A New Old Measure of Intergenerational Mobility: Iowa 1915 to 1940*

James J. Feigenbaum†

February 20, 2015

Abstract

Was economic mobility high during the first half of the twentieth century in the United States? I combine two historical data sources to estimate intergenerational income mobility between 1915 and 1940. 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. In my sample of fathers and sons, I estimate a lower intergenerational elasticity of income than is found in modern studies of the United States, suggesting higher levels of income mobility. Income mobility measured with relative income ranks also show higher mobility historically. Intergenerational mobility of education is higher in my sample than in modern measures as well. I find sons in rural counties in 1915 to have more mobility of both income and education than urban sons. Lacking data on income, past studies of historical intergenerational mobility have relied on occupation transition data for fathers and sons to measure mobility. When I compute standard measures of occupational mobility for my sample, I find levels of mobility between 1915 and 1940 to be larger than modern estimates, confirming the higher mobility I find in income measures. This suggests that the standard estimates of historical occupational mobility may be accurate substitutes for measures of income mobility when income data does not exist.

*I thank Claudia Goldin, Lawrence Katz, Rick Hornbeck, Nathan Nunn, 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, 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).

†PhD Candidate, Department of Economics, Harvard University, email: jfeigenb@fas.harvard.edu, site: scholar.harvard.edu/jfeigenbaum.

1 Introduction

The history of income inequality throughout the twentieth century is well known (Piketty and Saez 2003). Much less, however, is known about economic mobility in the United States in the first half of the twentieth century. How strong was the link between a child's outcomes in adulthood and the accident of his or her birth? And how does economic mobility in this earlier period 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? The concepts of inequality and intergenerational mobility are strongly linked, but inequality does not determine mobility, or vice versa, and so few clues are available in the inequality literature. The recent landmark studies on trends in intergenerational mobility are unable to trace income mobility before the 1980s (Lee and Solon 2009; Chetty et al. 2014b). Moreover, historical analysis of economic mobility has long been limited by the available data sources. 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. Historians have collected detailed place-specific data on intergenerational mobility and transfers (see Thernstrom 1964, 1973, on Newburyport, MA and Boston, for example), but these studies are often constrained by their inability to track those individuals who moved away from the original study site.

To measure economic mobility in 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. I estimate intergenerational mobility along three dimensions: income, education, and occupation. The estimates I present are the earliest intergenerational mobility parameters for both income and education in the United States.[1] I use these data to address the question of how intergenerational mobility changed both between 1915 and 1940 in the United States, as well as between 1915 and the present. In addition, because my sample includes intergenerational data on income, education, and occupation for the same individuals, I can determine whether these measures of intergenerational mobility all show consistent trends.

Table 1 summarises my primary results. I find a lower intergenerational income elasticity during the first half of the twentieth century in the US than modern studies find in the second half of the century. This result implies that there was more mobility of income during my study period than there is today. I also measure intergenerational mobility using the rank-rank parameter (Chetty et al. 2014a) and similarly find more mobility historically. Such differences between modern and historical mobility could be spurious,

[1] Parman (2011) also draws on the 1915 Iowa State Census to measure intergenerational mobility. However, data constraints limit the broad interpretability of his results. He was only able to match fathers to sons within 1915 Iowa. Thus his estimate is biased by omitting any sons who move out of the state. Even more problematic, the average age of the fathers in his sample is between 57 and 65 and his sons are between 25 and 30. Corak (2006) shows that mobility parameters estimated with such old fathers and young sons are biased down to a large degree. These points are addressed in more detail later in this paper.

Table 1: Intergenerational Mobility Results Summary

Intergenerational Mobility Measure	Estimates: 1915 to 1940	Estimates: Modern	Modern Source
Intergenerational Elasticity of Income	0.249	0.36 to 0.54	Lee and Solon (2009)
Income Rank-Rank Coefficient	0.210	0.307 to 0.317	Chetty et al. (2014)
Educational Persistence	0.187	0.46	Hertz et al. (2007)
Occupation Score Elasticity (1915 Basis)	0.234	.	
Occupation Score Elasticity (1950 Basis)	0.391	.	
Altham-Ferrie Occupation Transition Statistic	16.03	20.76	Ferrie (2005)

All measures of intergenerational mobility will be explained in detail in the text of this paper. Across all measures, higher estimates imply less mobility. The intergenerational elasticity of income is the regression coefficient on log of father's annual income, with son's annual income as the dependent variable. The income rank-rank coefficient is the regression coefficient on the father's income percentile, with the son's income percentile as the dependent variable. Educational persistence is the the regression coefficient on the father's years of education, with the son's years of education as the dependent variable. Occupation score elasticity is the regression coefficient on the father's occupation score, with the son's occupation score as the dependent variable, both in logs. The occupation scores are defined as the median income across all respondents in a given occupation in a given base year. The Altham-Ferrie occupation transition statistic relates the distance from a given occupation category transition matrix to the complete mobility matrix.

Sources: 1915 Iowa State Census Sample; 1940 Federal Census; Lee and Solon (2009); Chetty et al. (2014a); Hertz et al. (2007); Ferrie (2005)

driven by measurement error or differences in sample construction. However, I show that the estimated differences between modern and historical mobility remain large after adjusting the modern sample to mirror the historical sample in measurement noise and demographic and geographic composition.

The results for education are broadly similar: there was more mobility in education in the early twentieth century than today as well. This is also the case for occupational mobility when measured with the standard transition matrices. My results indicate that mobility is higher for the early twentieth century than just after mid-century; I also find less occupational mobility during the twentieth century than others have found for the nineteenth century.[2]

Overall, the various intergenerational mobility measures point to one, main conclusion: there was more economic mobility in the early 20th century than there is today.

The paper proceeds as follows. In the second section, I discuss the historical data that I draw on and my data collection and census-linking procedures. In section three, I review past measurements of intergenerational mobility, both in modern and historic samples. In particular, I focus on sources of measurement error that may bias the estimates of mobility up or down in historic data relative to modern data. In

[2] This point contrasts somewhat with findings in the modern intergenerational mobility literature. Jencks and Tach (2006) suggest that intergenerational correlations of earnings and occupational rank are not good substitutes. They note, in particular, that in the US earnings correlation is higher than in other rich democracies but occupational rank correlation is low relative to such countries. For historical study, I find that occupational and income mobility measures are relatively similar.

the fourth section, I present my estimates of intergenerational mobility in the early twentieth century for income, education, and occupation, and compare these results to estimates for the modern period. I also consider heterogeneity in the mobility parameters across my sample as well as geographic mobility. Section five concludes the paper.

2 Data

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. The 1915 Iowa Census was a complete survey of all 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.[3] 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.[4]

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, 6,071 of whom have fathers in their households and the requisite data on both the father's education and income. Each of those 7,580 observations is a son in 1915 Iowa.

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.[5] Once

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

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

[5]I generate the training data used to train the algorithm by manually comparing a subset of sons from 1915 to the set

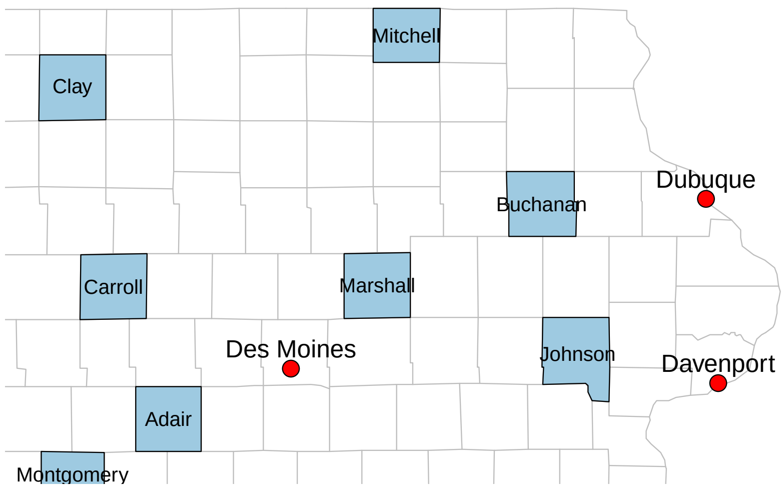


Figure 1: Map of Iowa 1915 Cities and County Sample

the matched sons are identified, I record the pertinent data from the 1940 records. The 1940 Census was the first federal census to collect data on incomes, weeks of work, and years of education of the entire population.[6] 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 is in line with the rates of previous literature linking between censuses.[7] Table 2 shows my match rates for the rural and urban samples; match rates are comparable between the two samples. My sample size of 4,478 father-son pairs is also comparable to many other projects measuring intergenerational mobility, both historically (Long and Ferrie 2013) and recently (Lee

of possible links in 1940 and determining which records are in fact matches for the same person. The match algorithm is used to reduce between-researcher variability in match quality, to speed up the matching process, and ensure data replication. 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. 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%. For a detailed description of the matching algorithm, see a technical write up on my website: http://scholar.harvard.edu/jfeigenbaum/publications/automated-census-record-linking.

[6]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 is 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 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 5, 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.

[7]Parman (2011) reports match rates of just below 50%. Guest et al. (1989) match at 39.4%.

and Solon 2009).

Table 2: Sample Match Rates

	Rural	Urban	Total
Found	2657	1821	4478
	(60.36)	(57.30)	(59.08)
Total	4402	3178	7580
	(100.00)	(100.00)	(100.00)

Any project linking historical data is subject to possible bias due to the difficulty of making matches between datasets. I present an analysis of potential bias in the matching below, which suggests that the final sample does not suffer any crucial construction defects. 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.

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.[8] 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 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 secularly more likely to be mis-transcribed than others: s is quite visually unique while l and n are both visually similar to other letters.[9] 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.[10] I use a matrix of letter visual similarity from Simpson et al. (2013) to compute, for first and last names, a similarity score.[11]

[8] 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.

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

[10] Recall matches are made using census indices transcribed by Ancestry.com and deposited with the NBER.

[11] 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.

Table 3: 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)
Normalized letter similarity score, first name			0.019***		0.024***
			(0.007)		(0.007)
Normalized letter similarity score, last name			0.006		0.005
			(0.007)		(0.007)
Normalized scrabble score, first name				−0.001	−0.002
				(0.006)	(0.007)
Normalized 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).

Sources: 1915 Iowa State Census Sample; 1940 Federal Census

Finally, I calculate a name's Scrabble score as an alternative measure of both name commonness and name simplicity.[12] 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). I use standardised z-scores 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.

Table 3 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.[13] I include controls for all of these name string properties in all subsequent analysis.

[12]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 20th century. Scrabble point values were based, originally, on the frequency of letters on US newspaper front pages.

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

Table 4: Effects of Family Covariates on the Probability of Matching Records from 1915 to 1940

X	β	SE	Predicted Match Rate with X at 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 3, there are 7580 observations and 4731 clusters, clustering standard errors by family.

Sources: 1915 Iowa State Census Sample; 1940 Federal Census

More serious issues could be generated by differential matching rates according to father, son, or family characteristics in 1915. In Table 4, 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.[14] 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 points 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. Confirming the results presented in Table 2, I am also less likely to match sons in the urban sample. I am also more likely to link sons born in Iowa, even after conditioning on name string characteristics.[15] 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.

The first two columns of Table 5 present summary statistics for the fathers of children between 3 and 17 in the Goldin-Katz Iowa State Census sample. Observation counts are smaller than my full sample because fathers of multiple children or sons are not double counted. The fathers in the sample are restricted to fathers of sons between 3 and 17; fathers found are the father 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

[14] 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.

[15] 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.

Table 5: 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		1358.1
	(591.5)	(587.6)		(931.2)
Log Yearly Earnings	6.743	6.747		6.955
	(0.604)	(0.597)		(0.802)
Log Weekly Earnings	2.858	2.859		3.183
	(0.566)	(0.563)		(0.708)
Years of Education	8.491	8.507		10.61
	(2.837)	(2.803)		(3.092)
Age (1915)	41.92	41.73	9.740	9.365
	(9.262)	(9.307)	(4.350)	(4.353)
Born in Iowa	0.473	0.505	0.859	0.869
	(0.499)	(0.500)	(0.348)	(0.337)
Urban (1915)	0.452	0.436	0.419	0.490
	(0.498)	(0.496)	(0.493)	(0.500)
Observations	3713	2204	7580	2940

All summary statistics are based on those fathers and sons with complete data for all listed variables. This restriction reduces the number of observations from the count of all sons found, as presented in Table 2. 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

Census earned very slightly more, though not significantly so, measured either in levels, logs, and weekly earnings.

The final two columns of Table 5 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 3,284. The located sons earned nearly $1400 in 1940, which is lower in real terms than the average earnings for their fathers in 1915, likely due to the lingering effects of the Great Depression.[16] Also notable in the summary statistics is the fact that the sons had on average two more years of schooling than their fathers. 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), Parman (2011), and others.

How does my sample compare to the rest of the US in 1915 or in 1940? In Table 6, I compare the sons in my sample to the national population in 1940, focusing on men aged 28 to 42. The second column is the

[16] 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.

Table 6: Summary Statistics: Sons in 1940

	Iowa Linked Sample	1940 IPUMS Sample	
	Sons Found	Unweighted	Weighted by State of Birth
Yearly Earnings	1358.1	1237.2	1256.4
	(931.2)	(889.4)	(898.0)
Log Yearly Earnings	6.955	6.819	6.851
	(0.802)	(0.887)	(0.863)
Log Weekly Earnings	3.183	3.102	3.100
	(0.708)	(0.758)	(0.757)
Years of Education	10.61	9.151	10.29
	(3.092)	(3.482)	(2.980)
Age (1940)	34.31	34.65	34.76
	(4.409)	(4.337)	(4.377)
Married	0.804	0.804	0.815
	(0.397)	(0.397)	(0.389)
Born in Iowa	0.869	0.0190	0.869
	(0.337)	(0.136)	(0.337)
Share White	0.993	0.902	0.991
	(0.0822)	(0.297)	(0.0922)
Observations	2940	112309	99105

All variables measured in 1940. After weighing the 1940 1% IPUMS sample of the census for state of birth, the matched sample of sons is comparable to the 1940 population of men between 28 and 42.

Sources: 1940 Federal Census

IPUMS 1% sample of the 1940 census; the third column reweights the 1% sample by state of birth to match the states of birth in my Iowa sons sample. I find that my sample has slightly higher income than either the general population or the reweighted sample. After reweighting, however, my sample is representative in terms of education, age, marital status, and race.[17]

3 Past Estimates of Intergenerational Mobility

3.1 Intergenerational Income Mobility

The most frequent measure of intergenerational economic mobility used in the literature 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 modern IGE literature.[18] 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

[17]In a later analysis of geographic mobility, I show in Table 12 that my sample's state of residence distribution in 1940 is roughly similar to the geographic distribution of all men born in Iowa between 1898 and 1912.

[18]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.

establish a correct measure of IGE well before the period previously studied in this literature.[19]

Corak's preferred measure of IGE in the US is 0.47,[20] in line with the reviews presented by both Solon and Black and Devereaux. Corak (2006) also documents large variations between US studies measuring the intergenerational elasticity of income.[21] 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.[22] 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.[23] Previously, I compared my estimated IGE parameter for income to the results in Lee and Solon's paper (see Table 1), showing that my estimates of mobility for the early twentieth century were higher than the modern estimates.

When researchers estimate the IGE, they impose a very particular functional form relationship between son's and father's earnings.[24] Chetty et al. (2014a) present evidence from modern 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 father's log income and 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 modern 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

[19] 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 same race men in the son's state of birth. While that is a possible proxy for father's income, it does not seem sufficiently detailed or granular to detect small shifts in the intergenerational transmission of income.

[20] Corak also gives lower and upper bounds of between 0.40 and 0.52.

[21] 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 son's income at age 30 on a three year average of 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.

[22] I will 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.

[23] Lee and Solon (2009) do find a trend in intergenerational mobility for daughters.

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

their sample farther back in time.

I address several measurement issues in order to produce unbiased and consistent estimates of mobility in my historical sample.[25] The first is simple measurement error of the income variables. Solon (1989) presents the classic treatment of this bias. One would like to estimate β, the intergenerational elasticity of income:

$$y_i = \beta X_i + \epsilon_i$$

where y_i is the child's adult income and X_i is the parent's adult income. The ideal measurements of y_i and X_i would be permanent income.[26] However, most studies, including mine, observe only a single year of earnings for both parents and children. Classic measurement error arises because a single year of income is a noisy proxy for permanent income. As with any classical measurement error, this biases coefficients towards zero. Drawing on notation from both Solon (1989) and Solon (1999), let the noisy signal of permanent income be:

$$\begin{aligned} y_{it} &= y_i + w_{it} \\ X_{it} &= X_i + w_{is} \end{aligned}$$

with observations of income of the parent in year t and income of the child in year s. Rearranging the above expressions and running the regression suggested by the definition of β yields a $\hat{\beta}$ that is biased:

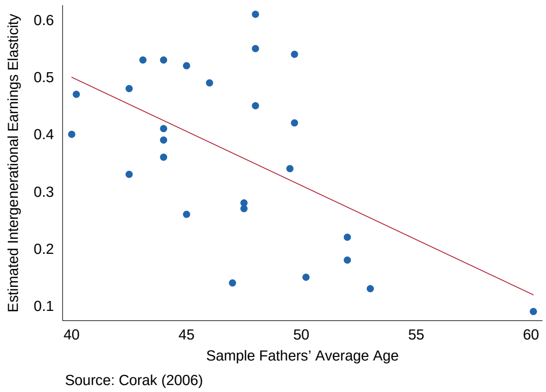
$$plim\hat{\beta} = \beta \frac{\sigma_x^2}{(\sigma_x^2 + \sigma_s^2)} < \beta$$

Solon (1989) also argues that more homogenous samples (relative to the study population) will tend to bias down IGE results. Though Iowa in 1915 was relatively representative of the nation, it is far more homogenous in terms of race than the country is today. Although it is impossible make my sample more racially heterogeneous, I can construct a racially homogenous sample using modern data and compare it to my earlier Iowa samples.

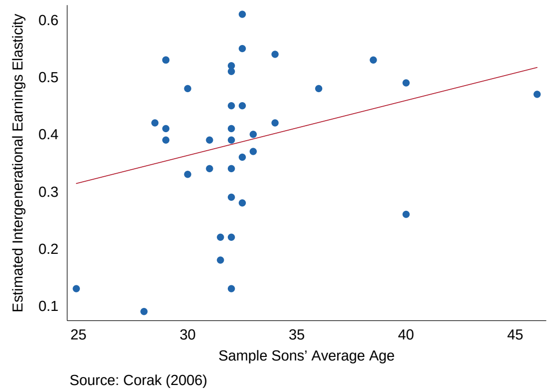
Haider and Solon (2006) raise a more complex problem owing to trends in life cycle earnings. The literature on life cycle earnings suggests that people with higher permanent income experience more earnings growth earlier in their career. Rather than treating current annual income as a noisy signal of permanent

[25]Though the literature has considered these measurement issues with respect to calculating the IGE parameter, in theory these same problems could bias the rank-rank parameter as well. However, Chetty et al. (2014a) document the general robustness of the rank-rank estimtes to single year samples, variation in father and son ages, and included control variables.

[26]Even standard permanent income may be the wrong measure. The parental income that matters for children's outcomes may be income while the child is in utero, young, or in school, depending on one's reading of the critical years literature.



(a) Correlation between estimated IGE and sample fathers' ages from literature reviewed by Corak (2006)



(b) Correlation between estimated IGE and sample sons' ages from literature reviewed by Corak (2006)

Figure 2

income, suppose there is a simple linear relationship between current and permanent income:

$$y_{it} = \lambda_t y_i + v_{it}$$
$$X_{is} = \lambda_s X_i + v_{is}$$

If there is only measurement error on the left hand side (that is, the parent's permanent income is observed), then $plim\hat{\beta} = \lambda_t \beta$. If $\lambda_t \neq 1$, as in the textbook case of measurement error, $\hat{\beta}$ will be inconsistent as well, with the inconsistency a function of age at the time income is observed. If, however, there is measurement error only on the right hand side (that is, the child's permanent income is observed but not the parent's), Haider and Solon show that the probability limit of β will be $\theta_s \beta$, where $\theta_s = \frac{\lambda_s \sigma_x^2}{\lambda_s^2 \sigma_x^2 + \sigma_s^2}$ is the inconsistency factor or the reliability ratio. Depending on the relative sizes of λ_s and the variances, this can amplify or attenuate the estimate of β.

One particular consequence of this severe bias is shown in Figure 2, based on the American IGE literature surveyed by Corak (2006). As the figure on the left shows, the older the average age of the fathers in the study samples, the lower the estimated intergenerational elasticity. The figure on the right shows a similar but weaker relationship holding in the opposite direction between estimated IGEs and the average age of sons in the sample.[27]

Parman (2011) also draws on data from the 1915 Iowa State Census. He matches adult men in the Iowa sample backwards in time to the 1900 Federal Census to construct childhood households. Though most

[27] Both of these best fit lines are statistically significant in the univariate regression, but clearly 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.

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.[28]

However, Parman's results are constrained by data limitations in two important ways.[29] First, all income data in Parman's study are drawn from a single state census in one year. Any fathers or sons leaving the state between 1900 and 1915 are omitted from the final dataset. The direction and magnitude of this bias is *a priori* unclear. If the moving sons are more likely to be unlike their fathers, father-son pairs with weaker relationships between their outcomes will be removed from the sample, thereby biasing up the IGE estimates. In this case, the early twentieth century may have been even more mobile than Parman suggests. Alternatively, the selection of which sons leave the state may make the final sample more homogenous; following the arguments in Solon (1989), this biases IGE estimates downward. Generally, the uncertain selection of out of state movers in Parman's sample complicates the interpretation of his results.

Figure 2 demonstrates the other major constraint on the interpretation of the IGE results in Parman (2011). 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 very likely to present a very low IGE, due to life-cycle-induced measurement errors (see Haider and Solon (2006) for the detailed econometric treatment of this issue). The average age of the sons in Parman's sample is between 25 and 30, and this may also, to an extent, bias his results towards a very low IGE. However, the strength of the bias based on the relationship presented in Figure 2b is less clear.

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.

3.2 Intergenerational Education Mobility

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).[30]

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

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

[30] 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.

Due to data constraints, there has been little work on educational mobility in the US historically.[31] Outside of the US, Checchi et al. (2008) 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 modern period.

Though measurement error is a concern for education, as it was with income, the specific concerns are quite different. Education observed in a given year during adulthood is a much better proxy for permanent education.[32] However, other measurement error issues might include faulty reporting of years completed and artificial heaping at milestone numbers (four years of high school or college, for example). Kane et al. (1999) have shown how non-classical measurement error in years of education can lead to potential biases of Mincerian regressions on the return to education.

The inability to measure only quantity rather than quality of education is also a potential issue. It may be the case that years of education is simply a noisy signal of true education. Any noise will bias the estimated mobility coefficient downwards, but to the extent that education quantity is always a noisy measure of education quality, this does not present a strong challenge to comparing results over different time periods.

3.3 Intergenerational Occupational Mobility

Due to data limitations, the study of historical intergenerational mobility has focused on the study of occupational mobility. Early work on this topic is 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) begins the modern literature in sociology on intergenerational mobility. He 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 19th 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 19th 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,

[31] Parman (2011) measures the effects of public education on income mobility, but does not estimate father to son educational mobility directly.

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

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 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:

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

Altham and Ferrie (2007) use the $d(P,Q)$ notation to convey the sense in which the Altham statistics are distance measures.[33] $d(P,I)$, 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.[34] Comparison of mobility measures across time is complicated—perhaps even driven—by the secular 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 19th century, but they suggest that this reflects the high-heritability of farming and the declining shares of farmers since the late 19th 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 modern 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 father and son

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

[34] 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.

from farming to another occupational category may represent an increase or a decrease in income. Further, to what extent is the wide variation in 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.

4 Results

In Figure 3, I present the raw correlations between fathers' and sons' outcomes: log annual income (3a), years of education (3b), and occupation score based on 1950 scores (3c) and on 1915 scores (3d).[35] Quite clearly, there is a strong positive relationship between outcomes for fathers in 1915 and sons in 1915, but the exact measure of the respective slopes of these lines—and how those slopes compare with the estimates of mobility from modern studies—is the key question of this paper.

4.1 Intergenerational Mobility of Income

I measure mobility 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:

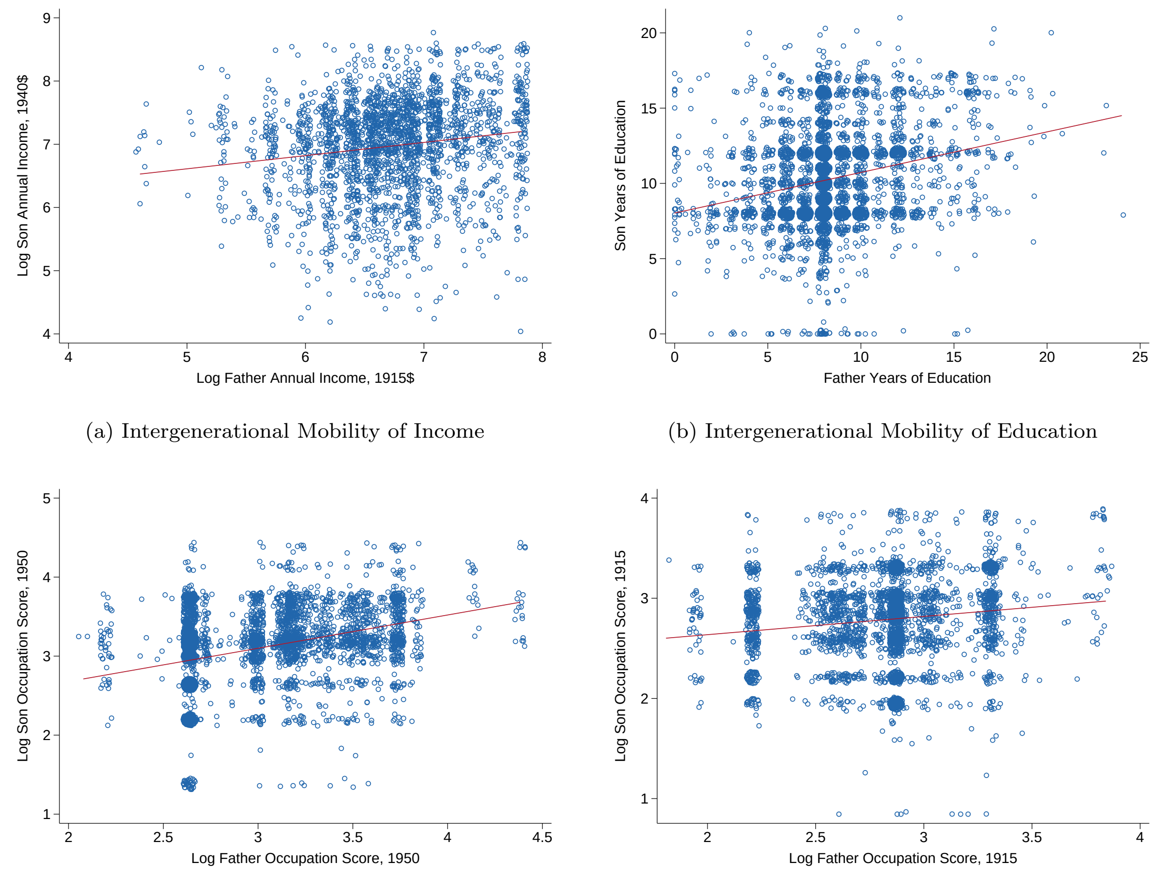
$$Y^s_{i,1940} = \alpha + \beta \cdot Y^f_{i,1915} + Q^s(age^s_{i,1915}) + Q^f(age^f_{i,1915}) + Q^s(age^s_{i,1915}) \times Y^f_{i,1915} + \epsilon_i$$

Larger estimates of the IGE parameter mean a tighter link between father and son and thus less mobility.[36]

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 is 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

[35]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.

[36]The literature often defines $1 - \beta$ as the mobility parameter.



(a) Intergenerational Mobility of Income

(b) Intergenerational Mobility of Education

(c) Intergenerational Mobility of Occupation, 1950 Basis

(d) Intergenerational Mobility of Occupation, 1915 Basis

Figure 3

range as the fathers in my sample. Ranks are scaled as percentiles between 0 and 1; a rank of .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, 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).[37] 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 (and therefore have the same father data), I cluster standard errors at the family level. I also include county fixed effects (which subsumes a control for urban or rural sample). The results are robust to the inclusion of controls for family-size effects, county fixed effects, and the name string control variables described previously.[38]

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) show that estimates are stable even with just one year of income observed for both fathers and sons. 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.

Panel A of Table 7 presents my estimates of the IGE of income across a variety of samples. Both father's and son's incomes are measured as annual log earnings.[39] The first specification is a simple univariate regression of son's log earnings on 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 normalised age and father log earnings to control for lifecycle measurement error (Haider and Solon 2006).

The first row of Table 7 presents the baseline estimates for the IGE parameter for the full sample of Iowa fathers and sons. The intergenerational mobility coefficient ranges from 0.191 to 0.249. The literature suggests an IGE of 0.47 for income in the United States today (Corak 2006). Lee and Solon (2009) argue

[37]With this normalization, the estimated β represents the relationship between son's and father's outcomes when the son is age 40. I follow Lee and Solon (2009) in normalizing to 40.

[38]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.

[39]To ensure comparability with modern estimates, I use annual earnings, not weekly earnings. Results using weekly earnings are similar and in fact lower than those presented in Table 7, suggesting even more mobility in the early twentieth century.

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.[40]

I present the rank-rank mobility estimates in Panel B of Table 7. The rank-rank parameter ranges from 0.163 to 0.210. 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. Similar to my IGE results, I find much more income mobility historically than today.

However, any measurement error will tend to bias down estimates of intergenerational mobility (Solon 1999). Further, though Iowa is broadly representative of the US in 1915 (Goldin and Katz 2000, 2008), the differences in my estimated IGE may reflect differences between Iowa and the rest of the country, not differences between time periods. In fact, according to modern data, children born in Iowa are among the most economically mobile in the entire country, across many measures (Chetty et al. 2013). To account for these threats to my analysis, I construct a sample of modern 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 (99.3% of my linked Iowa sample is white). I also limit the PSID to sons who grew up in the Midwest.[41] The results are presented in the second rows of each panel in Table 7 (Panel A for IGE, Panel B for rank-rank).

To calculate a comparable modern 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.[42] 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$ and logged, so as to interpret the estimated coefficients as intergenerational elasticities.[43] 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 and age controls.[44] Though the Iowa-like samples in the PSID are quite small, the results suggest that the lower mobility I find historically is driven neither by the demographic composition of my data nor

[40]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 son's income) than smaller changes farther 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.

[41]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.

[42]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.

[43]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.

[44]Given the litany of measurement concerns in the IGE literature, Specification 3, which includes the recommended controls, is likely the best measure of the Midwestern IGE.

Table 7: Intergenerational Mobility Estimates

	Specification (1)	(2)	(3)	Observations	Clusters
A. Intergenerational Elasticity (IGE)					
Full Sample	0.209	0.199	0.258	2041	1669
	(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	0.168	0.233	1037	845
	(0.040)	(0.041)	(0.113)		
Excluding Sons of Farmers	0.302	0.259	0.391	1454	1201
	(0.037)	(0.038)	(0.095)		
Including Sons with Imputed Income	0.209	0.199	0.258	2041	1669
	(0.032)	(0.031)	(0.081)		
Sons Remaining in Iowa	0.148	0.148	0.173	1185	1002
	(0.043)	(0.043)	(0.121)		
B. Intergenerational Rank Rank Parameter					
Full Sample	0.175	0.169	0.219	2041	1669
	(0.022)	(0.021)	(0.046)		
PSID Iowa-Like Sample	0.258	0.240	0.323	3680	356
	(0.049)	(0.064)	(0.084)		
Urban Sample	0.221	0.211	0.219	1004	824
	(0.032)	(0.034)	(0.071)		
Rural Sample	0.142	0.151	0.224	1037	845
	(0.028)	(0.028)	(0.061)		
Excluding Sons of Farmers	0.233	0.203	0.258	1454	1201
	(0.026)	(0.027)	(0.058)		
Including Sons with Imputed Income	0.175	0.169	0.219	2041	1669
	(0.022)	(0.021)	(0.046)		
Sons Remaining in Iowa	0.128	0.134	0.187	1185	1002
	(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

by the single year measurements of income.

In the second row of Panel B of Table 7, I measure the rank-rank mobility using the PSID sample. For the Midwest sample, I measure a lower parameter than is found nationally; however, these estimates are far larger than what I find historically in the full sample. For an alternative construction of a comparable modern rank-rank parameter, I use the county level results reported by Chetty et al. (2014a) in Online Data Table 3. 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 modern mobility parameter of 0.31, similar to the result from the modern PSID data and, more importantly, far larger than the rank-rank parameter of 0.163 to 0.210 that I find in my historical sample.[45]

In the appendix, I test the degree to which either false matches in my linking procedure between censuses or 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 modern sample. I introduce both mismatches and measurement error into my Iowa-like PSID sample considered above. Simulation tests suggest that neither source of error is likely to account for the differences in estimated IGEs. 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.[46] 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 earnings measured in the modern period to generate the large difference in IGEs.

Are the higher rates of mobility that I find historically driven by the secular movement off the farm in the early 20th century?[47] 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.[48] The results for these subsample analyses are presented in the third and fourth rows of Table 7. Only 15 of the urban sons has a father farmer in 1915 and only 65 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 modern data. The much higher levels of mobility for rural sons may be driven by the large increases in access to public

[45]The exact weighted average of the modern data is 0.3097. I can also split the sample between the urban and rural counties in my analysis. The weighted average of rank-rank mobility is 0.3538 in the three urban counties and 0.2714 in the 10 rural counties.

[46]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%.

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

[48]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.

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 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 panel A and B to only sons with fathers who were not farmers.[49] Again, mobility is lower than the modern 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, the difficulty of farmers to distinguish between net and gross income in census responses, or transitory income shocks (such as adverse weather or crop-destroying pests).

As noted previously, the 1940 census collected data only for wage and salary workers and not capital income.[50] Thus, it is impossible to include sons in 1940 who were either farmers or business owners; this excludes data from 13.7% of the sons in 1940 who were farm owners or operators without income. These observations are not included in the measures of mobility previously discussed in Table 7. However, in row 6, 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 higher levels of mobility than in my main results.[51]

Overall, my estimates suggest more mobility historically than today, measured both by IGE and with the rank-rank parameter. However, they are not as low as the results presented in Parman (2011). One key shortcoming of the data considered in that analysis—a sample with very old fathers and very young sons—was highlighted previously. Another data-driven limitation is that the Parman (2011) sample can only match fathers to sons who still live in Iowa as adults.[52] How large is the bias of this restriction and in what direction does it push the intergenerational income mobiility results? I 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. In the final rows of both panels A and B in Table 7, I calculate IGE and rank-rank parameters for this subset of sons 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. Because the state of residence in 1940 is, in part, jointly determined with

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

[50]The data collected is 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.

[51]For details on the imputation of capital income in 1940, see Appendix 5.

[52]Again, this is due to the fact that the 1940 Federal Census was not available until 2012, 72 years after the survey was originally conducted and Parman (2011) instead matched fathers and sons within the 1915 Iowa State Census.

Table 8: Income Rank Quintile Transition Matrix

	Father's Quintile					
Son's Quintile	1	2	3	4	5	Total
1	15.1	7.9	8.9	9.8	7.2	9.4
2	23.1	22.9	15.6	18.0	15.1	18.9
3	27.8	24.7	26.1	18.3	18.1	22.8
4	20.4	25.7	25.4	23.0	23.3	23.8
5	13.7	18.8	24.2	30.9	36.2	25.1
Total	100.0	100.0	100.0	100.0	100.0	100.0

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.
Sources: 1915 Iowa State Census Sample; 1940 Federal Census

the outcome of interest (income), controlling for it or splitting the sample based on it is problematic. In order to estimate an accurate IGE parameter for the early 20th century, we need to be able to observe sons that remain in their father's state of residence and sons that move elsewhere.[53]

What is the probability that a son born in a given quintile will be in the same or another quintile in 1940? Table 8, an income quintile transition matrix, can be used to answer questions of that form. 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).[54] A son whose father is in the bottom quintile in 1915 has only a 14.6% 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 is has only a 7% chance of being in the bottom quintile in 1940, while the odds are more than twice as high (nearly 15%) that a son born in the bottom quintile remains there. While the table clearly presents a degree of intergenerational immobility, when compared these percentages to a transition matrix for the modern period (sons born in the 1980-82 cohorts) from Chetty et al. (2014a), there is in fact more mobility between 1915 and 1940, at least at the lower end of the income distribution. Sons born in the bottom quintile are more than two times as likely to remain there as adults in the modern data than historically: 33.7% to 14.9%. Historically, sons born in the second quintile have a 69% chance of being in a higher quintile in adulthood; that probability is only 52% in modern data. At the top end, however, transition probabilities are similar between the two periods: sons born in the fourth or fifth quintiles have 23.7% and 36.2% probability of being in those quintiles as adults in the Iowa sample, compared with 24.4% and 36.5% probabilities in the Chetty et al. (2014a) sample.

[53]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.

[54]The columns sum to 100 because conditional on the father's quintile a son must be in one of the five groups. However, the rows do not sum to 100 because the sons' quintiles are based on the 1940 IPUMS 1% sample, not just the sons in my matched Iowa sample.

4.2 Alternative Measures of Intergenerational Mobility

I also estimate the intergenerational mobility of education and present these results in Panel A of Table 9.[55] In addition to serving as a (potentially) more accurately measured check on my income results, the education IGE estimate is a valuable and important historical parameter. 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 19th century and then through the high school movement in the early 20th 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 or not this massive public investment in education also reduced the strength of the relationship between a son's educational prospects and his father's educational outcomes can help scholars understand the role of public programs in shaping or changing inequality. The literature suggests an IGE of education of 0.46 (Hertz et al. 2007). As Table 9 shows, I find a much lower IGE parameter for schooling, between 0.187 and 0.275. This suggests that, like income, educational mobility in the US was higher in the early 20th century than it is today.[56] Similar to the results presented on income mobility, there is more mobility of education among rural sons than urban sons as well.

Though not my preferred measure of economic status or position, I can also estimate intergenerational mobility using occupation scores.[57] 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 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.[58]

The results, presented in Panels B and C of Table 9, suggest that, measured with occupation score,

[55] Though the regressions are estimated in levels, I will follow the literature in describing these relationships as intergenerational elasticities.

[56] 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.

[57] For research on economic outcomes in periods without earnings data, many economists have turned to such occupation score measures, including Olivetti and Paserman (2014) and Abramitzky et al. (2012, 2013, 2014).

[58] 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.

Table 9: 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	3378	2505
		(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	2095	1477
		(0.029)	(0.030)	(0.056)		
	Excluding Farmer Sons	0.317	0.303	0.338	2006	1590
		(0.027)	(0.028)	(0.046)		
B. 1915 Occupation Score Mobility						
Log Occupation Score 1915 Basis	Full Sample	0.167	0.162	0.265	3039	2280
		(0.030)	(0.030)	(0.052)		
	Urban Sample	0.229	0.226	0.308	1154	940
		(0.035)	(0.036)	(0.065)		
	Rural Sample	0.110	0.101	0.176	1885	1340
		(0.040)	(0.038)	(0.091)		
	Excluding Farmer Sons	0.180	0.182	0.289	1762	1415
		(0.030)	(0.030)	(0.053)		
C. 1950 Occupation Score Mobility						
Log Occupation Score 1950 Basis	Full Sample	0.441	0.366	0.386	3204	2375
		(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	0.419	0.401	1984	1395
		(0.030)	(0.030)	(0.070)		
	Excluding Farmer Sons	0.228	0.220	0.263	1869	1480
		(0.027)	(0.027)	(0.061)		

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

mobility was quite high in the early 20th century, corroborating my findings with income and education.[59] Specifically, I estimate an IGE parameter between 0.177 and 0.234 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. Though the confidence intervals still do not contain the point estimates for income mobility in the modern period, the differences are far smaller, suggesting only slightly more mobility in the early 20th century than today. 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 occupational score to determine intergenerational mobility, especially at a time of large structural change in the economy.

4.3 Occupation Results

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 the Altham statistics discussed above (Altham and Ferrie 2007). My occupational transition table from 1915 to 1940 is presented in Table 10. 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.[60] Sample sizes are different from previous portions of the analysis because not all occupation strings are matchable to occupation codes or broad occupation groups.

As detailed earlier, 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 10 yields $d(IA1940, I) = 16.03$. Long and Ferrie (2007) report $d(US1880, I) = 12.09$, $d(US1900, I) = 14.58$, and

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

[60]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.

Table 10: 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	220	190	1,078
Unskilled	332	179	196	91	798
White Collar	244	313	159	337	1,053
Total	1,463	928	619	649	3,659

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

$d(US1973, I) = 20.76$. The Altham statistic generated by my linked sample of fathers and sons from 1915 to 1940 is larger—thus indicating less mobility—than the measures presented by Long and Ferrie (2007) for the 19th 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 11. 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 modern estimates reported by Long and Ferrie.

The second and third columns of Table 11 present the Altham statistics for the urban and rural subsamples of my linked data. Relative to the full sample, I find more mobility among both the urban sample and the rural sample.[61] I also find more mobility in the urban sample than in the rural sample.[62] 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. The subsamples also confirm the previous finding of more mobility historically than today.

4.4 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 5 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).[63] Table 12 gives the

[61] 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 a very undesirable property of Altham statistics. In theory, an alternative statistic might possess a form of continuity and the intermediate value theorem.

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

[63] 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

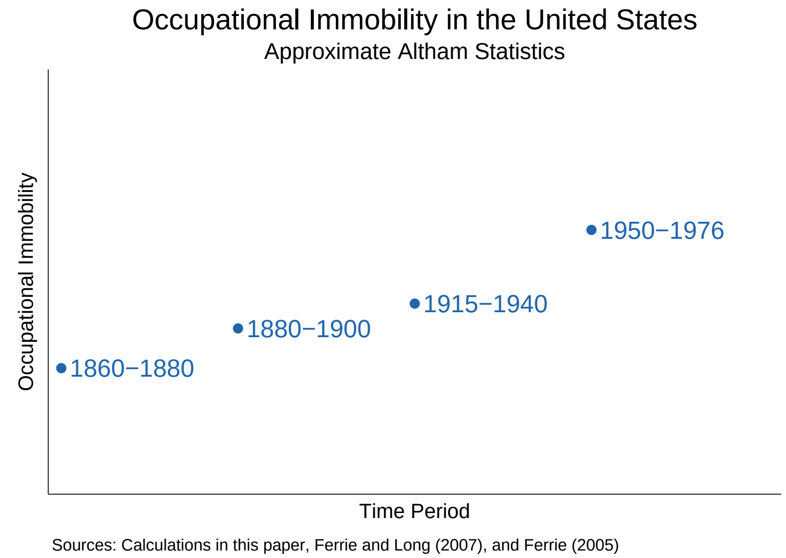


Figure 4: Occupational Immobility measured by Altham statistics from 1860 to 1976. This graph suggests that occupational immobility rose over time and that there is less mobility in the mid-20th century than there was between 1915 and 1940. The Altham statistics for each period presented in the plot are statistically different from one another. See Table 11 for exact values and distance tests.

Table 11: 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
$Pr(d(P, IAsons) = 0)$	.	0.718	0.711
$Pr(d(P, IAurbansons) = 0)$	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)

Table 12: Sons of Iowa Residences in 1940

	Matched 1915-1940 Sample		1940 IPUMS Sample	
	Count	Share (%)	Count	Share (%)
Iowa	2859	63.8	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	498	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

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).[64] Nearly 64% remain in Iowa.[65] Neighboring 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.[66]

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 13 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 13.

1940 Federal Census sample used for son-matching.

[64]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.

[65]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.

[66]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.

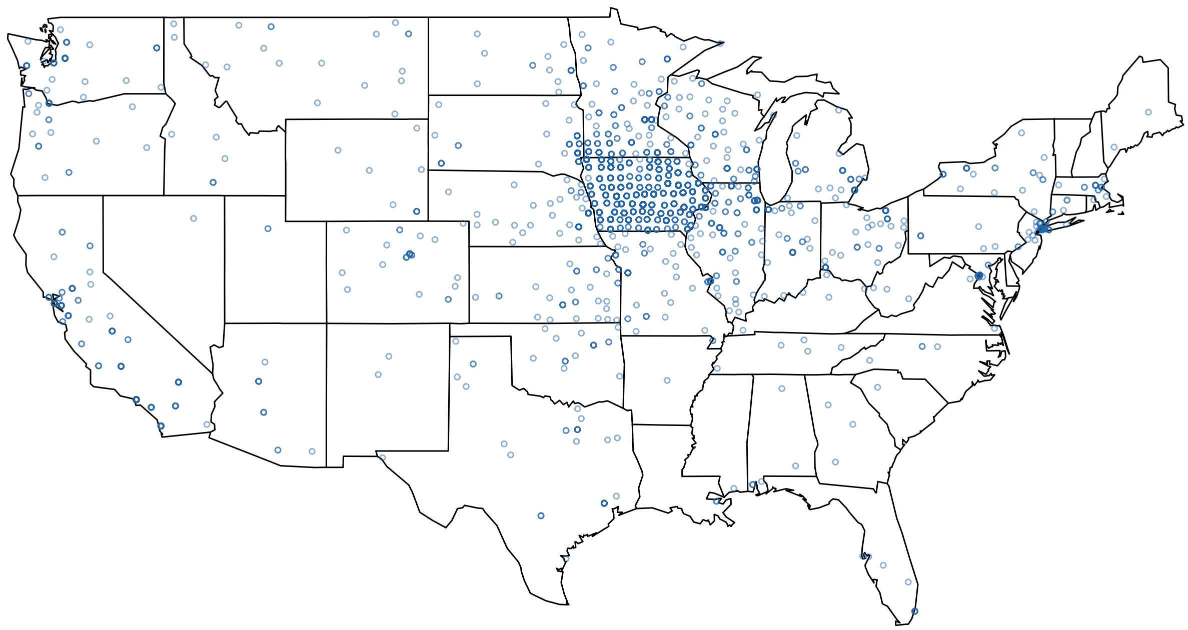


Figure 5: Sons County of Residence in 1940. The darker symbols implies greater density of points at a given latitude and longitude.

Table 13: Geographic Mobility: Miles Moved 1915 to 1940

	Full Sample			Urban		Rural	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Father Earnings	0.0474		-0.00973	0.252*		-0.102	
	(0.0616)		(0.0638)	(0.130)		(0.0738)	
Father Education		0.0578***	0.0478***	0.0419	0.0709***	0.0418**	0.0497***
		(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	3657	3937	3651	1479	1563	2172	2374
Number of Clusters	2629	2836	2624	1141	1207	1483	1629
R-squared	0.0252	0.0291	0.0292	0.0467	0.0448	0.0180	0.0156

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

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 20th century: the Great Depression and the Dust Bowl. The Dust Bowl may be a particularly important factor to consider when following the earnings and education trajectories of men growing up in Iowa during this time period. Huge dust storms during the 1930s blew topsoil off farms in the American plains, causing severe and lasting erosion to once highly productive and fertile farmland. Hornbeck (2012) uses variation in dust bowl severity to measure the effects of environmental catastrophes on economic outcomes. He finds that the Dust Bowl led to immediate and persistent reductions in agricultural land vales and production and to large population out-migrations. Dust Bowl severity in Iowa was highly varied throughout the state (Hornbeck 2012, Figure 2). While the central and northern portions of the state suffered little to no erosion, the southwestern and southern areas of the state endured high levels of topsoil erosion and loss. Of the rural counties included in the Goldin-Katz sample, Adair, Carroll, Johnson, Lyon, Montgomery, and Wayne were all in medium erosion zones; the other rural counties were in low erosion areas.

The regressions presented in Table 13 and used in all other analyses in this paper include county level fixed effects; these fixed effects control for differences in Dust Bowl exposure at the county level. However, it may be the case that fathers with higher income or higher levels of human capital (measured here by years of education) were better prepared to deal with the Dust Bowl (either through out-migration or changing their agricultural practices). As farms are often passed down from father to son, the probability of moving or the distance moved depends greatly for these sons on the interaction of the Dust Bowl and their father's attributes. Results on a sample restricted to fathers who were farmers in 1915 are limited.[67] It does not appear that the Dust Bowl interacted in any general way with either father's income or education to determine the son's geographic mobility.

5 Conclusion

This paper presents estimates of both income and educational mobility for men born in Iowa between 1900 and 1910, showing that their mobility rates were much higher than those of men born since 1960. I matched fathers from the Iowa State Census of 1915 to their sons in the 1940 Federal Census. In my sample of fathers and sons, I estimate a lower intergenerational elasticity of income than is found in modern studies of the United States, suggesting higher levels of income mobility. Estimates of an intergenerationl rank-rank parameter, relating the father's ranking in the income distribution in 1915 to the son's ranking in the income distribution in 1940, also show more mobility historically than in more recent settings. Intergenerational

[67] See Table A.1 in the appendix.

mobility of education is higher in my sample than in modern measures as well. I find sons in rural counties in 1915 to have more mobility of both income and education than urban sons. When I compute the standard measures of occupational mobility for my sample, I find generally higher levels of mobility between 1915 and 1940 than is found in modern estimates as well.

In addition to constructing estimates of intergenerational mobility of income, education, and occupation, my dataset also enables me to measure other forms of intergenerational mobility. Because I match into the Federal Census of 1940, I am not restricted to considering non-migrants. Thus, I am able to measure the relationship between parental outcomes and children's geographic mobility. Many of the sons in my sample remained in Iowa, but a large portion moved within the state to urban areas or to larger urban centers outside the state, such as Chicago and Los Angeles. I can also measure differential geographic mobility between rural and urban sons. My results suggest that the sons of more educated farmers are more mobile, especially those raised in urban Iowa, but that there is little to no effect of father's income once controlling for father's education.

This paper demonstrates that there was more intergenerational mobility in the early twentieth century than there is today. An important question is why. The high school movement and the huge expansion of access to public education could have been one driver. 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—which is unlikely if free schooling had not been widely available. 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. Even without a change in the underlying mobility parameters, the changing composition of the country—fewer rural residents and farmers in each subsequent generation—would lower mobility over time.[68] There were high levels of geographic mobility throughout the country in the early twentieth century, both to the west and to urban areas. Geographic mobility and economic mobility are likely correlated, but this pushes the question back a step. What determined geographic mobility? My results suggest that the odds of a son moving or his distance traveled may have been related to his father's outcomes, but that overall an individual's migration decision is difficult to predict. Finally, the Great Depression almost certainly altered the economic fortunes of most sons in my sample, perhaps by changing the relative returns to different professions or skills in unpredictable ways. However, the change induced by the dislocation of the Great Depression is unlikely to be the full explanation, given that I find high levels of educational mobility as well. Even the youngest sons in my sample were 19 in 1929, and the vast majority had already completed their

[68] 42% of the fathers in my sample were farmers compared to only 21% of the sons.

schooling before the downturn.

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 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? Was there variation across the US in this change? And, finally, what caused this shift? Based on the results in this paper, it appears unlikely that the Great Depression drove this change, at least among the generation of sons born before but employed after it. Did the Great Compression, documented by Goldin and Margo (1992), induce a new era of lower mobility? 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 exist but much of it is not yet accessible. With the 2022 release of the full non-anonymous 1950 census—and the 2032 release fo the 1960 census—it will be possible to track intergenerational mobility through the middle of the 20th century.

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.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2014. "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." *Journal of Political Economy* 122:467–717.

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.

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.

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. 2008. "Intergenerational Persistence in Educational Attainment in Italy." http://www.econstor.eu/dspace/bitstream/10419/34814/1/575205105.pdf.

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.

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. Emerald Group Publishing Limited.

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.

Ferrie, Joseph P. 2005. "History Lessons: The End of American Exceptionalism? Mobility in the United States Since 1850." *Journal of Economic Perspectives* 19:199–215.

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.

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, 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.

Jencks, Christopher and Laura Tach. 2006. "Would Equal Opportunity Mean More Mobility?" In *Mobility and Inequality: Frontiers of Research in Sociology and Economics*, edited by Stephen Morgan, David Grusky, and Gary Fields, pp. 22–58. Stanford University Press.

Kane, Thomas J, Cecilia Rouse, and Douglas Staiger. 1999. "Estimating Returns to Schooling When Schooling is Misreported." http://www.nber.org/papers/w7235.

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." *Economic Journal* 117:61–71.

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.

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. 5 – 23. 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.

Olivetti, Claudia and M Daniele Paserman. 2014. “In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1930.” http://people.bu.edu/olivetti/papers/Olivetti-Paserman_NameOfTheSon_July2014.pdf.

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

Piketty, Thomas and Emmanuel Saez. 2003. “Income Inequality in the United States.” *Quarterly Journal of Economics* 118:1–39.

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.

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.

Appendix[69]

Additional Tables and Figures

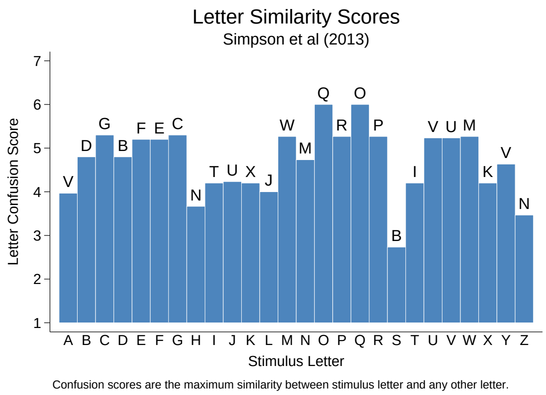
Table A.1: Geographic Mobility and the Dust Bowl, 1915 to 1940

	Log Miles Moved		In Iowa in 1940		In Same County in 1940	
	(1)	(2)	(3)	(4)	(5)	(6)
Log Father Earnings	-0.194		-0.00134		0.0586	
	(0.147)		(0.0355)		(0.0366)	
Father Education		-0.00533		0.00340		0.00677
		(0.0298)		(0.00753)		(0.00799)
Dust Bowl Severity	-1.720	-0.0596	0.119	-0.00497	0.502	0.138
	(1.222)	(0.356)	(0.299)	(0.0879)	(0.318)	(0.0933)
Dust Bowl Severity × Earnings	0.248		-0.0171		-0.0675	
	(0.178)		(0.0433)		(0.0462)	
Dust Bowl Severity × Education		0.000415		0.00188		-0.0117
		(0.0415)		(0.0103)		(0.0110)
Son Age Quartic	Yes	Yes	Yes	Yes	Yes	Yes
Father Age Quartic	Yes	Yes	Yes	Yes	Yes	Yes
Name String Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	1421	1513	1421	1513	1421	1513
Number of Clusters	930	987	930	987	930	987
R-squared	0.00805	0.00312	0.00412	0.000813	0.00107	-0.0000557

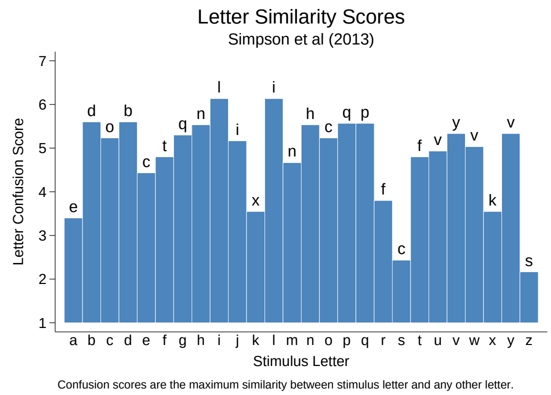
In the first two columns, the dependent variable is the log of 1 + the number of miles between the son's county in 1940 and in 1915. In the second two columns, the dependent variable is an indicator variable for whether or not the son lives in Iowa in 1940. In the final two columns, the dependent variable is an indicator variable for whether or not the son lives in the same county in 1940 as he lived in 1915. Standard errors clustered by family. Dust Bowl severity is coded as 0 (low erosion), 1 (medium erosion), or 2 (high erosion), following Hornbeck (2012), measured at the county level. 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

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



(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.

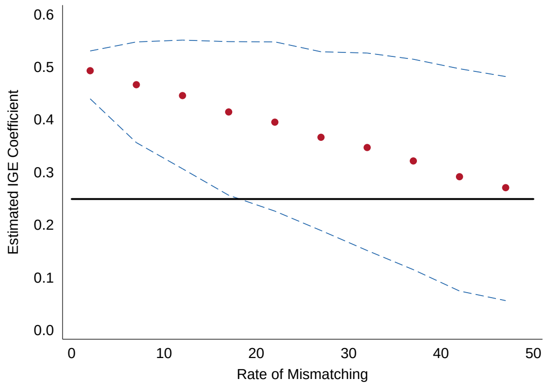
Matching Bias and Measurement Error

Matching Bias

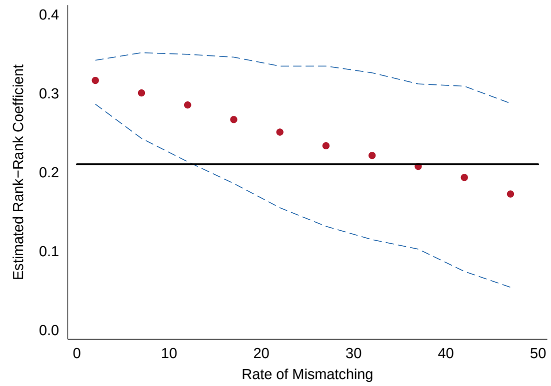
The analysis that I conduct in this paper requires the construction of a dataset that links fathers and sons over time, between two censuses. The linking procedure, though carefully conducted, likely introduces a type of measurement error and bias to the estimation of IGE parameters for historical periods that might not be present in modern 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 modern 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.[70] 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 modern 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, year of birth, and, where applicable, parent's states of birth. However,

[70]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.



(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

despite my best efforts at ensuring a unique and correct match, I may identify the "wrong" son in 1940.[71]

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.[72] Using this new data I estimate 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 percent. 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 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 $\beta = 0.249$ represents my largest estimate of the IGE between 1915 and 1940 from Panel A of Table 7. 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 modern samples.[73] 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 $\beta = .210$, the largest estimate of the rank-rank parameter for the full sample in Panel B of Table 7. In this case, the mismatch rate would have to be at

[71] 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).

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

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

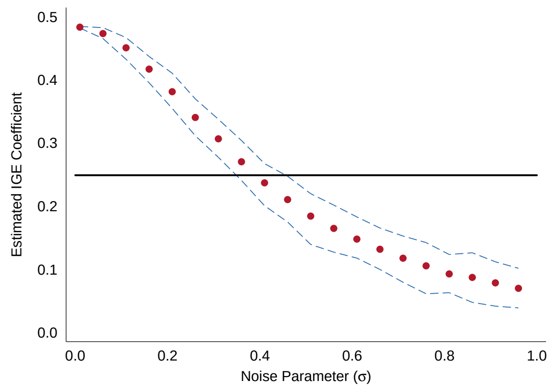


Figure A.3: Simulated Intergenerational Elasticity of Income in the PSID as the noise in earnings varies

least 30% to induce such a low rank-rank estimate of intergenerational mobility in the modern data as I find historically.

Measurement Error

As Haider and Solon (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 modern 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 intergenerational mobility was in fact lower in the early 20th 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.

Farmer Income in 1940

For each individual listed in the 1940 Federal Census, annual wage and salary earnings and weeks of work are reported.[74] However, the census does not include information on either business or farm income as in later censuses. In practice, 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 2529 matched sons in my sample, 675 report zero earnings in 1940. Of these, more than half (348) are farmers or farm owners or farm operators. The other 327 are a variety of occupations, including labourers (16), proprietors (16), managers (15), owners (14). amd various forms of doctors and lawyers.[75]

In the main results presented in Panels A and B of Table 7, 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 mechanization, 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.[76] The estimated IGE and rank-rank parameters including sons with imputed capital income are presented in Table A.2 and suggest even more mobility historically, relative to modern 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

[74] 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

[75] The most common occupation among the non-farmers without earnings is actually to have no occupation listed (45 of the 327). These men are likely unemployed and not working for the WPA and thus the zero reported earnings are a correct measure not unreported data.

[76] I do this only the 348 farmers in my sample in 1940 without reported earnings. For the 327 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.

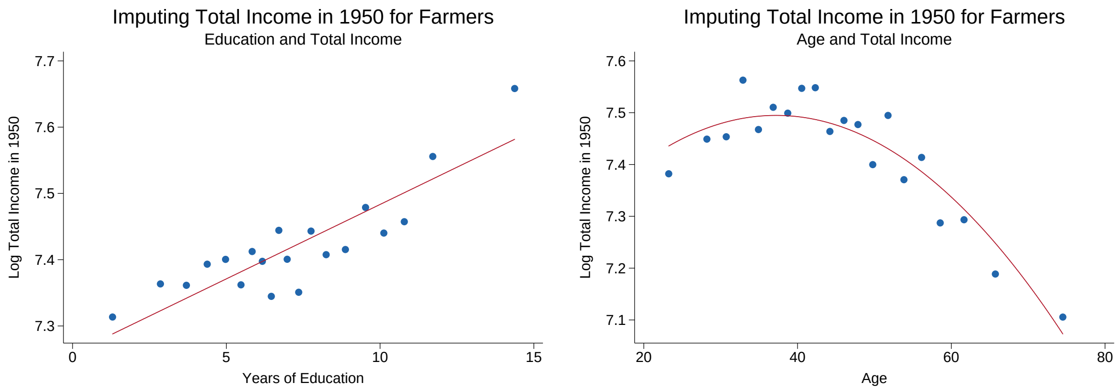


Figure A.4: Binscatter graphs presenting the correlation between years of education and age with log total income in 1950 for farmers from the IPUMS 1% sample. Both figures include controls for weeks worked, state of birth, state of residence, and education or age (when the variable is not on the x-axis).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

business and farm income. In a sample of only farmers in 1950, I regress total income on years of education,[77] a quartic in age, number of weeks worked, and indicator variables for state of birth and state of residence in 1950. 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%,[78] while the relationship between age and total income is non-linear. 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.

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

[78]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.

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.[79] 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. The results are presented below. I report the true census relationship from IPUMS across the table and my imputed family relationship down the table.[80] 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.

	True Census Relation				
	child	father	mother	other	Total
child	18265	17	2	569	18853
father	38	6391	0	38	6467
mother	40	0	6521	84	6645
other	151	69	47	804	1071
Total	18494	6477	6570	1495	33036

[79]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 modern 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.

[80]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.

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 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.2, I show that normalizing earnings to be weekly, rather than annual, increases historical levels of mobility.[81]

In Panel B of Table A.2, I present intergenerational "elasticity" estimations[82] with the income variables in levels. Panel C of Table A.2 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 modern 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.[83] 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.2 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.

Intergenerational Mobility using Family Income

To generate the modern 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 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

[81] 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.

[82] 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.

[83] 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.

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.2 underscore that expectation.

Table A.2: Intergenerational Mobility Estimates: Alternative Function Forms

	Specification			Observations	Clusters
	(1)	(2)	(3)		
	A. Intergenerational Elasticity (IGE), Log Weekly Earnings				
Full Sample	0.207	0.194	0.266	2039	1667
	(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	1035	843
	(0.040)	(0.041)	(0.116)		
	B. Annual Income in Levels				
Full Sample	0.319	0.319	0.438	2116	1731
	(0.044)	(0.042)	(0.104)		
Urban Sample	0.541	0.548	0.558	1025	842
	(0.082)	(0.084)	(0.208)		
Rural Sample	0.193	0.203	0.352	1091	889
	(0.045)	(0.045)	(0.107)		
	C. Square Root of Annual Income				
Full Sample	0.189	0.184	0.251	2116	1731
	(0.031)	(0.030)	(0.069)		
Urban Sample	0.298	0.298	0.286	1025	842
	(0.057)	(0.058)	(0.141)		
Rural Sample	0.120	0.132	0.217	1091	889
	(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	2437	1959
	(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	1330	1059
	(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 depdendent variable. In Panel D, the depdendent 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

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 20th century.[84] 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 technology, supply, and demand are likely to grow over time.

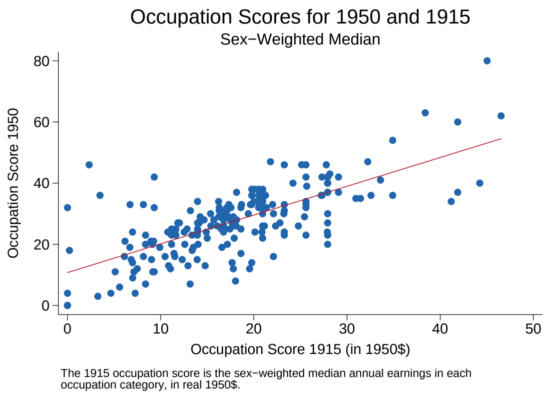
Taking advantage of the 1915 Iowa State Census, which was the earliest census in the US to record respondents' incomes, occupation, and education level, I construct two variants on the traditional measures of occupation score, measured not in 1950 but in 1915. While these measure are highly correlated 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 20th century.

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

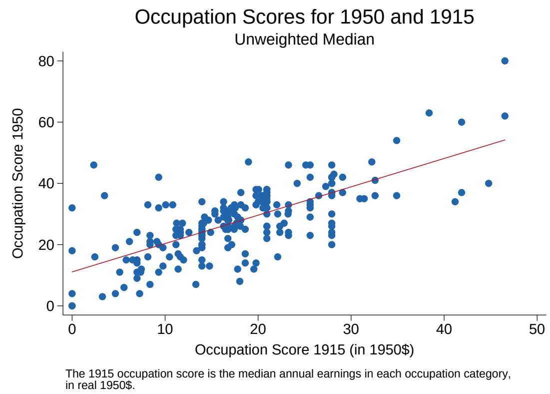
> The occupational income score indicates the median total income – in hundreds of dollars – of the personas [sic] in each occupation in 1950. It is calculated using data from a published 1950 census report. For the post-1950 period, the score reflects the weighted average income of the 1950 occupational components of each contemporary occupation. In practice, this has only a small effect, but it means that the measure can vary slightly across census years for a given occupation.[85]

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.

[84] On the historical occupation tasks of accounting and bookkeeping specifically, see Rosenthal (2013).
[85] https://usa.IPUMS.org/usa/chapter4/chapter4.shtml



(a)



(b)

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

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.[86] 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.[87]

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

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".[88] "Mechanical engineers" and "Power station operators" are both relatively low-paid positions in 1915, but by 1950 they are in the upper

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

[87] 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.

[88] 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.

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.