Economics* 102:2–14.

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

Altham, P.M.E. and Joseph P Ferrie. 2007. "Comparing Contingency Tables Tools for Analyzing Data from Two Groups Cross-Classified by Two Characteristics." *Historical Methods* 40:3–17.

Becker, Gary S. and Nigel Tomes. 1979. "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility." *Journal of Political Economy* 87:1153–1189.

Biavaschi, Costanza, Corrado Giulietti, and Zahra Siddique. 2013. "The Economic Payoff of Name Americanization." https://www.econstor.eu/dspace/bitstream/10419/89864/1/dp7725.pdf.

Bjorklund, Anders, Jesper Roine, and Daniel Waldenstrom. 2012. "Intergenerational top income mobility in Sweden: Capitalist dynasties in the land of equal opportunity?" *Journal of Public Economics* 96:474–484.

Black, Sandra E. and Paul J. Devereux. 2011. "Recent Developments in Intergenerational Mobility." In *Handbook of Labor Economics*, edited by David Card and Orley Ashenfelter, volume 4, chapter 16, pp. 1487–1541. Elsevier B.V., volume 4b edition.

Checchi, Daniele, Carlo V. Fiorio, and Marco Leonardi. 2013. "Intergenerational persistence of educational attainment in Italy." *Economics Letters* 118:229–232.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2013. "The Equality of Opportunity Project." http://www.equality-of-opportunity.org/.

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

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. 2014b. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." *American Economic Review: Papers & Proceedings* 104:141–147.

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

Clark, Gregory and Neil Cummins. 2015. "Intergenerational wealth mobility in England, 1858-2012: Surnames and social mobility." *Economic Journal* 125:61–85.

Clark, Gregory, Daniel Marcin, Firas Abu-sneneh, Wilfred M Chow, Kuk Mo, Ariel M Marek, and Kevin M Williams. 2012. "Social Mobility Rates in the USA , 1920-2010 : Surname Analysis A." .

Corak, Miles. 2006. "Do poor children become poor adults? Lessons from a cross country comparison of generational earnings mobility." In *Dynamics of Inequality and Poverty (Research on Economic Inequality, Volume 13)*, edited by John Creedy and Guyonne Kalb, pp. 143–188. The Netherlands: Elsevier Press.

Corak, Miles and Andrew Heisz. 1999. "The intergenerational earnings and income mobility of Canadian men: Evidence from longitudinal income tax data." *Journal of Human Resources* 34:504–533.

Dahl, MW and T DeLeire. 2008. "The association between children's earnings and fathers' lifetime earnings: estimates using administrative data." http://irp.wisc.edu/publications/dps/pdfs/dp134208.pdf.

Duncan, Otis Dudley. 1965. "The Trend of Occupational Mobility in the United States." *American Sociological Review* 30:491–498.

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

Feigenbaum, James J. 2016b. "Intergenerational Mobility during the Great Depression." http://jamesfeigenbaum.github.io/research/pdf/jmp.pdf.

Gentzkow, Matthew, Jesse M. Shapiro, and Michael Sinkinson. 2011. "The Effect of Newspaper Entry and Exit on Electoral Politics." *American Economic Review* 101:2980–3018.

Goldin, Claudia. 2006. "The Quiet Revolution That Transformed Women's Employment, Education, and Family." *American Economic Review* 96:1–21.

Goldin, Claudia and Lawrence F. Katz. 2000. "Education and Income in the Early Twentieth Century: Evidence from the Prairies." *Journal of Economic History* 60:782–818.

Goldin, Claudia and Lawrence F. Katz. 2008. *The Race Between Education and Technology*. Cambridge, MA: The Belknap Press of Harvard University Press.

Goldin, Claudia and Lawrence F. Katz. 2011. "Mass Secondary Schooling and the State: The Role of State Compulsion in the High School Movement." In *Understanding Long Run Economic Growth*, edited by Dora L. Costa and Naomi R Lamoreaux. Chicago, IL.

Goldin, Claudia and Robert A. Margo. 1992. "The Great Compression: The Wage Structure in the United States at Mid-Century." *Quarterly Journal of Economics* 107:1–34.

Grawe, Nathan D. 2006. "Lifecycle bias in estimates of intergenerational earnings persistence." *Labour Economics* 13:551–570.

Guell, M., J. V. Rodriguez Mora, and Christopher I. Telmer. 2015. "The Informational Content of Surnames, the Evolution of Intergenerational Mobility, and Assortative Mating." *The Review of Economic Studies* 82:693–735.

Guest, Avery M., Nancy S. Landale, and James C. McCann. 1989. "Intergenerational Occupational Mobility in the Late 19th Century United States." *Social Forces* 68:351–378.

Haider, Steven and Gary Solon. 2006. "Life-Cycle Variance in the Association between Current and Lifetime Earnings." *American Economic Review* 96:1308–1320.

Hertz, Tom. 2009. "Rags, Riches, and Race: The Intergenerational Economic Mobiltiy of Black and White Families in the United States." In *Unequal Chances: Family Background and Economic Success*, pp. 165–191.

Hertz, Tom, Tamara Jayasundera, Patrizio Piraino, Sibel Selcuk, Nicole Smith, and Alina Verashchagina. 2007. "The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends." *The B.E. Journal of Economic Analysis & Policy* 7.

Hornbeck, Richard. 2012. "The Enduring Impact of the American Dust Bowl: Short- and Long-Run Adjustments to Environmental Catastrophe." *American Economic Review* 102:1477–1507.

Hornbeck, Richard and Pinar Keskin. 2014. "The Historically Evolving Impact of the Ogallala Aquifer: Agricultural Adaptation to Groundwater and Drought." *American Economic Journal: Applied Economics* 6:190–219.

Krueger, Alan B. 2012. "The Rise and Consequences of Inequality."

Lee, CI and Gary Solon. 2009. "Trends in intergenerational income mobility." *Review of Economics and Statistics* 91:766–772.

Lemieux, Thomas. 2006. "The Mincer Equation Thirty Years After Schooling, Experience and Earnings." In *Jacob Mincer A Pioneer of Modern Labor Economics*, edited by Shoshana Grossbard, volume 79, pp. 127–145. Springer US.

Levine, David I and Bhashkar Mazumder. 2002. "Choosing the Right Parents: Changes in the Intergenerational Transmission of Inequality Between 1980 and the Early 1990s." http://www.ssrn.com/abstract=323881.

Long, Jason and Joseph P Ferrie. 2004. "Geographic and Occupational Mobility in Britain and the U.S., 1850-1881." http://www.colby.edu/economics/faculty/jmlong/research/usbritainmobility.pdf.

Long, Jason and Joseph P Ferrie. 2007. "The Path to Convergence: Intergenerational Occupational Mobility in Britain and the US in Three Eras."

Long, Jason and Joseph P Ferrie. 2008. "Intergenerational Occupational Mobility in Britain and the US Since 1850."

Long, Jason and Joseph P Ferrie. 2013. "Intergenerational Occupational Mobility in Great Britain and the United States Since 1850." *The American Economic Review* 103:1109–1137.

Mayer, Susan E and Leonard M Lopoo. 2005. "Has the Intergenerational Transmission of Economic Status Changed?" *Journal of Human Resources* 40:169–185.

Mazumder, Bhashkar. 2009. "The Apple Falls Even Closer to the Tree Than We Thought." In *Unequal Chances: Family Background and Economic Success*, pp. 80–99.

Mazumder, Bhashkar. 2015. "Estimating the Intergenerational Elasticity and Rank Association in the US: Overcoming the Current Limitations of Tax Data." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2647277.

McGraw, Gary, John Rehling, and Robert Goldstone. 1994. "Letter Perception: Toward a conceptual approach." In *Sixteenth Annual Conference of the Cognitive Science Society*, pp. 613–618, Atlanta, GA.

Mincer, Jacob A. 1974. "Individual Acquisition of Earning Power." In *Schooling, Experience, and Earnings*, edited by Jacob A. Mincer, volume I, pp. 7–22. New York, NY: Columbia University Press.

Nybom, Martin and Jan Stuhler. 2014a. "Biases in Standard Measures of Intergenerational Income Dependence." https://janstuhler.files.wordpress.com/2013/06/lifecycle_bias_pt2_november_2014.pdf.

Nybom, Martin and Jan Stuhler. 2014b. "Interpreting Trends in Intergenerational Mobility." http://www.sofi.su.se/polopoly_fs/1.170138.1394463038!/menu/standard/file/WP14no3.pdf http://www.homepages.ucl.ac.uk/ uctpjst/stuhler_interpreting_trends_(JMP).pdf.

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

Parman, John M. 2011. "American Mobility and the Expansion of Public Education." *The Journal of Economic History* 71:105–132.

Perez, Santiago. 2016a. "Moving to Opportunity: Railroads, migrations and economic mobility in 19th century Argentina." http://web.stanford.edu/ santip/.

Perez, Santiago. 2016b. "The (South) American Dream: Mobility and Economic Outcomes of First and Second Generation Immigrants in 19th Century Argentina." http://web.stanford.edu/ santip/Perez_South_American.pdf.

Rockoff, Hugh. 2008. "US Economy in World War I."

Rosenthal, Caitlin C. 2013. *From Memory to Mastery: Accounting for Control in America, 1750-1880*. Ph.D. thesis, Harvard University.

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, MN: Minnesota Population Center [producer and distributor].

Simpson, Ian C, Petroula Mousikou, Juan Manuel Montoya, and Sylvia Defior. 2013. "A letter visual-similarity matrix for Latin-based alphabets." *Behavior research methods* 45:431–9.

Solon, Gary. 1989. "Biases in the Estimation of Intergenerational Earnings Correlations." *Review of Economics and Statistics* 71:172–174.

Solon, Gary. 1999. "Intergenerational Mobility in the Labor Market." *Handbook of Labor Economics* 3.

Tassava, Christopher. 2008. "The American Economy during World War II."

Thernstrom, Stephan. 1964. *Poverty and progress; social mobility in a nineteenth century city*. Cambridge, MA: Harvard University Press.

Thernstrom, Stephan. 1973. *The other Bostonians : poverty and progress in the American metropolis, 1880-1970*. Cambridge, MA: Harvard University Press.

Xie, Yu and Alexandra Killewald. 2013. "Intergenerational Occupational Mobility in Great Britain and the United States Since 1850: Comment." *American Economic Review* 103:2003–2020.

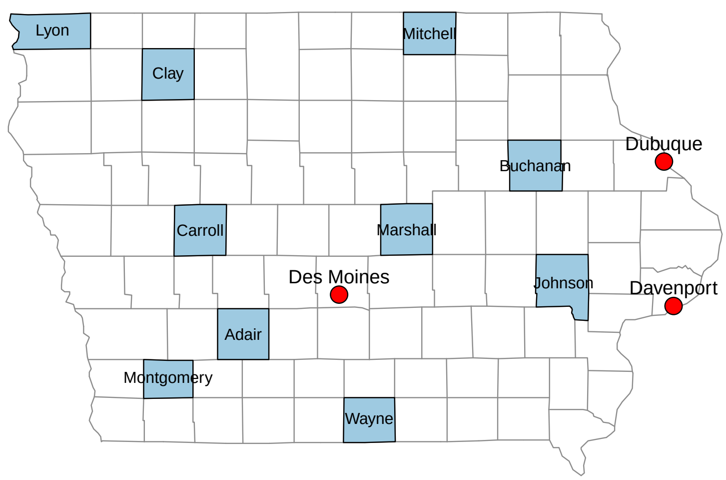
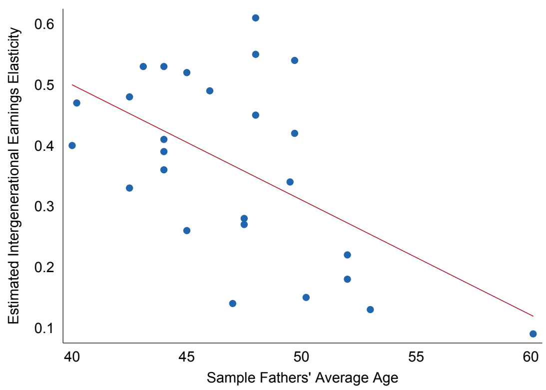
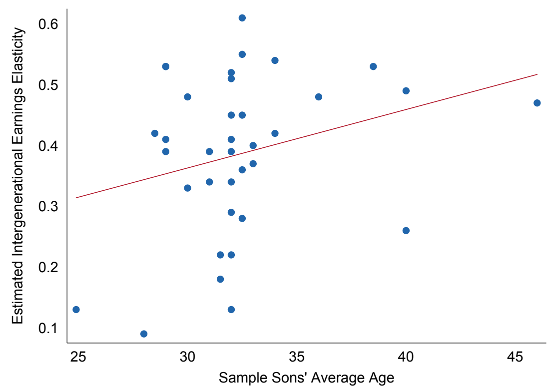


Figure 1: Map of Iowa 1915 Cities and County Sample

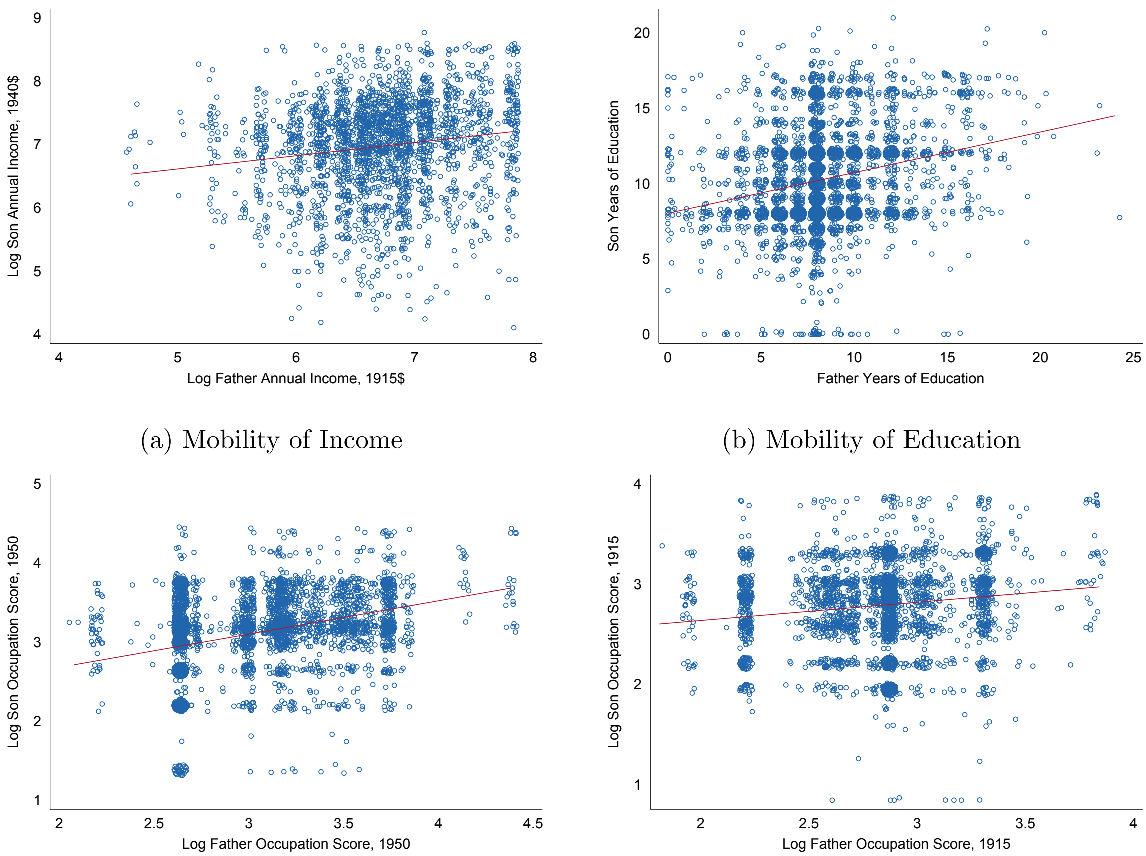


(a) Correlation between estimated IGE and sample fathers' ages



(b) Correlation between estimated IGE and sample sons' ages

Figure 2: The average ages of both fathers and sons sampled may bias the estimated intergenerational mobility elasticity. Each point in the two scatterplots represent the estimated elasticity of income from studies of American intergenerational mobility reviewed by Corak (2006), plotted against the average age of either fathers or sons in the samples.



(a) Mobility of Income

(b) Mobility of Education

(c) Mobility of Occupation, 1950 Basis

(d) Mobility of Occupation, 1915 Basis

Figure 3: Intergenerational Mobility for Iowa 1915 to 1940.

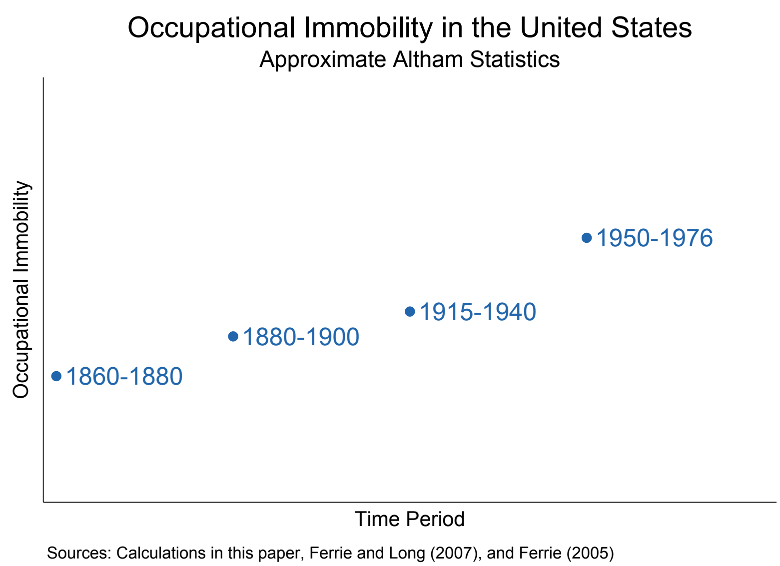


Figure 4: Occupational Immobility measured by Altham statistics from 1860 to 1976. This graph suggests that occupational immobility rose over time and that there is less mobility in the mid-twentieth century than there was between 1915 and 1940. The Altham statistics for each period presented in the plot are statistically different from one another. See Table 9 for exact values and distance tests.

Table 1: Probability of Matching a Record from Iowa 1915 to the Federal Census 1940

	(1)	(2)	(3)	(4)	(5)
Name commonness, first name	0.041^{**}				0.056^{***}
	(0.017)				(0.020)
Name commonness, last name	-0.122^{***}				-0.121^{***}
	(0.039)				(0.039)
String length, first name		0.013^{***}			0.020^{***}
		(0.004)			(0.004)
String length, last name		−0.002			−0.002
		(0.004)			(0.004)
Normalised letter similarity score, first name			0.019^{***}		0.024^{***}
			(0.007)		(0.007)
Normalised letter similarity score, last name			0.006		0.005
			(0.007)		(0.007)
Normalised scrabble score, first name				−0.001	−0.002
				(0.006)	(0.007)
Normalised scrabble score, last name				0.009	0.008
				(0.006)	(0.006)
Observations	7580	7580	7580	7580	7580
Clusters	4731	4731	4731	4731	4731
Adjusted R^2	0.002	0.002	0.001	0.000	0.007

Linear probability model with an indicator variable for a successful match as the outcome. Standard errors are clustered by family. Results are consistent using a probit or logit model as well. Name commonness is measured as the share of 100 men in the 1910 and 1920 IPUMS sample with the same first or last name. Name length is the number of characters in the first or last name. Name similarity scores are based on character typology similarity from Simpson et al. (2013). Standardised z-scores are used for both the visual similarity scores and the Scrabble scores; the z-scores are based on the distribution of visual similarity scores and Scrabble scores within the pooled sample of my Iowa sons and the 1910 and 1920 censuses.

Sources: 1915 Iowa State Census Sample; 1940 Federal Census

Table 2: Effects of Family Covariates on the Probability of Matching Records from 1915 to 1940

			Predicted Match Rate with X at	
X	β	SE	25th Percentile	75th Percentile
Father Log Earnings	0.013	0.011	59.6	60.6
Father Education	0.004	0.002	59.2	60.0
Mother Education	0.003	0.003	59.8	60.3
Urban in 1915	-0.034	0.012	60.5	57.1
Son Born in IA	0.138	0.018	61.0	61.0
Father Foreign Born	-0.063	0.013	61.2	54.8

This table presents the coefficients from a series of linear probability regressions with X as the primary independent variable, controlling for first and last name commonness, length, letter similarity, and Scrabble score. As in Table 1, there are 7580 observations and 4731 clusters, clustering standard errors by family.

Sources: 1915 Iowa State Census Sample; 1940 Federal Census

Table 3: Iowa and the Nation in 1910 and 1920

	1910				1920			
	Iowa	Midwest	Nation	Iowa Rank	Iowa	Midwest	Nation	Iowa Rank
Total Population (100k)	22.25	298.89	919.72	15	24.04	340.20	1057.11	16
Population Density (people pre sq mi)	40.02	39.52	30.93	24	43.25	44.98	35.55	24
Urbanisation Rate	30.57	45.12	45.66	25	36.42	52.25	51.23	24
Rural Rate	69.43	54.86	53.66	23	63.58	47.75	48.77	25
White Share	99.30	97.96	88.87	5	99.17	97.49	89.70	6
Black Share	0.67	1.82	10.69	32	0.79	2.33	9.90	34
Foreign Born Share	12.29	15.66	14.51	28	9.39	13.51	12.97	27
Native Born with Foreign Parents Share	28.42	27.85	20.55	14	26.21	27.35	21.46	19
Sex Ratio (M/F)	1.07	1.08	1.06	24	1.05	1.06	1.04	24
Farms (per 100 people)	9.76	7.47	6.92	16	8.88	6.41	6.10	17
Share of Farm Acres Improved	86.92	72.23	54.44	1	85.46	69.20	52.63	1
Literacy Rate (ages 10+)	98.66	97.47	94.00	1	99.14	97.98	95.33	1
School Enrollment Rate (ages 6 to 20)	69.57	66.47	62.34	5	72.39	70.33	68.10	13
Newspaper Circulation per thousand	220.63	270.57	237.21	18	313.68	310.62	265.55	11
Daily Newspapers per 100k	2.79	2.95	2.42	15	2.16	2.33	2.02	20
Gini Coefficient (based on farm sizes)	0.38	0.47	0.54	47	0.35	0.47	0.56	47

The Midwest region is defined as the states of Illinois, Indiana, Michigan, Ohio, Wisconsin, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota and South Dakota. The newspaper data, drawn from Gentzkow et al. (2011), was only digitised for presidential election years, so the 1910 data reflects newspaper counts and circulation for 1912. The native born with foreign parents share includes both the native born with two foreign parents and the native born with one foreign parent.

Sources: 1910 and 1920 Federal Censuses; Gentzkow et al. (2011)

Table 4: Intergenerational Mobility Estimates

	Specification (1)	(2)	(3)	Observations	Clusters
A. Intergenerational Elasticity (IGE)					
Full Sample	0.208	0.199	0.258	2039	1668
	(0.032)	(0.031)	(0.081)		
PSID Iowa-Like Sample	0.330	0.350	0.502	3449	346
	(0.056)	(0.080)	(0.166)		
Urban Sample	0.287	0.275	0.310	1004	824
	(0.045)	(0.050)	(0.102)		
Rural Sample	0.156	0.167	0.233	1035	844
	(0.040)	(0.041)	(0.113)		
Excluding Sons of Farmers	0.301	0.259	0.391	1452	1200
	(0.037)	(0.038)	(0.095)		
Only Native Grandparents Sample	0.287	0.277	0.411	714	600
	(0.045)	(0.046)	(0.116)		
Only Foreign Grandparents Sample	0.145	0.159	0.160	641	516
	(0.059)	(0.059)	(0.123)		
Including Sons with Imputed Income	0.162	0.156	0.217	2864	2161
	(0.021)	(0.022)	(0.050)		
All Sons, Imputed 1950 Income	0.066	0.066	0.061	3144	2324
	(0.010)	(0.010)	(0.019)		
Sons Remaining in Iowa	0.147	0.148	0.173	1184	1001
	(0.043)	(0.043)	(0.121)		
B. Intergenerational Rank Rank Parameter					
Full Sample	0.172	0.167	0.217	2039	1668
	(0.022)	(0.021)	(0.045)		
PSID Iowa-Like Sample	0.258	0.240	0.323	3680	356
	(0.049)	(0.064)	(0.084)		
Urban Sample	0.217	0.206	0.214	1004	824
	(0.032)	(0.034)	(0.071)		
Rural Sample	0.141	0.150	0.224	1035	844
	(0.028)	(0.028)	(0.060)		
Excluding Sons of Farmers	0.230	0.201	0.254	1452	1200
	(0.026)	(0.027)	(0.058)		
Only Native Grandparents Sample	0.240	0.224	0.312	714	600
	(0.034)	(0.035)	(0.078)		
Only Foreign Grandparents Sample	0.140	0.159	0.178	641	516
	(0.041)	(0.038)	(0.081)		
Including Sons with Imputed Income	0.140	0.136	0.181	2864	2161
	(0.016)	(0.016)	(0.030)		
All Sons, Imputed 1950 Rank	0.046	0.046	0.037	3266	2417
	(0.009)	(0.009)	(0.017)		
Sons Remaining in Iowa	0.126	0.132	0.187	1184	1001
	(0.028)	(0.027)	(0.058)		

Standard errors clustered by family in all regressions. In Panel A, son's annual log earnings in 1940 is the dependent variable. In Panel B, the son's rank in the income distribution in 1940 is the dependent variable. The income distribution in 1940 calculated using the 1940 IPUMS 1% sample. Specification 1 is a univariate regression of son's outcome on father's outcome (log earnings or income rank). Specification 2 adds name string controls, 1915 county fixed effects, and quartic controls in father and son age. Specification 3 adds an interaction between father's outcome and son's normalised age. Name string controls: first and last name commonness, length, letter similarity, and Scrabble scores. Son's ages are normalised relative to age 40 in 1940.

Sources: 1915 Iowa State Census Sample; 1940 Federal Census

Table 5: Income Rank Quintile Transition Matrix

A. Iowa 1915 to 1940 Earnings Quintile Transition						
	Father's Quintile					
Son's Quintile	1	2	3	4	5	Total
1	15.0	7.2	8.7	10.0	7.2	9.2
2	23.1	22.7	15.7	17.9	15.1	18.6
3	27.7	24.3	25.4	19.1	18.1	22.8
4	20.5	26.5	24.2	24.0	23.3	24.3
5	13.8	19.3	26.1	29.1	36.2	25.1
Total	17.6	21.3	19.7	20.7	20.7	
B. Contemporary Iowa Earnings Quintile Transition						
	Father's Quintile					
Son's Quintile	1	2	3	4	5	Total
1	26.1	15.9	10.8	7.7	7.1	12.1
2	23.0	18.9	15.0	12.0	9.9	15.1
3	20.2	21.5	21.1	19.1	16.1	19.7
4	17.7	23.6	26.5	27.8	25.6	25.2
5	12.9	20.0	26.6	33.5	41.3	27.9
Total	12.3	19.1	26.8	25.1	16.7	

In Panel A, the cells in this table report the probability that a son with a father in a given income quintile in 1915 (column) will be a given income quintile in 1940 (row). The income distribution in 1940 calculated using the 1940 IPUMS 1% sample. The income distribution in 1915 is calculated using the Goldin-Katz 1915 Iowa State Census sample. In Panel B, the cells report the probability that a son with a father in a given income quintile in the Chetty et al. (2014a) data will be a given income quintile as an adult (row). Panel B replicates the distribution of children in my Iowa 1915 sample, weighting the transition probabilities by the number of sons in each county in my sample in 1915 (and linking counties to commuting zones, as reported by Chetty et al. (2014a)).

Sources: 1915 Iowa State Census Sample; 1940 Federal Census; Chetty et al. (2014a)

Table 6: Alternative Intergenerational Mobility Estimates

	Sample	Specification (1)	(2)	(3)	Observations	Clusters
A. Education Mobility						
Years of Education	Full Sample	0.264	0.241	0.206	3376	2504
		(0.023)	(0.023)	(0.040)		
	Urban Sample	0.297	0.269	0.301	1283	1028
		(0.035)	(0.035)	(0.058)		
	Rural Sample	0.235	0.223	0.142	2093	1476
		(0.029)	(0.030)	(0.056)		
	Excluding Sons of Farmers	0.317	0.303	0.338	2004	1589
		(0.027)	(0.028)	(0.046)		
	Only Native Grandparents Sample	0.323	0.313	0.308	1117	859
		(0.036)	(0.036)	(0.062)		
	Only Foreign Grandparents Sample	0.161	0.154	0.076	1139	819
		(0.050)	(0.047)	(0.080)		
B. 1915 Occupation Score Mobility						
Log Occupation Score 1915 Basis	Full Sample	0.190	0.184	0.265	3037	2279
		(0.026)	(0.026)	(0.052)		
	Urban Sample	0.229	0.226	0.308	1154	940
		(0.035)	(0.036)	(0.065)		
	Rural Sample	0.145	0.133	0.175	1883	1339
		(0.039)	(0.038)	(0.091)		
	Excluding Sons of Farmers	0.203	0.204	0.288	1760	1414
		(0.026)	(0.026)	(0.053)		
	Only Native Grandparents Sample	0.242	0.242	0.412	1002	776
		(0.040)	(0.040)	(0.084)		
	Only Foreign Grandparents Sample	0.123	0.130	0.103	1038	756
		(0.048)	(0.048)	(0.081)		
C. 1950 Occupation Score Mobility						
Log Occupation Score 1950 Basis	Full Sample	0.441	0.366	0.386	3202	2374
		(0.021)	(0.024)	(0.045)		
	Urban Sample	0.258	0.247	0.291	1220	980
		(0.036)	(0.036)	(0.068)		
	Rural Sample	0.424	0.418	0.401	1982	1394
		(0.030)	(0.030)	(0.070)		
	Excluding Sons of Farmers	0.229	0.220	0.263	1867	1479
		(0.027)	(0.027)	(0.061)		
	Only Native Grandparents Sample	0.351	0.298	0.302	1053	807
		(0.034)	(0.037)	(0.081)		
	Only Foreign Grandparents Sample	0.505	0.361	0.369	1084	780
		(0.037)	(0.044)	(0.075)		

Son's completed years of education in 1940 is the dependent variable in panel A. The log of the son's occupation score, using either the 1915 or 1950 occupation score measures, is the dependent variable in panel B (1915) and C (1950). Standard errors clustered by family. Specification 1 is a univariate regression of son's outcome on father's outcome. Specification 2 adds name string controls, 1915 county fixed effects, and quartic controls in father and son age. Specification 3 adds an interaction between father's outcome and son's normalised age. Name string controls: first and last name commonness, length, letter similarity, and Scrabble scores. Son's ages are normalised relative to age 40 in 1940.

Sources: 1915 Iowa State Census Sample; 1940 Federal Census

Table 7: Intergenerational Mobility Based on First Names

	Sample	Specification (1)	(2)	(3)	Observations	Clusters
		A. Log Earnings				
First Name Based Father's Occupation Score	Full Sample	0.300	0.177	0.404	2942	2248
		(0.102)	(0.103)	(0.250)		
	Urban Sample	0.314	0.333	0.433	1156	945
		(0.170)	(0.172)	(0.470)		
	Rural Sample	0.123	0.094	0.367	1786	1303
		(0.126)	(0.129)	(0.309)		
	Excluding Farmer Sons	0.382	0.309	0.741	1792	1457
		(0.134)	(0.136)	(0.416)		
	Only Native Grandparents Sample	0.310	0.260	0.157	950	758
		(0.182)	(0.185)	(0.510)		
	Only Foreign Grandparents Sample	0.279	0.122	0.324	1000	743
		(0.190)	(0.190)	(0.408)		
		B. 1950 Occupation Score Mobility				
First Name Based Father's Occupation Score	Full Sample	0.353	0.178	0.318	3158	2372
		(0.074)	(0.076)	(0.148)		
	Urban Sample	0.171	0.184	0.120	1205	977
		(0.106)	(0.112)	(0.216)		
	Rural Sample	0.164	0.183	0.430	1953	1395
		(0.097)	(0.102)	(0.206)		
	Excluding Farmer Sons	0.160	0.165	0.190	1867	1498
		(0.083)	(0.087)	(0.187)		
	Only Native Grandparents Sample	0.237	0.060	0.155	1032	804
		(0.121)	(0.127)	(0.319)		
	Only Foreign Grandparents Sample	0.638	0.306	0.546	1078	787
		(0.123)	(0.118)	(0.214)		

Log of son's earnings in 1940 is the dependent variable in panel A. The log of the son's occupation score using the 1950 occupation score measure is the dependent variable in panel B. Standard errors clustered by family. The main independent variable is the father's imputed status, based on the son's first name, following Olivetti and Paserman (2015). Specifically, the father's status is the mean of log occupation score among all father's with sons of a given first name, based on the 1910 IPUMS sample of the census. Specification 1 is a univariate regression of son's outcome on father's name-based imputed occupation score. Specification 2 adds name string controls, 1915 county fixed effects, and quartic controls in father and son age. Specification 3 adds an interaction between father's outcome and son's normalised age. Name string controls: first and last name commonness, length, letter similarity, and Scrabble scores. Son's ages are normalised relative to age 40 in 1940.

Sources: 1915 Iowa State Census Sample; 1940 Federal Census; Olivetti and Paserman (2015)

Table 8: Occupation Tranisitions Table for Fathers and Sons, 1915 to 1940

	Father's Occupation				
Son's Occupation	Farmer	Skilled	Unskilled	White Collar	Total
Farmer	599	56	44	31	730
Skilled	288	380	219	190	1,077
Unskilled	332	179	195	91	797
White Collar	244	313	159	337	1,053
Total	1,463	928	617	649	3,657

Father's occupation categories are determined from the 1915 Iowa State Census. Son's occupation categories are determined from the 1940 Federal Census. Total counts do not match previous totals for fathers and sons in other tables because some observations contain information on wages or education but with occupation descriptions that cannot be linked to broad categories.

Sources: 1915 Iowa State Census Sample; 1940 Federal Census

Table 9: Altham Statistic Summary, Iowa 1915 to 1940

	Sample (P)		
	All Sons	Urban Sons	Rural Sons
Altham Statistic ($d(P, I)$)	16.14	10.74	14.34
$Pr(d(P, US1880) = 0)$	0.000	0.627	0.141
$Pr(d(P, US1900) = 0)$	0.000	0.017	0.511
$Pr(d(P, US1973) = 0)$	0.000	0.003	0.004
$Pr(d(P, IAsons) = 0)$	.	1.000	0.718
$Pr(d(P, IAurbansons) = 0)$	0.718	.	0.786
$Pr(d(P, IAruralsons) = 0)$	0.711	0.786	.

Father's occupation categories are determined from the 1915 Iowa State Census. Son's occupation categories are determined from the 1940 Federal Census. The Altham-Ferrie statistic is a distance metric; the distance from the identity matrix I can be interpreted as a measure of mobility with higher values implying less mobility. The distance metric can also compare two occupation transition matrices. Altham-Ferrie statistics for US1880 (a father-son linked sample between fathers in 1860 and sons in 1880) is 12.09, for US1900 (fathers in 1880 and sons in 1900) it is 14.58, and for US1973 (fathers in 1950 and sons in 1973) it is 20.76. The above results reject that occupation category transitions were the same between 1915 and 1940 and any of the other periods. In particular, there was more occupation transition mobility in the nineteenth century than the early twentieth century and more mobility in the early twentieth century than between 1950 and 1973.

Sources: 1915 Iowa State Census Sample; 1940 Federal Census; Long and Ferrie (2007)

A Appendix

To be included as a web-only appendix with additional tables, figures, and robustness checks.

A.1 Additional Tables and Figures

[Table A.1 about here.]

[Table A.2 about here.]

[Table A.3 about here.]

[Figure A.1 about here.]

A.2 Matching Bias and Measurement Error

A.2.1 Matching Bias

The analysis that I conduct in this paper requires the construction of a dataset that links fathers and sons over time, between two censuses. The linking procedure, though carefully conducted, likely introduces a type of measurement error and bias to the estimation of IGE parameters for historical periods that might not be present in contemporary data. As these errors (or the mismatching rate) grow, the likely estimate of the IGE or of the rank-rank parameter will fall. To some extent, this could explain why an estimate of the historical IGE is smaller than a contemporary estimate of the same parameter, even if the true values are the same. I follow a mismatching simulation procedure based on the one used by Parman (2011) to gauge the magnitude of these biases.[79] While it is impossible to know exactly the rate of mismatches in the my linked sample between Iowa 1915 and the Federal 1940 census, I can introduce different levels of mismatch into the PSID data and measure the effect on estimated IGE parameters for matching error.

To determine the appropriate mismatching simulation for my data, I begin by reexamining the actual matching procedure used between the 1915 and 1940 censuses. I observe families with fathers and sons in 1915 where sons are between 3 and 17 years of age. There is no restriction on father's ages in these sample families. I then search for the sons in the 1940 Federal Census, using uniquely identifying information such as first and last name, state of birth, and year of birth. However, despite my best efforts at ensuring a unique and correct match, I may identify the 'wrong' son in 1940.[80]

Suppose the match error rate is π. That is, if I make 100 matches, then $\pi \times 100$ matches will be erroneous. To replicate a π share of matching errors in the PSID, I drop the son's income and education data for π of the father-son pairs. Then, I randomly draw new son outcome data (income or education, independently), conditional on the true son's age.[81] Using this new data I estimate

[79] Though Parman also draws on the 1915 Iowa Census, the construction of my dataset of linked fathers and sons varies somewhat from that used by Parman (2011). Thus, my simulation method differs from his so as to properly replicate the possible points of measurement error in the matching.

[80] Wrong sons, in this case, would be a man with the same name, state of birth, and year of birth (within a 1 or 2 year bandwidth). This is not a common name or 'John Smith' problem, as there are likely too many John Smiths born in a given state and year. Rather this might be a 'John Smitherson' problem if there are two John Smithersons but one is not found in 1940 (possibly because he is dead, out of the country, or had his name transcribed incorrectly into the census as, for example, John Smithson).

[81] I do this all conditional on the son's age because I observe the true son's age in the original 1915 sample.

an IGE parameter, following the regression specified in the main empirical section of this paper. I simulate 1000 draws for each π and, following Parman (2011), I determine the IGE for π ranging from 2 to 50%. I repeat the same procedure (with a new simulation of mismatches) to test the stability of the rank-rank parameter as well.

[Figure A.2 about here.]

Figure A.2 presents the estimated β parameter from these mismatching tests, with the rate of mismatching, π, on the x-axis. In Figure A.2a, the solid horizontal line at $\beta = 0.258$ represents my largest estimate of the IGE between 1915 and 1940 from Panel A of Table 4. These tests suggest that a mismatching rate of more than 50% would be required to generate an IGE as low as I find in the historical period, if the true IGE were the same as in contemporary samples.[82] Given that matches are made on first and last names, states of birth, and years of birth, such a high rate of mismatch seems extremely unlikely. Similarly, Figure A.2b presents the same tests but for the rank-rank parameter, following Dahl and DeLeire (2008) and Chetty et al. (2014b). In this case, the solid horizontal line is drawn at $\beta = 0.217$, the largest estimate of the rank-rank parameter for the full sample in Panel B of Table 4. In this case, the mismatch rate would have to be at least 30% to induce such a low rank-rank estimate of intergenerational mobility in the recent data as I find historically.

A.2.2 Measurement Error

As Haider and Solon (2006); Grawe (2006); Solon (1989) show, measurement error will cause serious problems for estimates of IGE parameters. I have attempted to minimise these issues with age quartic controls, age quartic interaction controls, and by sampling fathers and sons at the middle of both lifecycles. In addition, I compared my historical estimates to contemporary estimates generated with just a single year of income data observed for fathers and sons, and the difference in the results remained. Finally, my results are quite consistent between several measures of father and son outcomes—income, education, and occupational standing. These measures all suffer from their own measurement problems, but taken together the consistent results are reassuring that

[82] Doing a similar test, Parman similarly finds a mismatching rate of 50% would be required to overturn his IGE findings.

intergenerational mobility was in fact lower in the early twentieth century than it is in the recent period.

As a further test of the measurement error effects, I introduce measurement error into the presumably well-measured PSID data. Let ζ be a $N(0, \sigma^2)$ shock. I add this random noise to either the father's income, the son's income, or both (in this case, the shocks are uncorrelated). I then reestimate the IGE parameter. I simulate 1000 draws for each σ and let σ vary from 0 to 1. Figure A.3 presents the estimated betas for measurement error in both measures of the father's and son's earnings.

[Figure A.3 about here.]

A.3 Farmer Income in 1940

For each individual listed in the 1940 Federal Census, annual wage and salary earnings and weeks of work are reported.[83] However, the census does not include information on either business or farm income as in later censuses. In practise, farm owners and other business proprietors reported working a full year (52 weeks) and having zero income in 1940. Thus, for any observed sons in 1940 who are either farm owners or business proprietors, I do not observe any measure of earnings in 1940. This restriction does not apply to farm labourers: farm labour income is reported in the same way as any other form of wage or salary income. Of the 4,478 matched sons in my sample, 1,177 report zero earnings in 1940. Of these, more than half (610) are farmers or farm owners or farm operators. The other 567 are a variety of occupations, including proprietors (36), operators (31), labourers (29), owners (23). and various forms of doctors and lawyers.[84]

In the main results presented in Panels A and B of Table 4, I drop all of these observations with no earnings in 1940. However, to the extent that sons with either very high or very low intergenerational mobility select into farming in 1940, this restriction could bias my estimates. It may be the case that the sons are farmers in 1940 because they have inherited the family farm from their fathers and thus their incomes, driven perhaps in large part by the productivity in the same plot of land, are highly correlated. Given the large changes in agriculture during this period, owing both to mechanisation, the discovery of new irrigation sources in the Ogallala Aquifer, and especially the Dust Bowl (Hornbeck and Keskin, 2014; Hornbeck, 2012), this correlation may not be as strong in the early twentieth century as during other eras.

My results are consistent across other measures of mobility, particularly educational mobility, which do not suffer from this same missing data problem in 1940. As a further robustness check, I impute farm income in 1940 and re-compute the main results on intergenerational mobility below.[85] The estimated IGE and rank-rank parameters including sons with imputed capital income are

[83] IPUMS reports the specific enumeration instructions. The entry should be the 'total amount of money wages or salary' but 'Do not include the earning of businessmen, farmers, or professional persons derived from business profits, sale of corps, or fees'. For more, see `https://usa.ipums.org/usa/voliii/inst1940.shtml#584`.

[84] The most common occupation among the non-farmers without earnings is actually to have no occupation listed (83 of the 567). These men are likely unemployed and not working for the WPA and thus the zero reported earnings are a correct measure not unreported data.

[85] I do this only the 610 farmers in my sample in 1940 without reported earnings. For the other proprietors and occupations, imputation would be much less accurate given the smaller sample sizes of these various occupations in 1950 and the (perhaps) more idiosyncratic nature of earnings in these professions.

presented in the sixth rows of Table 4 and suggest even more mobility historically, relative to contemporary estimates, than do my baseline results. Thus, the fact that I have to exclude the sons who are farmers in 1940 in the main results because I do not observe their incomes is not driving the measured result that income mobility was higher in the early twentieth century than it is today.

Here, I detail the imputation process. First, I collect data from the 1950 IPUMS 1% sample, which includes measures of both wage and salary income (which I observe in 1940), as well as total income and business and farm income. In a sample of only farmers in 1950, I regress total income on years of education, a quartic in age, number of weeks worked, and indicator variables for state of birth and state of residence in 1950.[86] Using the results of this regression, I impute the total income of farmers in my 1940 sample, assuming that the earnings function in 1940 resembles that in 1950 with respect to the effects of education, age (experience), weeks of work, and location fixed effects. An additional week of work in 1950 increases total income by 0.22%. As Figure A.4 shows, the relationship between years of education and the log of total income is nearly linear, with a slope of 2%, while the relationship between age and total income is non-linear.[87] Both the state of residence and state of birth fixed effects are quite strong as well. Overall, the R^2 is 15.52. I convert the imputed 1950 earnings to 1940 earnings with the price deflator.

[Figure A.4 about here.]

[86]With a large sample in 1950, I measure the effect of education nonparametrically with indicator variables for each year of education.

[87]The returns to education are quite low in this imputation because I am also controlling for state of residence. Among farmers, differential mobility and location choice is likely one of the channels through which education determines earnings.

A.4 Imputing Relationships

The 1915 Iowa State Census lacks household relationships. The raw data is stored not in tabular form, as is the case for federal censuses, but rather, in card form, with one card for each individual in the state. Goldin and Katz (2008) create household families in the data based on card position and last name and address matching. However, to link fathers and sons between 1915 and 1940, the existing family identification is insufficient. Instead, I need to assign roles within each family, to identify the father, mother, and children, as well as any other non-nuclear family members in the household. I create an algorithm to assign probable family roles to each member. 85% of families observed in the Iowa 1915 data have two married people and the rest single, with the married people of the opposite sex and the other family members younger than the married couple. In these cases, assigning roles is trivial.[88] For the rest, I use simple rules based on marital status of family members, sex, and age.

I test my algorithm on the IPUMS 1% samples for 1910 and 1920 for Iowa that does include household roles (Table A.4). I report the true census relationship from IPUMS across the table and my imputed family relationship down the table.[89] Generally, my family relationship imputation does quite well in replicating the family positions for citizens of Iowa in 1910 and 1920. Among fathers in the 1910 and 1920 IPUMS samples, my imputation algorithm identifies 98.7% as fathers and only 1.2% of the identified fathers are false positives. Among children, 98.8% are identified properly and only 3.1% of the identified children are false positives.

[Table A.4 about here.]

A.5 Intergenerational Mobility under Alternative Function Forms

Measuring intergenerational mobility of income using the log-log specification, as is done in the main section of this paper and in the intergenerational mobility literature, constrains the relationship between the incomes of successive generations to have a very particular function form. The log

[88] Of course, some of those children could be step-siblings or half-siblings or live-in cousins. Unfortunately, there is no way for me to know this with any certainty. In terms of comparability with recent data however, studies of intergenerational mobility using the PSID, for example, rely not on measures of biological fathers', but on income of the male head of household in the child's house during childhood. Thus, misassignment of step-fathers as fathers is not a major problem.

[89] I should note that the so-called true census relationship are in fact imputed by IPUMS as well, based on family roles relative to the head of household reported on the census, as well as age, sex, and name.

function assigns the same weights to percentage changes in income, rather than to absolute changes in income. Thus very small changes in income at the bottom of the distribution are given the same weight as much larger change in income elsewhere in the distribution. To the extent that farmers were growing their own food and not selling production on the market, their reported or cash incomes would understate their true incomes or status.

However, I show here that my standard results are robust to alternative transformations of annual income. First, in Panel A of Table A.5, I show that normalising earnings to be weekly, rather than annual, increases historical levels of mobility.[90]

In Panel B of Table A.5, I present intergenerational 'elasticity' estimations with the income variables in levels.[91] Panel C of Table A.5 uses a square root transformation.

I also recompute the intergenerational elasticity of education using a log-log specification. In this case, the weighting implied by a log transformation is somewhat unnatural. Lemieux (2006) argues that in contemporary data the returns to education in a traditional Mincerian framework are convex, suggesting that each additional year of schooling is actually more valuable than the previous year.[92] By logging both the father's and the son's years of education, this specification implies that the return to each year of schooling for the father is decreasing (where the 'returns' are measured as the years of completed for the son, rather than in wages, as is usual). The estimates presented in Panel E of Table A.5 suggest that the IGE parameter is smaller than I found earlier in this paper. Using the more traditional levels version, I found an IGE for education of between 0.187 and 0.275. Here, in logs, the IGE is between 0.10 and 0.19 and the confidence intervals for these sets of estimates do overlap.

A.5.1 Intergenerational Mobility using Family Income

To generate the contemporary estimates of the intergenerational elasticity of income, researchers typically measure income at the family level rather than at the individual level, on both the right

[90] In 1915, the Iowa State Census measured the number of months unemployed for respondents. In 1940, the Federal Census measured the number of weeks employed. Using these variables, I can easily construct earnings per week employed.

[91] This is a slight abuse of notation common to the IGE literature. When the father's and son's incomes are no longer logged, the parameters are not truly elasticities.

[92] There is no work, that I am aware of, in the spirit of Lemieux that reconsiders the exact polynomial function of education that best fits the data in a Mincerian wage regression. Mincer (1974) uses untransformed years of schooling in his canonical study.

hand side (fathers) and on the left hand side (sons) (for example Lee and Solon, 2009). This is a necessary definition of earnings when the goal is to measure the relation, broadly, between outcomes from one generation to the next. However, given historical patterns of female labour force participation (Goldin, 2006), I have chosen in this paper to measure income at the individual level. First, this more accurately replicates the occupational mobility literature, which measures the occupational categories of fathers and sons, ignoring mothers and spouses. Second, the collection of wives' income from the 1940 census would have added an additional round of costly data collection and little usable data, given how few married women worked in this period. In the Goldin-Katz 1915 Iowa sample, the overall correlation between family income and the income of the head of household is 0.9951; among my sample of fathers (limited to those with matched sons in 1940), the correlation is 0.9976. Thus, while for some families, possibly those with disabled or sick fathers, the mother's income could be a valuable resource, in practice the father's income and the family's income are nearly identical. The results presented in Panel D of Table A.5 underscore that expectation.

[Table A.5 about here.]

A.6 Construction of Occupational Score from 1915

Prior to 1940, the United States Federal Census did not ask respondents to report annual income. Economic historians and others interested in income and occupational standing have instead used reported occupations to measure social status, linking the occupations to median income by occupation from 1950. These so-called occscores are provided by IPUMS in all census data extracts before 1950. However, while such occscores likely provide some information on the expected income of a given census respondent, the signal to noise ratio falls as the analysis shifts to earlier census data. This occurs for two main reasons. First, the measurement error in matching occupations across time increases with time. While the tasks performed by an accountant or bookkeeper were very similar between 1950 and 1940, they are far different from the tasks performed by accountants at the turn of the twentieth century.[93] Second, in the response to both uneven technological change over time as well as shifting supply and demand for various types of labour, the returns to some occupations will fall and the returns to other will rise. The magnitudes of these changes to

[93]On the historical occupation tasks of accounting and bookkeeping specifically, see Rosenthal (2013).

technology, supply, and demand are likely to grow over time.

Taking advantage of the 1915 Iowa State Census, which was the earliest census in the US to record respondents' incomes, occupation, and education level, I construct two variants on the traditional measures of occupation score, measured not in 1950 but in 1915. While these measure are highly correlated with the occupation score provided by IPUMS based on the 1950 census, they vary in important ways and likely allow for a more accurate assessment of the income in a given occupation in the United States in the early twentieth century.

IPUMS defines the 'OCCSCORE' on a 1950 basis as:

> The occupational income score indicates the median total income – in hundreds of dollars – of the personas [sic] in each occupation in 1950. It is calculated using data from a published 1950 census report. For the post-1950 period, the score reflects the weighted average income of the 1950 occupational components of each contemporary occupation. In practise, this has only a small effect, but it means that the measure can vary slightly across census years for a given occupation.[94]

The 1950 census source used by IPUMS includes a median income for men and for women. IPUMS then weighs these medians by the sex share in each occupation to get one score for a given occupation. Using the Goldin and Katz (2008) sample of the 1915 Iowa State Census, I can create a similar median wage for each occupation group.

The Goldin and Katz (2008) Iowa sample includes a variable linking each observation to the 1940 occupation codes used in IPUMS. To generate a crosswalk between the 1940 and 1950 occupation codes, I collect the IPUMS 1940 1% sample of the census and contract the data by 1940-occupation and 1950-occupation.[95] Merging this crosswalk onto the 1915 Iowa data allows me to link observations of income in the Iowa data to occupation categories in the 1950 data. I then calculate both the simple median and sex-weighted median income within each occupation group.[96]

Figure A.5 presents scatter plots of the occupation scores for 1950 and 1915. Both measures of

[94] `https://usa.IPUMS.org/usa/chapter4/chapter4.shtml`

[95] The exact variables in the IPUMS sample are `occ` and `occ1915`.

[96] To find the sex-weighted median, I first calculate the median income for each occupation category by sex. Then I calculated the weighted average of these two medians, where the weights are the shares of men or women in each occupation. For the occupations where all observations are the same sex, the simple median and the sex-weighted median are the same. For example, in Iowa 1915, of the 100 civil engineers observed, none are women. Conversely, of the 27 private family laundresses in the sample, all are women.

occupational score in 1915 are highly correlated with the 1950 measure: the sex-weighted median is correlated at 0.7091 and the simple median at 0.7059. Given the high correlation with the traditional occupation score measure and the high correlation between my two constructed measures, I will focus on the sex-weighted median, particularly because the construction of that variable follows the IPUMS construction of the 1950 occupation score variable.

[Figure A.5 about here.]

The points farthest from the best fit line may be of some interest. These are the occupations for which the returns changed the most between 1915 and 1950. Potentially consistent with increasing returns to human capital or education, the two of the occupation categories with the largest difference between the occupation score in 1950 and 1915 are 'Physicians and surgeons' and 'Optometrists'.[97] 'Mechanical engineers' and 'Power station operators' are both relatively low-paid positions in 1915, but by 1950 they are in the upper quartile of incomes. 'Attendants, recreation and amusement' is an occupation that was relatively middle-ranked in 1915, but by 1950 is towards the bottom of the occupational ladder.

A.7 Geographic Mobility

In addition to the standard measures of mobility considered thus far, my linked 1915 and 1940 samples also allow me to estimate the correlations of father's income or education with son's geographic mobility. Figure A.6 presents a map of the residences, in 1940, of the sons included in my sample. The sons are located in almost every state in the US and most territories (territories not pictured on the map).[98] Table A.2 gives the percentage living in each of the most common states and compares these results with the geographic locations among 28-42 year old white male Iowa natives in the 1940 IPUMS 1% sample (Ruggles et al., 2010).[99] Nearly 64% remain in Iowa.[100]

[97]An alternative story for these divergences would be the difference between the rural, agrarian economy of Iowa in 1915 versus the whole US economy in 1950.

[98]Recall that I am matching from the 1915 Iowa State Census to the 1940 Federal Census. Thus, while I will be able to find sons in any of the 48 states or other territories included in the census, sons leaving the country will not be matched. There are no sons living in Delaware, New Hampshire, or Vermont. Hawaii and Alaska were not yet states and are not covered by the 1940 Federal Census sample used for son-matching.

[99]The conceptual construction of the 1% 1940 IPUMS sample does not match my sample exactly because not all sons in Iowa in 1915 (in my sample) were born in the state, but they are roughly similar.

[100]Long and Ferrie (2004) estimate geographic mobility in the US between 1850 and 1880 and find identical results for the earlier period: 64.7% of young men in their matched sample remain in the same state from 1850 to 1880.

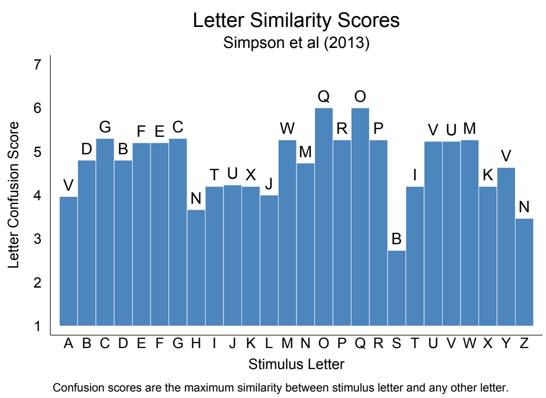
Neighbouring states, especially Illinois and Minnesota, account for 17.3% of sons. Los Angeles county is the most prominent urban destination for the sons who left Iowa, with 3.9% of the sample population, followed by Cook county, Illinois (Chicago, 3.3%), Rock Island county, Illinois (1%), Douglas county, Nebraska (Omaha, 0.9%), and Hennepin county, Minnesota (Minneapolis, 0.9%); few travel farther east than Detroit.[101]

[Figure A.6 about here.]

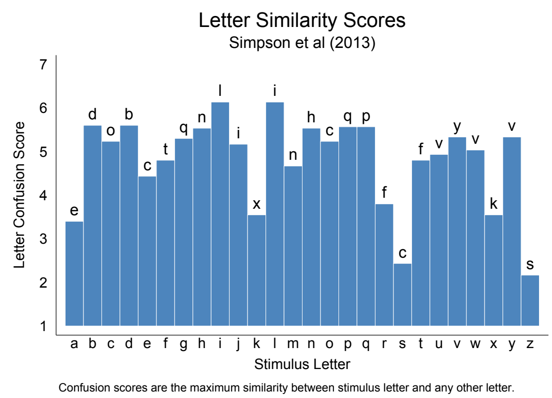
I also measure geographic mobility as the distance that the sons had moved between when they are first observed in 1915 and when they are observed again in 1940. Table A.6 suggests that distance moved increases with either the father's income or father's education, though in column 3, with both father-level variables included, only education has a significant effect. An additional year of education for the father increases the number of miles moved by the son by between 4.8% and 5.8% in the full sample. This relationship appears to be stronger for the urban sons (nearly 7% per year of education) relative to the rural sample (4% per year). While these results are more speculative, they suggest that enabling higher levels of geographic mobility may be one way in which better educated fathers (or richer fathers) improve potential outcomes for their sons. However, the importance of geographic mobility should not be overstated; earlier in this paper, I found that rural sons had more economic mobility even though they had less geographic mobility in Table A.6.

[Table A.6 about here.]

[101] Rock Island, Illinois is across the Mississippi River from Davenport, Iowa; some sons remaining in Iowa travel fewer miles than those sons moving from Davenport to Rock Island, IL or Moline, IL.

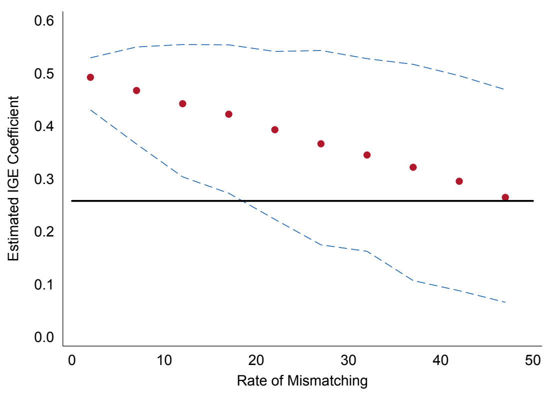


(a) Upper Case Letters

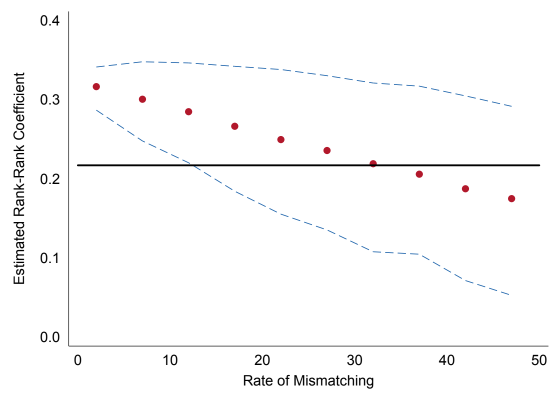


(b) Lower Case Letters

Figure A.1: Letter Similarity Scores used to calculate typographical errors. The letters listed on the x-axis are most similar to the letters printed on the column chart. For example, O and Q are most similar upper case letter pair, with a score of 6. S is the upper case letter least likely to be confused as its most similar match is B with only a score of 2.73. Among lower case letters, l and i are most similar (score of 6.13); z is the most distinct.



(a) Simulated Intergenerational Elasticity of Income



(b) Simulated Intergenerational Rank-Rank Correlation of Income

Figure A.2: Simulated Intergenerational Mobility of Income in the PSID Iowa-like sample as the rate of mismatch between fathers and sons varies

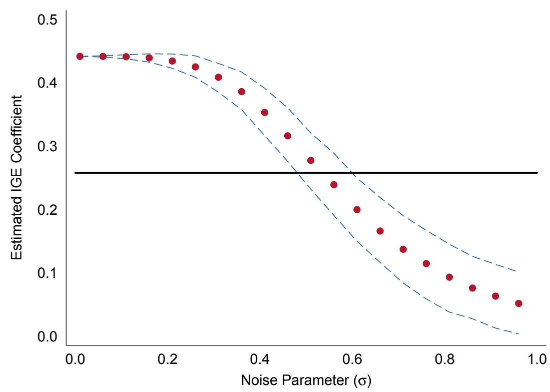


Figure A.3: Simulated Intergenerational Elasticity of Income in the PSID as the noise in earnings varies

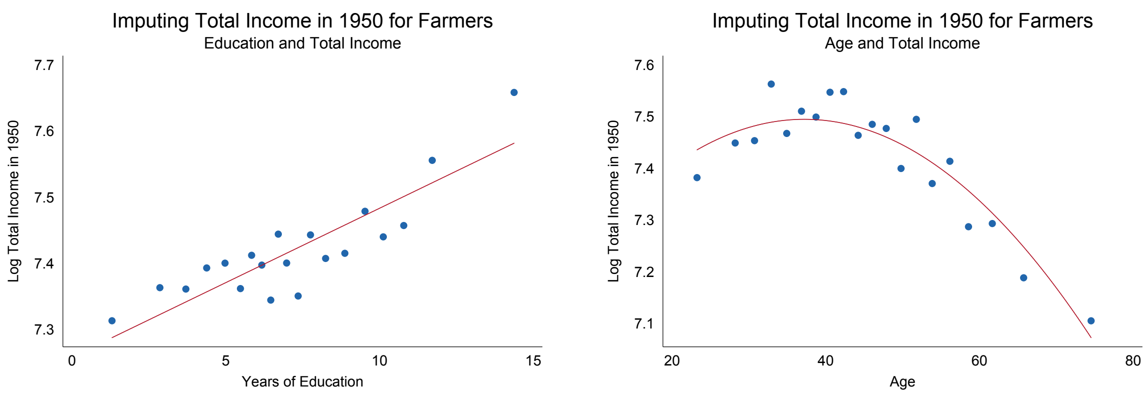
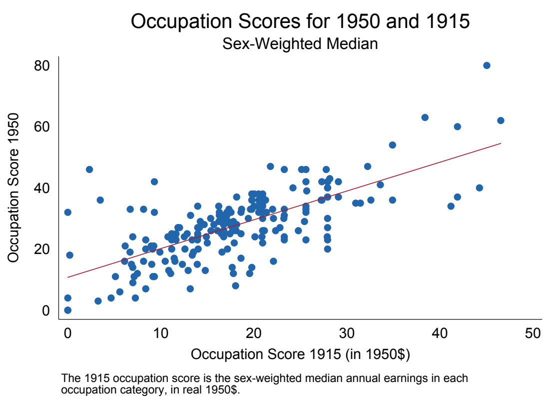


Figure A.4: Binscatter graphs presenting the correlation between years of education and age with log total income in 1950 for farmers from the IPUMS 1% sample. Both figures include controls for weeks worked, state of birth, state of residence, and education or age (when the variable is not on the x-axis). The slope in the figure on the left is approximately 0.0225.



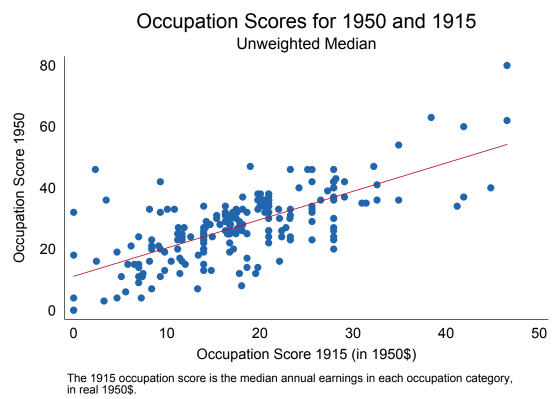


Figure A.5: Comparing Occupation Score Measures between 1915 and 1950

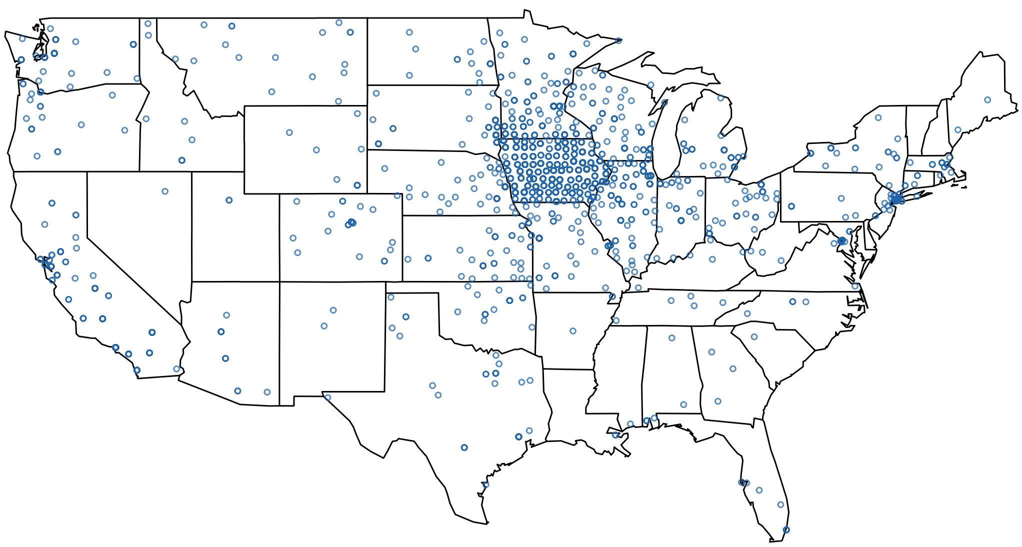


Figure A.6: Sons County of Residence in 1940. The darker symbols implies greater density of points at a given latitude and longitude.

Table A.1: Summary Statistics: Fathers in 1915 and Sons in 1915 and 1940

	Fathers		Sons	
	Fathers in Sample	Found Fathers	Sons in Sample	Sons Found
Yearly Earnings	1005.6	1007.1		1417.1
	(591.5)	(587.6)		(818.9)
Log Yearly Earnings	6.743	6.747		7.059
	(0.604)	(0.597)		(0.716)
Log Weekly Earnings	2.858	2.859		3.250
	(0.566)	(0.563)		(0.633)
Years of Education	8.491	8.507		10.40
	(2.837)	(2.803)		(3.068)
Age (1915)	41.92	41.73	9.740	9.605
	(9.262)	(9.307)	(4.350)	(4.357)
Born in Iowa	0.473	0.505	0.859	0.886
	(0.499)	(0.500)	(0.348)	(0.318)
Urban (1915)	0.452	0.436	0.419	0.415
	(0.498)	(0.496)	(0.493)	(0.493)
Observations	3713	2204	7580	3971

All summary statistics are based on those fathers and sons with complete data for all listed variables. The sample fathers include only men with sons between the ages of 3 and 17 in 1915. The found fathers are only those men with sons matched into the 1940 census. All sons includes any boys aged 3 to 17 in the Iowa sample in 1915; the found sons are only those boys linked from 1915 to 1940. For fathers, earnings, education, age, and urban status are measured in the 1915 Iowa State Census. For sons, earnings and education are measured in the 1940 Federal Census, while age and urban status are measured in the 1915 Iowa State Census.

Sources: 1915 Iowa State Census Sample; 1940 Federal Census

Table A.2: Sons of Iowa Residences in 1940

	Matched 1915-1940 Sample		1940 IPUMS Sample	
	Count	Share (%)	Count	Share (%)
Iowa	2858	63.9	1933	58.9
Illinois	301	6.7	163	5.0
California	292	6.5	221	6.7
Minnesota	177	4.0	159	4.8
Wisconsin	94	2.1	61	1.9
Nebraska	78	1.7	82	2.5
Missouri	66	1.5	71	2.2
South Dakota	61	1.4	70	2.1
New York	52	1.2	33	1.0
Other	497	11.1	487	14.8

This table compares the state residences of the sons matched between the 1915 Iowa State Census and the 1940 Federal Census with state residences of all men born in Iowa between 1898 and 1912 in the IPUMS 1% sample of the 1940 census.

Sources: 1915 Iowa State Census Sample; 1940 Federal Census

Table A.3: Census Matching Variables

	Probit	Logit
First and Last name match	0.632^{***}	1.129^{***}
	(0.086)	(0.168)
First name distance, Jaro-Winkler	-6.071^{***}	-11.543^{***}
	(0.525)	(0.994)
Last name distance, Jaro-Winkler	-10.285^{***}	-19.145^{***}
	(0.487)	(0.954)
Absolute Value Difference in Year of Birth is 1	-0.708^{***}	-1.308^{***}
	(0.044)	(0.083)
Absolute Value Difference in Year of Birth is 2	-1.562^{***}	-2.893^{***}
	(0.065)	(0.126)
Absolute Value Difference in Year of Birth is 3	-2.316^{***}	-4.370^{***}
	(0.102)	(0.208)
First name Soundex match	0.153^{***}	0.294^{***}
	(0.054)	(0.100)
Last name Soundex match	0.698^{***}	1.341^{***}
	(0.069)	(0.135)
Hits	-0.064^{***}	-0.123^{***}
	(0.002)	(0.005)
Hits-squared	0.0003^{***}	0.001^{***}
	(0.00002)	(0.00004)
More than one match for first and last name	-1.690^{***}	-3.217^{***}
	(0.093)	(0.183)
First letter of first name matches	0.871^{***}	1.593^{***}
	(0.130)	(0.245)
First letter of last name matches	0.886^{***}	2.003^{***}
	(0.148)	(0.356)
Last letter of first name matches	0.147^{***}	0.312^{***}
	(0.053)	(0.101)
Last letter of last name matches	0.649^{***}	1.239^{***}
	(0.070)	(0.139)
Middle Initial matches, if there is a middle initial	0.537^{***}	0.908^{***}
	(0.097)	(0.186)
Constant	-1.479^{***}	-3.087^{***}
	(0.225)	(0.480)
Observations	38,091	38,091
Log Likelihood	−2,440.877	−2,444.649
Akaike Inf. Crit.	4,915.753	4,923.298

The results of two different matching algorithms—a probit model and a logit model—trained on the 30% sample of Iowa sons link to the 1940 census, constructed manually by trained research assistants. Each observation is a possible link between a son in 1915 and records in 1940. I use the probit model to generate a score for each possible link. Possible links are coded as actual links if (1) the score is the top score for the given son in 1915, (2) the score is larger than 0.14, and (3) the ratio of the top score to the second best score is larger than 1.2. These parameters were chosen to maximise the accuracy and efficiency of the model through cross-validation. For more details, including tests of the algorithm against other machine learning procedures, please see `http://jamesfeigenbaum.github.io/research/census-link-ml.html`. If there are any sons in 1915 with multiple exact matches in 1940—that is exactly the same first name string, last name string, and year of birth—then I am unable to pick between these possible matches. All possible matches are equally as likely to be the true match. Instead, I score any record links of this type with failure it is not used directly in the prediction algorithm.

Table A.4: Applying Family Relationship Imputation Devised for 1915 Iowa State Census Sample to the 1910 and 1920 IPUMS samples from Iowa

	True Census Relation				
	child	father	mother	other	Total
child	18274	17	2	574	18867
father	38	6396	0	41	6475
mother	40	0	6527	87	6654
other	149	70	47	803	1069
Total	18501	6483	6576	1505	33065

Sources: 1910 and 1920 IPUMS samples (Iowa only); 1915 Iowa State Census Sample

Table A.5: Intergenerational Mobility Estimates: Alternative Function Forms

	Specification				
	(1)	(2)	(3)	Observations	Clusters
A. Intergenerational Elasticity (IGE), Log Weekly Earnings					
Full Sample	0.207	0.194	0.266	2037	1666
	(0.032)	(0.032)	(0.084)		
Urban Sample	0.286	0.280	0.311	1004	824
	(0.046)	(0.052)	(0.105)		
Rural Sample	0.147	0.161	0.244	1033	842
	(0.040)	(0.041)	(0.116)		
B. Annual Income in Levels					
Full Sample	0.318	0.319	0.438	2114	1730
	(0.044)	(0.042)	(0.104)		
Urban Sample	0.541	0.548	0.558	1025	842
	(0.082)	(0.084)	(0.208)		
Rural Sample	0.192	0.202	0.352	1089	888
	(0.045)	(0.045)	(0.107)		
C. Square Root of Annual Income					
Full Sample	0.189	0.184	0.251	2114	1730
	(0.031)	(0.030)	(0.069)		
Urban Sample	0.298	0.298	0.286	1025	842
	(0.057)	(0.058)	(0.141)		
Rural Sample	0.119	0.131	0.217	1089	888
	(0.035)	(0.035)	(0.076)		
D. Log Family Annual Income					
Full Sample	0.208	0.195	0.260	1955	1595
	(0.032)	(0.032)	(0.085)		
Urban Sample	0.280	0.265	0.292	946	774
	(0.046)	(0.052)	(0.109)		
Rural Sample	0.157	0.168	0.246	1009	821
	(0.041)	(0.041)	(0.118)		
E. Log Years of Education					
Full Sample	0.186	0.170	0.100	2435	1958
	(0.020)	(0.020)	(0.049)		
Urban Sample	0.200	0.179	0.235	1107	900
	(0.034)	(0.034)	(0.060)		
Rural Sample	0.175	0.166	0.027	1328	1058
	(0.025)	(0.026)	(0.058)		

Standard errors clustered by family in all regressions. In Panel A, son's weekly log earnings in 1940 is the dependent variable. In Panel B, son's annual income (in levels not logs) is the dependent variable. In Panel C, the square root of the son's annual income is the dependent variable. In Panel D, the dependent variable is the son's individual earnings in 1940 in logs; the key independent variable is the log of 1915 family income, rather than father's income. In Panel E, educations is measured as the log of years of completed schooling. Specification 1 is a univariate regression of son's outcome on father's outcome. Specification 2 adds name string controls, 1915 county fixed effects, and quartic controls in father and son age. Specification 3 adds an interaction between father's outcome and son's normalised age. Name string controls: first and last name commonness, length, letter similarity, and Scrabble scores. Son's ages are normalised relative to age 40 in 1940.

Sources: 1915 Iowa State Census Sample; 1940 Federal Census

Table A.6: Geographic Mobility: Miles Moved 1915 to 1940

	Full Sample			Urban		Rural	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Father Earnings	0.0487		-0.00830	0.252^{*}		-0.0999	
	(0.0616)		(0.0638)	(0.130)		(0.0738)	
Father Education		0.0577^{***}	0.0477^{***}	0.0419	0.0709^{***}	0.0416^{**}	0.0496^{***}
		(0.0132)	(0.0145)	(0.0267)	(0.0227)	(0.0172)	(0.0161)
Son Age Quartic	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Father Age Quartic	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Name String Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	3655	3935	3649	1479	1563	2170	2372
Number of Clusters	2628	2835	2623	1141	1207	1482	1628
R-squared	0.0255	0.0293	0.0295	0.0467	0.0448	0.0184	0.0158

The log of 1 + miles moved by the son from 1915 to 1940 is the dependent variable. It is necessary to add one to the number of miles to avoid dropping sons who did not move counties between 1915 and 1940 from the analysis. Standard errors clustered by family. Name string controls: first and last name commonness, length, letter similarity, and Scrabble scores. Son's ages are normalised relative to age 40 in 1940.

Sources: 1915 Iowa State Census Sample; 1940 Federal Census