

## MULTIPLE MEASURES OF HISTORICAL INTERGENERATIONAL MOBILITY: IOWA 1915 TO 1940\*

*James J. Feigenbaum*

Was intergenerational economic mobility high in the early twentieth century in the US? Comparisons of mobility across time are complicated by the constraints of the data available. I match fathers from the Iowa State Census of 1915 to their sons in the 1940 Federal Census, the first state and federal censuses with data on income and years of education. I can estimate intergenerational mobility between 1915 and 1940 based on earnings, education, occupation, and names. Across all these measures, I document broad consensus that rates of persistence were low in Iowa in the early twentieth century.

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

I take advantage of two historical data sources that enable me to measure mobility in many different ways – including intergenerational elasticity of earnings, rank–rank persistence, occupational score elasticity, occupational transitions, education persistence, imputed status based on both given names and rare surnames. Because I can estimate these various mobility measures within the same source, I shed light on how well these various measures agree with one another, at least for the early twentieth century.

\* Corresponding author: James J. Feigenbaum, Department of Economics, 270 Bay State Road, Boston University, Boston, MA 02215, USA. Email: jamesf@bu.edu.

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

<sup>1</sup> 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.

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

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.<sup>3</sup>

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

Was there more mobility historically than today? I find a lower intergenerational income elasticity during the first half of the twentieth century in the US than studies find in the second half of the century. I also measure intergenerational mobility using the rank–rank parameter (Chetty *et al.*, 2014a) and similarly find more mobility historically. The results for education, occupational mobility – measured either using transition matrices or occupation score – and name-based mobility all point in the same direction. But such differences between contemporary and historical mobility could be spurious. For one, my sample is not a random draw of the American population. By population density, urbanisation, and the share foreign born, Iowa is nearly the median state. On other dimensions, Iowa in 1915 is an outlier: it was also almost universally white, invested more and earlier in education than other states, and had relatively low levels of inequality. In addition, measurement error, either due to the historical nature of my data, the difficulty of creating longitudinal linked samples, or the single year I am able to observe fathers and sons in the census may push me to find excess mobility. I work to mollify these concerns in a few ways. In the earnings data, I show that the estimated differences between contemporary and historical mobility remain after adjusting the contemporary sample to mirror the historical sample in measurement noise and demographic and geographic composition. The agreement across many measures of mobility, some based on outcomes like education and occupation that should be less noisy than a single year of earnings, also sharpens the comparative result. Ultimately, my results do not prove there is less mobility today than in the past, but, taken along with other evidence comparing mobility over the twentieth century (Parman, 2011; Long and Ferrie, 2013), strengthens the belief that there was more mobility in the early twentieth century.

The article proceeds as follows. In the first Section, I discuss the historical data that I draw on and my data collection and census linking procedures. I describe what, if any, bias the census linking procedure may induce and compare Iowa in 1915 to the rest of the nation. In Section 2, I review the various methods of measuring intergenerational mobility that I am able to apply to my sample of linked fathers and sons. In Section 3,

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

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

I present my estimates of intergenerational mobility in the early twentieth century for income, education, occupation and name-based status. Section 4 concludes the article.

## 1. Data and Record Linkage

### 1.1. *Linking the Iowa 1915 and Federal 1940 Censuses*

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

The 1915 Iowa Census enumerated the 2.3 million Iowa residents in 1914. It was the first American census of any kind to include data on both annual income and years of education in addition to more traditional census measures, and it also includes respondent name, age, place of residence, birthplace, marital status, race and occupation. I use the Iowa State Census sample digitised by Claudia Goldin and Lawrence Katz for their work on the historical returns to education (Goldin and Katz, 2000, 2008). The Goldin–Katz sample includes 26,768 urban residents (5.5% of the total urban population of Iowa in 1915) and 33,305 rural residents (1.8% of the total rural population). Figure 1 presents a map of the counties and cities included in the Goldin–Katz sample. The three large Iowa cities sampled are Des Moines, Davenport and Dubuque.<sup>5</sup> In 1915, the population of Des Moines was approximately 97,000

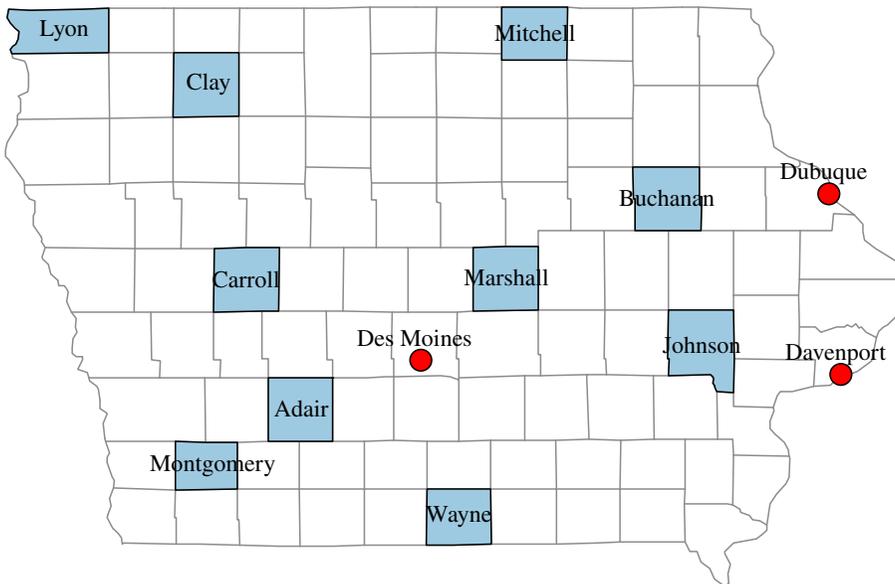


Fig. 1. *Map of Iowa 1915 Cities and County Sample*

Note. Colour Figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)

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

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.<sup>6</sup>

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 the annual income. I restrict my analysis to sons in 1915, because name changes make it impossible to locate most daughters in the 1940 Census. This leaves me with a sample of 7,580 boys in Iowa in 1915, 6,071 of whom have fathers in their households, and the requisite data on both the father’s education and income.

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

Once the matched sons are identified, I record the pertinent data from the 1940 census, enumerated in the shadow of the Great Depression and on the eve of WWII.<sup>8</sup> I use the 1940 census because it is the only census suitable for tracking the sons of

<sup>6</sup> For more details on the construction of the Goldin–Katz sample, see Goldin and Katz (2000).

<sup>7</sup> The machine learning approach, which I detail in Feigenbaum (2016a), teaches an algorithm how to replicate the careful hand-linking work of a researcher but at scale and with extreme consistency. I generate the training data for the algorithm by manually linking a 30% random sample of sons from Iowa in 1915 into the 1940 census. To be considered as a possible link, records must first match exactly on state of birth, be within three years’ difference in the year of birth, and within 0.3 Jaro–Winkler string distance in both first and last names. Then, within that filtered list of possible matches, the records are double-entry matched to the 1940 census manually by trained research assistants. Records determined to match are marked as such, records without a clear and certain match in the 1940 census are marked as unmatched. With this corpus of links, I then train a match algorithm. The match algorithm is used to reduce between-researcher variability in match quality, to speed up the matching process, and to ensure data replication. The method improves on previous efforts based on phonetic matching because typos and transcription errors will not cripple the matching. The matching algorithm uses Jaro–Winkler string distances in first and last names, exact matches on state of birth, absolute difference in year of birth, Soundex matches for first and last names, middle initial matches, matching first and last letters of first and last names and other record-based variables to predict whether a record is a true or false match. The algorithm also factors in the match quality of other possible matches for the given record searched for, only making matches when a record is a significantly better match than other possibilities. Based on cross-validated out of sample predictions within my training data, the match algorithm has a true positive rate of nearly 90% and a positive prediction rate of 86%. In Table A3 of the online Appendix, I detail the exact weights on the match algorithm used.

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

Iowa. The 1940 Census was the first federal census to collect data on incomes, weeks of work, and years of education of the entire population.<sup>9</sup> While such data was also collected in 1950 and 1960, federal censuses are privacy-restricted for 72 year after enumeration. The data on names required for linking the 1950 census will not be accessible until 2022, 2032 for the 1960 census.<sup>10</sup> Because it is a national sample, I do not have to worry about losing many sons to out-migration, which might otherwise bias my estimates.

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

## 1.2. *What Predicts Matches?*

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

First, I determine the name commonness of both the first and last name, relative to all names in the pooled IPUMS sample of the 1910 and 1920 censuses.<sup>12</sup> 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.

<sup>9</sup> The earnings in the 1940 census are top-coded at \$5,000. However, only 44 of the sons in my sample report such high earnings; I code them as earning \$5,000 for my analyses. Past federal censuses record contemporary school enrolment for each person (child) but not years of schooling completed for adults no longer in school. Earnings data were collected in 1940 only for wage and salary workers. The data collected are the ‘total amount of money wages or salary’ but enumerators were instructed: ‘Do not include the earning of businessmen, farmers, or professional persons derived from business profits, sale of corps, or fees’. For more, see <https://usa.ipums.org/usa/voliii/inst1940.shtml#584>. The importance of these 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 online Appendix A.3, I impute earnings for farmers using the 1950 census, which did collect data on capital income and non-wage and salary earnings. Using these imputed earnings, I estimate even higher levels of mobility than in my main results.

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

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

<sup>12</sup> 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 one 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.

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

Finally, I calculate a name's Scrabble score as an alternative measure of both name commonness and name simplicity.<sup>16</sup> 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.*, forthcoming).

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

More serious issues could be generated by differential matching rates according to father, son, or family characteristics in 1915. In my sample, I find little evidence that such characteristics strongly affect the probability of matching. In Table 2, I present the estimated effects of a set of variables observed for fathers and sons in 1915 on the probability of positively locating the son in the 1940 Census.<sup>18</sup> 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

<sup>13</sup> *l* is likely to be confused with *f* and *i* for example, while *n* is similar to both *h* and *m*.

<sup>14</sup> Recall matches are made using census indices transcribed by Ancestry.com.

<sup>15</sup> 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 A1 in the online 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.* (2013) 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.

<sup>16</sup> Biavaschi *et al.* (forthcoming) introduce the use of Scrabble scores into the economic literature. They use this measure to predict name changes by immigrants to the US during the early twentieth century. Scrabble point values were based, originally, on the frequency of letters on US newspaper front pages.

<sup>17</sup> With controls for commonness and length, the Scrabble scores do not seem to relate to match rates.

<sup>18</sup> 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.

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

	(1)	(2)	(3)	(4)	(5)
Name commonness, first name	0.041** (0.017)				0.056*** (0.020)
Name commonness, last name	-0.122*** (0.039)				-0.121*** (0.039)
String length, first name		0.013*** (0.004)			0.020*** (0.004)
String length, last name		-0.002 (0.004)			-0.002 (0.004)
Normalised letter similarity score, first name			0.019*** (0.007)		0.024*** (0.007)
Normalised letter similarity score, last name			0.006 (0.007)		0.005 (0.007)
Normalised scrabble score, first name				-0.001 (0.006)	-0.002 (0.007)
Normalised scrabble score, last name				0.009 (0.006)	0.008 (0.006)
Observations	7,580	7,580	7,580	7,580	7,580
Clusters	4,731	4,731	4,731	4,731	4,731
Adjusted R <sup>2</sup>	0.002	0.002	0.001	0.000	0.007

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

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

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

<i>X</i>	$\beta$	SE	Predicted match rate with <i>X</i> 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 Iowa	0.138	0.018	61.0	61.0
Father foreign born	-0.063	0.013	61.2	54.8

*Notes.* This Table presents the coefficients from a series of linear probability regressions with *X* as the primary independent variable, controlling for first and last name commonness, length, letter similarity, and Scrabble score. As in Table 1, there are 7,580 observations and 4,731 clusters, clustering standard errors by family.

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

Scrabble score. I am slightly more likely to match sons who had higher income or more educated fathers (or mothers) in 1915 but these effects are both economically and statistically insignificant. For example, the probability of matching a son with a father at

the 25th percentile of income is only 1 percentage point lower than matching a son with a father at the 75th percentile of income. Similarly small effects of both father's and mother's education can be seen as well. I am also less likely to match sons in the urban sample which follows from the slightly higher match rate among rural sons. I am also more likely to link sons born in Iowa, even after conditioning on name string characteristics.<sup>19</sup> All analysis undertaken in this article will include controls for son's place of birth, place of residence in 1915, and, where appropriate, father's place of birth.

### 1.3. *How Does Iowa 1915 Compare to the US in 1915?*

Iowa undertook an enumeration of its population in 1915 unlike any other in American history. In addition to gathering reliable 'characteristics of population [and] agricultural and other industries', the census also collected data on the earnings of its citizens and the extent of education in population. The census report announced that the finding 'happily confirms the claim of very high rank for Iowa in educational standing'.<sup>20</sup> That Iowa collected such earnings and education data makes the present study possible, but its rarity among contemporary censuses also raises a question: how unique was Iowa and how might that affect my estimated mobility rates?

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

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

<sup>19</sup> 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.

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

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

<sup>22</sup> I aggregate the census totals for native born with two foreign parents and native born with one foreign parent.

Table 3  
*Iowa and the Nation in 1910 and 1920*

	1910				1920			
	Iowa	Midwest	Nation	Iowa rank	Iowa	Midwest	Nation	Iowa rank
Total population (100k)	22.25	298.89	919.72	15	24.04	340.20	1,057.11	16
Population density (people per miles <sup>2</sup> )	40.02	39.52	30.93	24	43.25	44.98	35.55	24
Urbanisation rate	30.57	45.12	45.66	25	36.42	52.25	51.23	24
Rural rate	69.43	54.86	53.66	23	63.58	47.75	48.77	25
White share	99.30	97.96	88.87	5	99.17	97.49	89.70	6
Black share	0.67	1.82	10.69	32	0.79	2.33	9.90	34
Foreign born share	12.29	15.66	14.51	28	9.39	13.51	12.97	27
Native born with foreign parents share	28.42	27.85	20.55	14	26.21	27.35	21.46	19
Sex ratio (M/F)	1.07	1.08	1.06	24	1.05	1.06	1.04	24
Farms (per 100 people)	9.76	7.47	6.92	16	8.88	6.41	6.10	17
Share of farm acres improved	86.92	72.23	54.44	1	85.46	69.20	52.63	1
Literacy rate (ages 10+)	98.66	97.47	94.00	1	99.14	97.98	95.33	1
School enrollment rate (ages 6 to 20)	69.57	66.47	62.34	5	72.39	70.33	68.10	13
Newspaper circulation per thousand	220.63	270.57	237.21	18	313.68	310.62	265.55	11
Daily newspapers per 100k	2.79	2.95	2.42	15	2.16	2.33	2.02	20
Gini coefficient (based on farm sizes)	0.38	0.47	0.54	47	0.35	0.47	0.56	47

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

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

when I calculate inequality using the distribution of farm sizes in the census, Iowa is the second most equal state in the continental US in both 1910 and 1920.

How might the differences between Iowa in 1915 and the rest of the nation affect my analysis?<sup>23</sup> Iowa's racial homogeneity may create more mobility than in the rest of the country. If the Gatsby Curve – as Krueger (2012) dubbed the correlation in OECD countries in the last decade between intergenerational mobility and inequality – held in the early twentieth century, again, perhaps Iowa will demonstrate more mobility than other states in this era because of its low inequality levels. Alternatively, the relatively low share of the foreign born may reduce mobility, if the children of recent immigrants are likely to experience fast upward mobility as they assimilate. In the

<sup>23</sup> Could any uniqueness about the year 1915 affect my analysis? I find that unlikely. The US did not enter The First World War until April 1917 and the draft did not begin until May 1917. Fighting in Europe began in July 1914 but little if any economic effects of the first six months of the war would have been felt by the residents of Iowa. Rockoff (2008) documents no rise in the industrial production or the stock market in response to the beginning months of European fighting. Changes in taxing, spending and borrowing, particularly the famous liberty bond programme, were not introduced until October 1917. None of the fathers in my sample had been mobilised and any fathers working in industries that might later be subject to government price or wage controls were not subject to them in 1914. In the nearly 800 page report on the 1915 Iowa Census, published in 1916 by the enumerators, the ongoing European war was not mentioned.

analysis that follows, I take Iowa's uniqueness into account when comparing mobility over time and attempt to contrast my findings not with mobility overall but with contemporary estimates of mobility in an Iowa-like sample.

In the online Appendix, I present summary statistics on the men in my linked sample, exploring how the fathers and the sons in the sample differ from the matched fathers and sons. The first two columns of Table A1 present summary statistics for the fathers of children between 3 and 17 in the Goldin–Katz Iowa State Census sample. Fathers found are the fathers for whom sons were located in the 1940 Census through Ancestry.com. Average yearly earnings for the fathers are approximately \$1,000 in 1915 dollars. The average father had a half year more than a common school education (eight years) and was approximately 42 years old in 1915. Of the fathers in my sample, those fathers for whom I matched a son into the 1940 Census earned very slightly more, though not significantly so, measured either in levels, logs, and weekly earnings. The final two columns of Table A1 present summary statistics for the Iowa sons in my sample. Only summary data for sons with complete information in the 1940 Federal Census is reported in the Table, which lowers the number of observations in the final column to 3,971. The located sons earned more than \$1,400 in 1940, which is roughly the same in real terms as the average earnings for their fathers in 1915; the lingering effects of the Great Depression may have reduced any real income gains overall.<sup>24</sup> Also notable 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) and Parman (2011).

## 2. Measures of Intergenerational Mobility

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

### 2.1. *Intergenerational Income Mobility*

The most common measure of intergenerational economic mobility is the intergenerational elasticity of income (IGE), estimated by regressing the son's log income against his father's log income. Solon (1999), Corak (2006) and Black and Devereux (2011)

<sup>24</sup> I measure all dollar amounts in this article 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.

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

present thorough reviews of the IGE literature.<sup>26</sup> These reviews all indicate a lack of historical data on intergenerational mobility: Corak (2006) documents 41 studies of the US IGE, none of which presents data before 1980. One aim of my project is to establish a correct measure of IGE well before the period previously studied in this literature.<sup>27</sup>

Corak's preferred measure of IGE in the US is 0.47, in line with the reviews presented by both Solon and Black and Devereaux.<sup>28</sup> Hertz (2009), who is able to split the sample by race, estimates an IGE of between 0.39 and 0.44 for whites, the relevant comparison here given is the high white share in my Iowa sample. Corak (2006) also documents large variations between US studies measuring the intergenerational elasticity of income.<sup>29</sup>

However, the IGE imposes a very particular functional form relationship between the son's and father's earnings.<sup>30</sup> Chetty *et al.* (2014a) present evidence from recent US tax data that suggests this assumption is false; Corak and Heisz (1999) show the same with Canadian tax data. At both tails of the income distribution, the linear relationship between the father's log income and the son's log income breaks down. Following Dahl and DeLeire (2008) and Chetty *et al.* (2014a, b) estimate a rank–rank parameter of intergenerational mobility, regressing the son's income percentile (within his cohort) against the father's income percentile (within his cohort). The graphical evidence presented in Chetty *et al.* (2014a) suggests that the implied linearity in the rank–rank specification is a more accurate fit of the data. For their sample of the US, Chetty *et al.* (2014a) estimate a rank–rank parameter of 0.341 overall and 0.336 for sons. However, similar to the IGE literature, there are few historical estimates of the

<sup>26</sup> The estimated elasticity of income between one generation and the next is commonly referred to as an IGE and I use that abbreviation here.

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

<sup>28</sup> Corak also gives lower and upper bounds of between 0.40 and 0.52. Mazumder (2015) argues for a parameter of 0.6.

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

<sup>30</sup> Specifically, a linear relationship between log father's income and log son's income.

rank–rank parameter: Chetty *et al.* (2014*b*) plot trends in mobility for sons born between 1971 and 1993 but they cannot extend their sample further back in time.

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

My study is not the first to use the earnings data in the Iowa 1915 census to estimate intergenerational mobility in the early twentieth century. Parman (2011) uses a clever multiple match technique to create father–son links within the Iowa 1915 census, even though the adult men in the Iowa sample are no longer in their parent’s households. Parman matches the men backwards in time to the 1900 Federal Census to construct childhood households. Though most are heads of household in 1915, these are Parman’s ‘sons’ in the analysis. The reconstructed households yield the name, state of birth, and other demographic characteristics of the ‘fathers’ in 1900. Parman then matches these fathers forward in time to the 1915 sample, thus observing both fathers and sons in Iowa in 1915 and estimating IGEs based on income reported in the Iowa State Census. Parman finds an IGE of approximately 0.11 for all father–son pairs or 0.17 for non-farmer father–son pairs. These low estimates paint a picture of high levels of mobility in the early twentieth century.<sup>31</sup>

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

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

<sup>31</sup> These estimates are similar but much lower than the income IGE estimates I present later in this article.

<sup>32</sup> Full access to the 1940 Federal Census, including all citizens’ names, was not available until April 2012, well after Parman completed his research.

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

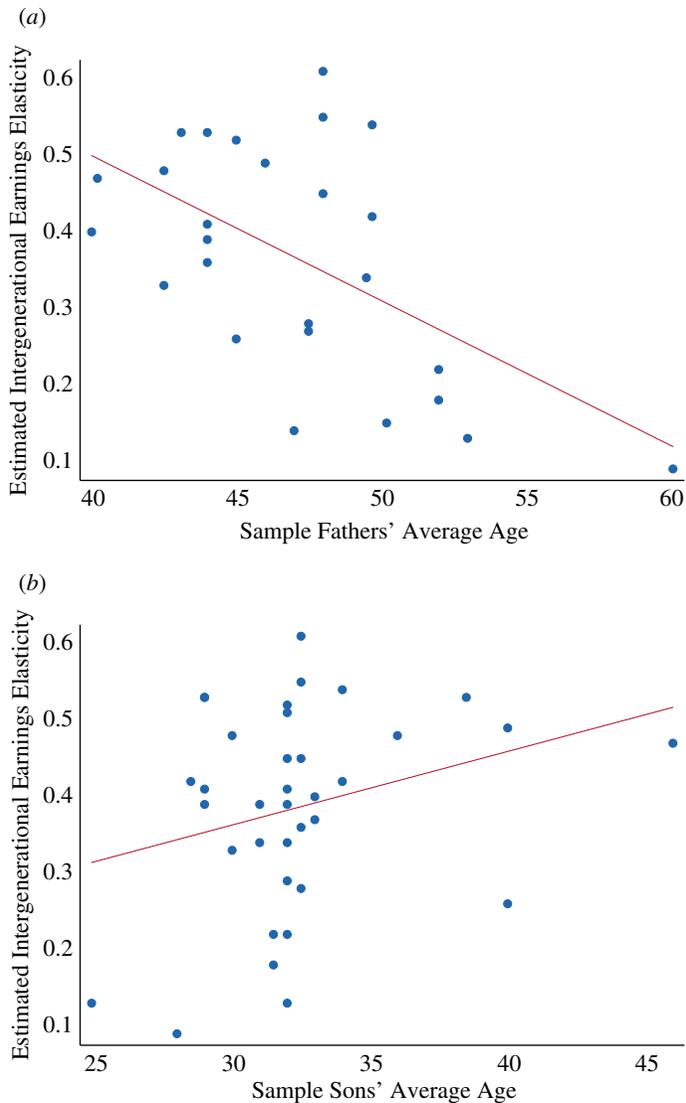


Fig. 2. *The Average Ages of Both Fathers and Sons Sampled May Bias the Estimated Intergenerational Mobility Elasticity*

Notes. Each point in the two scatterplots represent the estimated elasticity of income from studies of American intergenerational mobility reviewed by Corak (2006), plotted against the average age of either fathers or sons in the samples. (a) Correlation between estimated IGE and sample fathers' ages. (b) Correlation between estimated IGE and sample sons' ages. Colour Figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).

The second data-driven limitation is that the Parman (2011) sample is restricted to fathers and sons both living as adults in Iowa in 1915 because the 1940 Federal Census was not available until 2012. How large is the bias of this restriction, and in what direction does it push the intergenerational income mobility results? In the results

section that follows, I am able to use my sample to understand the magnitude and direction of the problem better by limiting my sample to only those sons who still live in Iowa in 1940. To preview, I find that the bias is large and negative: selecting the sample using only sons remaining in Iowa in adulthood reduces both the measured IGE and rank–rank parameters.<sup>34</sup> Because the state of residence in adulthood is, in part, jointly determined with the outcome of interest (earnings), controlling for it or splitting the sample based on it is problematic. In order to estimate an accurate IGE parameter for the early twentieth century, one needs to be able to observe sons that remain in their father’s state of residence and sons that move elsewhere.<sup>35</sup>

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

## 2.2. *Intergenerational Education Mobility*

Due to data constraints, there has been little work on educational mobility in the US historically.<sup>36</sup> Hertz *et al.* (2008) 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).<sup>37</sup> With data collected on years of education in 1915 in Iowa and in 1940 nationally, I am able to construct the earliest educational mobility estimates for the United States. In addition, because education is significantly less noisy as a proxy for status than a single year of earnings, the educational mobility results can serve as a check on mobility based on earnings, which may be biased down due to noise.<sup>38</sup>

## 2.3. *Intergenerational Occupational Mobility*

The study of historical intergenerational mobility has focused on the study of occupational mobility because occupational data has historically been much more available. Early work on this topic was undertaken by Thernstrom (1964, 1973), studying the occupations of successive generations in Boston and Newburyport, MA. Thernstrom tends to find quite high upward mobility, but a lot of white collar stability as well. Duncan (1965) finds more upward and less downward mobility in 1962 relative to the occupational transition matrices of 1952, 1942 or 1932, relying on data gathered

<sup>34</sup> These findings are in the final rows of both panels (a) and (b) in Table 4.

<sup>35</sup> 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.

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

<sup>37</sup> 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.

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

from occupational changes in a generation (OCG). However, neither Duncan nor any of the subsequent work based entirely on the OCG data is able to measure occupational mobility for earlier periods.

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

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

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

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

One of the strongest criticisms of using occupations to study long-term trends in intergenerational mobility is the difficulty in classifying farmers (Xie and Killewald, 2013).<sup>40</sup> Comparison of mobility measures across time is complicated – perhaps even driven – by the movement out of agriculture in the US. For example, Guest *et al.* (1989) conclude that there was more social mobility in the post-WWII period in the US than there had been in the nineteenth century, but they suggest that this reflects the high-heritability of farming and the declining shares of farmers since the late nineteenth century. In this article, I attempt to control for that by comparing relatively homogeneous samples over time, particularly by constructing a sample in the PSID or other contemporary data that is as rural, white, and agricultural as my Iowa sample.

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

<sup>40</sup> 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.

I also focus more analysis on the urban Iowa sons, almost none of whom had farmers for fathers or became farmers themselves.

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

#### 2.4. *Name-based Intergenerational Mobility*

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

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

Surnames also convey information about the previous generation. In recent research, Greg Clark and coauthors use rare last names to measure mobility, both between parents and children (Clark and Cummins, 2015), as well as in the very long run (Clark, 2014). Guell *et al.* (2015) formalise such a method of measuring mobility based on surnames and show declining mobility in Catalonia over the century. One major advance with the rare surname method is its flexibility: whether based on admissions lists from elite academic institutions or probate records or voting records, any data containing both last names and a measure of status can be used to estimate mobility. Clark (2014) argues that these measures tend to suggest far less mobility than one generation income or occupational status mobility measures. Studying mobility based on the over or under-representation among doctors or lawyers of historically

prestigious surnames in the US, Clark *et al.* (2012) find advantages persisting for at least five generations. While my two-generation sample does not permit such long run estimates of mobility, I am able to apply Clark's method to the sons of Iowa to estimate the changing shares of elite surnames in certain occupations, focusing on doctors and lawyers as in Clark *et al.* (2012).<sup>41</sup>

### 3. Intergenerational Mobility Estimates for Iowa from 1915 to 1940

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

I begin by presenting raw intergenerational correlations of fathers' and sons' outcomes in Figure 3: log annual income (3), years of education (3), and occupation score based on 1950 scores (3) and on 1915 scores (3).<sup>42</sup> The intergenerational correlation is one possible measure of mobility and holds trends in the marginal distributions – the dispersion of earnings among fathers and among sons – fixed. In log earnings, the correlation is 0.14; the education correlation is a bit higher (0.24).<sup>43</sup> Clearly, there is a strong positive relationship between outcomes for fathers in 1915 and sons in 1940 but the exact measure of the respective slopes of these lines – and how those slopes compare across the various measures of mobility, as well as with the estimates of mobility in the recent period – is the key question of this article.

#### 3.1. Intergenerational Mobility of Earnings

I measure mobility of earnings in two primary ways. First, following the intergenerational mobility literature, I use intergenerational elasticities (IGE) (Solon, 1999; Corak, 2006; 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,

<sup>41</sup> Guell *et al.* (2015) develop an alternative metric for deriving the informational content of surnames.

<sup>42</sup> Naturally, there are many father–son pairs with the same outcome levels as other pairs. In an attempt to display this density at certain points on the graphs, I have used both hollow scatterplot markers and jittered the data. The best fit lines are, of course, drawn based on the full sample before jittering. The \$5,000 top coding in the 1940 census is apparent on the right side of Figure 3(a).

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

either (log) income or education. The model I estimate can be summarised as follows:

$$Y_{i,1940}^s = \alpha + \beta \times Y_{i,1915}^f + Q^s(\text{age}_{i,1915}^s) + Q^f(\text{age}_{i,1915}^f) + Q^s(\text{age}_{i,1915}^s) \times Y_{i,1915}^f + \epsilon_i, \quad (2)$$

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

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

To reduce any measurement error induced by life cycle income effects (Grawe, 2006), I follow Lee and Solon (2009) and include quartic age controls for both the father and the son, defined as  $Q^s$  and  $Q^f$  above, as well as an interaction between the son’s age and the father’s outcome. In the interaction term, I normalise son’s age in 1940 relative to age 40 (Haider and Solon, 2006).<sup>44</sup> The fact that I define my sample to observe sons between ages 28 and 42 in the 1940 Census also reduces life cycle driven measurement error. As some of my observed sons are brothers, I cluster standard errors at the family level. I also include an Iowa 1915 county fixed effect, subsuming an urban or rural control. The results are robust to the inclusion of controls for family-size effects, county fixed effects, and the name string control variables described previously.<sup>45</sup>

Panel (a) of Table 4 presents my estimates of the IGE of income across a variety of samples. Both the father’s and son’s incomes are measured as annual log earnings.<sup>46</sup> The first specification is a simple univariate regression of the son’s log earnings on the father’s log earnings. In specification two, I include controls for name string properties

<sup>44</sup> With this normalisation, the estimated  $\beta$  represents the relationship between son’s and father’s outcomes when the son is age 40. I follow Lee and Solon (2009) in normalising to 40.

<sup>45</sup> The county fixed effects indicate the county of residence when the son is observed in Iowa in 1915. The name string controls include first and last name commonness, length, letter similarity and Scrabble scores, all attempts to control for differential matching rates between the 1915 and 1940 censuses. While I include these various controls to reduce measurement error, both Chetty *et al.* (2014*a*) and Nybom and Stuhler (2014*a*) 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.* (2014*a*) argue that estimates are stable even with just one year of income observed for both fathers and sons, though this is still an unresolved issue in the literature. Further, they document that the exact age when fathers or sons are observed has very little effect on the measurement of mobility, so long as the fathers are observed between the ages of 30 and 55 and the sons are observed after age 30. Nybom and Stuhler (2014*a*) 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.

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

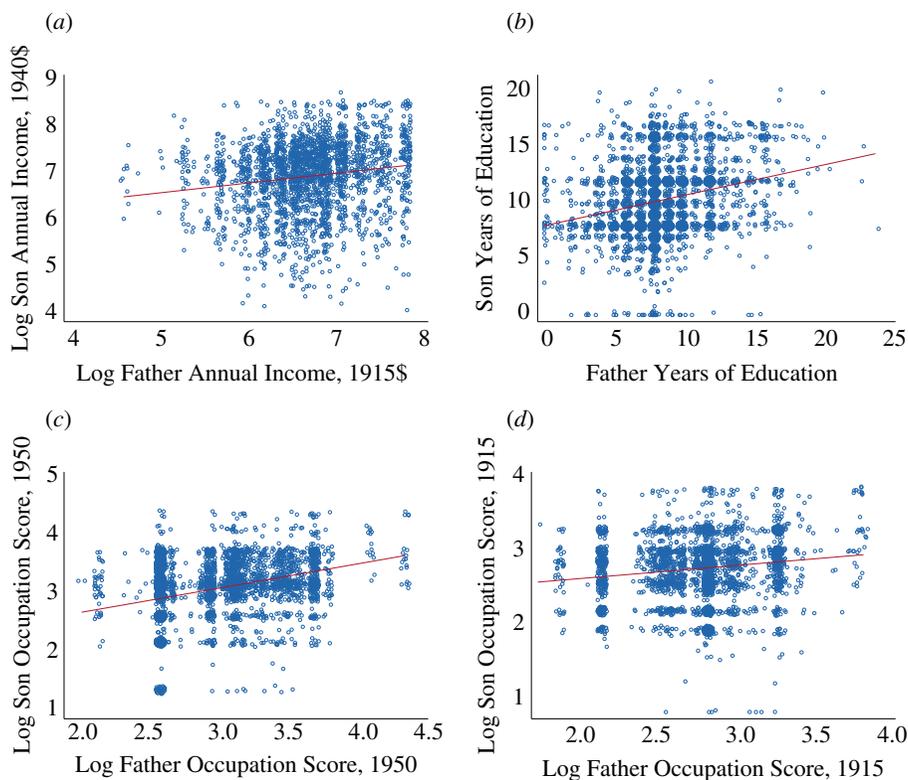


Fig. 3. *Intergenerational Mobility for Iowa 1915 to 1940. (a) Mobility of Income; (b) Mobility of Education; (c) Mobility of Occupation, 1950 Basis; (d) Mobility of Occupation, 1915 Basis*  
*Notes.* Colour Figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).

that might affect matching, 1915 county of residence fixed effects, and quartic controls in father and son age. In the third specification, I also include an interaction between son's normalised age and father's log earnings to control for life cycle measurement error (Grawe, 2006; Haider and Solon, 2006).

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

Similar to my IGE results, I find much more income mobility historically than today. The rank–rank parameter ranges from 0.167 to 0.217 in the first row of panel (b) of

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

Table 4  
*Intergenerational Mobility Estimates*

	Specification			Observations	Clusters
	(1)	(2)	(3)		
Panel (a): intergenerational elasticity (IGE)					
Full sample	0.208 (0.032)	0.199 (0.031)	0.258 (0.081)	2,039	1,668
PSID Iowa-like sample	0.330 (0.056)	0.350 (0.080)	0.502 (0.166)	3,449	346
Urban sample	0.287 (0.045)	0.275 (0.050)	0.310 (0.102)	1,004	824
Rural sample	0.156 (0.040)	0.167 (0.041)	0.233 (0.113)	1,035	844
Excluding sons of farmers	0.301 (0.037)	0.259 (0.038)	0.391 (0.095)	1,452	1,200
Only native grandparents sample	0.287 (0.045)	0.277 (0.046)	0.411 (0.116)	714	600
Only foreign grandparents sample	0.145 (0.059)	0.159 (0.059)	0.160 (0.123)	641	516
Including sons with imputed income	0.162 (0.021)	0.156 (0.022)	0.217 (0.050)	2,864	2,161
All sons, imputed 1950 income	0.066 (0.010)	0.066 (0.010)	0.061 (0.019)	3,144	2,324
Sons remaining in Iowa	0.147 (0.043)	0.148 (0.043)	0.173 (0.121)	1,184	1,001
Panel (b): intergenerational rank–rank parameter					
Full sample	0.172 (0.022)	0.167 (0.021)	0.217 (0.045)	2,039	1,668
PSID Iowa-like sample	0.258 (0.049)	0.240 (0.064)	0.323 (0.084)	3,680	356
Urban sample	0.217 (0.032)	0.206 (0.034)	0.214 (0.071)	1,004	824
Rural sample	0.141 (0.028)	0.150 (0.028)	0.224 (0.060)	1,035	844
Excluding sons of farmers	0.230 (0.026)	0.201 (0.027)	0.254 (0.058)	1,452	1,200
Only native grandparents sample	0.240 (0.034)	0.224 (0.035)	0.312 (0.078)	714	600
Only foreign grandparents sample	0.140 (0.041)	0.159 (0.038)	0.178 (0.081)	641	516
Including sons with imputed income	0.140 (0.016)	0.136 (0.016)	0.181 (0.030)	2,864	2,161
All sons, imputed 1950 rank	0.046 (0.009)	0.046 (0.009)	0.037 (0.017)	3,266	2,417
Sons remaining in Iowa	0.126 (0.028)	0.132 (0.027)	0.187 (0.058)	1,184	1,001

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

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

Table 4, the main sample with all linked father–son pairs. Chetty *et al.* (2014a) measure a rank–rank parameter of 0.341; among just male children, they find a rank–rank estimate between 0.307 and 0.317.

However, any measurement error will tend to bias down estimates of intergenerational mobility (Solon, 1999). Further, though Iowa is in some ways broadly representative of the US in 1915 – with respect to urban and rural shares – the differences in my estimated mobility may reflect differences between Iowa and the rest of the country, not differences between time periods. In fact, according to contemporary data, children born in Iowa are among the most economically mobile in the entire country, across many measures (Chetty *et al.*, 2014a).

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

To calculate a comparable contemporary IGE, rather than follow Lee and Solon (2009) and measure the father’s income as the average of his income when the matched son is between 15 and 17 years old, I use the father’s income when his son is 10.<sup>49</sup> 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 year 2000\$.<sup>50</sup> 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.<sup>51</sup> Though the Iowa-like samples in the PSID are small and contain repeated observations of the same set of father–son pairs over many years, the results suggest that the lower mobility I find historically is driven neither by the demographic composition of my data nor by the single-year measurements of income. I come to a similar conclusion when comparing my estimates of rank–rank mobility with contemporary measures. The first comparison, shown in the second row of panel (b) of Table 4, is the rank–rank mobility using the PSID sample. For the Midwest sample, I measure a lower parameter than is found nationally, suggesting the Midwest is more mobile; however, these estimates are far larger than what I find historically in the full sample.

As a second alternative construction of comparable recent mobility parameters, I use the county level results reported by Chetty *et al.* (2014a). Unfortunately, such a

<sup>48</sup> 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.

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

<sup>50</sup> 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.

<sup>51</sup> Specification 3 includes interactions with father’s outcomes and son’s ages. However, the PSID includes a small number of father–son pairs observed repeatedly which may complicate the interpretation of this estimate. Specification 1 and 2 are preferred and more conservative as comparisons.

comparison can only be made for the rank–rank parameter, as Chetty *et al.* (2014a) do not calculate IGE at any disaggregated geographies.<sup>52</sup> When I calculate the weighted average of rank–rank mobility, weighing by the shares of my sample living in each county in 1915, I find a rank–rank parameter of 0.303, similar to the result from the contemporary PSID data and, more importantly, far larger than the rank–rank parameter of 0.169 to 0.219 that I find in my historical sample.<sup>53</sup>

In the online Appendix, I show that neither false matches in my census linking procedure nor higher levels of measurement error in historical data could account for my estimates of lower IGE parameters (and thus higher mobility) in the 1915 to 1940 sample relative in the contemporary sample. Via simulation, I introduce both mismatches and measurement error into my Iowa-like PSID sample considered above. The share of false matches would have to approach 50% for mismatching to account for the estimated differences in IGE parameters, which seems highly unlikely.<sup>54</sup> 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 recent period to generate the large difference in IGEs.

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

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

<sup>53</sup> I can also split the sample between the urban and rural counties in my analysis. The weighted average of rank–rank mobility is 0.355 in the three urban counties and 0.268 in the 10 rural counties.

<sup>54</sup> 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%.

<sup>55</sup> Or are the results driven by the difficulty of accurately measuring income for farmers?

<sup>56</sup> 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.

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

I also find that the grandchildren on the foreign-born have more mobility, as shown in the sixth and seventh rows in Table 4. Drawing on data on each son’s grandparent’s place of birth in the 1915 census, I partition the sample into sons with four native-born grandparents and with four foreign-born grandparents.<sup>58</sup> This corresponds to the high rate of upward mobility Perez (2016b) finds in Argentina in the nineteenth century.

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

<sup>57</sup> This sample is made up of the urban sample and nearly half of the rural sample.

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

<sup>59</sup> As noted previously, only wage and salary earnings were recorded in the census. The data collected are the ‘total amount of money wages or salary’ and enumerators were instructed: ‘Do not include the earnings of businessmen, farmers, or professional persons derived from business profits, sale of crops, or fees’. For more, see <https://usa.ipums.org/usa/voliii/inst1940.shtml#584>.

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

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

row 8, in row 9 I impute total ‘1950’ earnings for every son.<sup>62</sup> This exercise suffers slightly with a difficult imputation as the variables I observe in the 1950 1% sample explain just under half of the variation in log total earnings in 1950. Nevertheless, I estimate a very low rate of persistence in both panels (a) and (b).

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

What is the probability that a son born in a given quintile will be in the same or another quintile in 1940? Panel (a) of Table 5, an income quintile transition matrix,

Table 5  
*Income Rank Quintile Transition Matrix*

Son’s quintile	Father’s quintile					Total
	1	2	3	4	5	
Panel (a): Iowa 1915 to 1940 earnings quintile transition						
1	15.0	7.2	8.7	10.0	7.2	9.2
2	23.1	22.7	15.7	17.9	15.1	18.6
3	27.7	24.3	25.4	19.1	18.1	22.8
4	20.5	26.5	24.2	24.0	23.3	24.3
5	13.8	19.3	26.1	29.1	36.2	25.1
Total	17.6	21.3	19.7	20.7	20.7	
Panel (b): contemporary Iowa earnings quintile transition						
1	26.1	15.9	10.8	7.7	7.1	12.1
2	23.0	18.9	15.0	12.0	9.9	15.1
3	20.2	21.5	21.1	19.1	16.1	19.7
4	17.7	23.6	26.5	27.8	25.6	25.2
5	12.9	20.0	26.6	33.5	41.3	27.9
Total	12.3	19.1	26.8	25.1	16.7	

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

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

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

can be used to answer such questions. Each cell represents the probability that a son with a father in a given income quintile (identified by the column) will be in a given income quintile in 1940 (given by the row). A son whose father is in the bottom quintile in 1915 has only a 13.8% chance of being in the top quintile in 1940, while the odds that a son born in the top quintile remains there in 1940 are 36.2%. What is the likelihood a son falls into the bottom quintile? Not surprisingly, those odds fall with father's rank: a son born in the top quintile has only a 7.2% chance of being in the bottom quintile in 1940, while the odds are more than twice as high (15.0%) that a son born in the bottom quintile remains there. While the Table clearly presents a degree of intergenerational immobility, when these percentages are compared to a transition matrix for the recent period (sons born in the 1980–2 cohorts) from Chetty *et al.* (2014a), there is in fact more mobility between 1915 and 1940. In panel (b) of Table 5, I calculate the quintile transition matrix for Iowa today, weighting the commuting zone level results from Chetty *et al.* (2014a) by the share of 1915 sons in each county. Sons born in the bottom quintile are ten points more likely to remain there as adults in the contemporary data than historically: 26.1% to 15.0%. Historically, sons born in the second quintile have a 70% chance of being in a higher quintile in adulthood; that probability is only 65% in contemporary data. At the top end, however, transition probabilities are similar between the two periods: sons born in the fourth or fifth quintiles have 24.0% and 36.2% probability of being in those quintiles as adults in the Iowa sample, compared with 27.8% and 41.3% probabilities in the Iowa 1915 weighted Chetty *et al.* (2014a) sample.

### 3.2. *Alternative Measures of Intergenerational Mobility*

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

In addition to serving as a (potentially) more accurately measured check on my income results, the education persistence estimates are a valuable and important historical parameter. I estimate the intergenerational mobility of education and present these results in panel (a) of Table 6.<sup>64</sup> 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 latter half of the nineteenth century and then through the high school movement in the early twentieth century, made education widely available and free (Goldin and Katz, 2008, 2011). Goldin and Katz (2008) also argue that this increase in human capital helped spur national growth and prosperity in the following century. Whether this massive public investment in education also reduced the link between a son's and his father's educational outcomes can help scholars understand the role of public programmes in shaping or changing inequality.

<sup>64</sup> The education persistence regressions are estimated in levels rather than logs and so they are not true intergenerational elasticities.

Table 6  
*Alternative Intergenerational Mobility Estimates*

Sample	Specification			Observations	Clusters
	(1)	(2)	(3)		
Panel (a): education mobility					
<i>Years of education</i>					
Full sample	0.264 (0.023)	0.241 (0.023)	0.206 (0.040)	3,376	2,504
Urban sample	0.297 (0.035)	0.269 (0.035)	0.301 (0.058)	1,283	1,028
Rural sample	0.235 (0.029)	0.223 (0.030)	0.142 (0.056)	2,093	1,476
Excluding sons of farmers	0.317 (0.027)	0.303 (0.028)	0.338 (0.046)	2,004	1,589
Only native grandparents sample	0.323 (0.036)	0.313 (0.036)	0.308 (0.062)	1,117	859
Only foreign grandparents sample	0.161 (0.050)	0.154 (0.047)	0.076 (0.080)	1,139	819
Panel (b): 1915 occupation score mobility					
<i>Log occupation score 1915 basis</i>					
Full sample	0.190 (0.026)	0.184 (0.026)	0.265 (0.052)	3,037	2,279
Urban sample	0.229 (0.035)	0.226 (0.036)	0.308 (0.065)	1,154	940
Rural sample	0.145 (0.039)	0.133 (0.038)	0.175 (0.091)	1,883	1,339
Excluding sons of farmers	0.203 (0.026)	0.204 (0.026)	0.288 (0.053)	1,760	1,414
Only native grandparents sample	0.242 (0.040)	0.242 (0.040)	0.412 (0.084)	1,002	776
Only foreign grandparents sample	0.123 (0.048)	0.130 (0.048)	0.103 (0.081)	1,038	756
Panel (c). 1950 occupation score mobility					
<i>Log occupation score 1950 basis</i>					
Full sample	0.441 (0.021)	0.366 (0.024)	0.386 (0.045)	3,202	2,374
Urban sample	0.258 (0.036)	0.247 (0.036)	0.291 (0.068)	1,220	980
Rural sample	0.424 (0.030)	0.418 (0.030)	0.401 (0.070)	1,982	1,394
Excluding sons of farmers	0.229 (0.027)	0.220 (0.027)	0.263 (0.061)	1,867	1,479
Only native grandparents sample	0.351 (0.034)	0.298 (0.037)	0.302 (0.081)	1,053	807
Only foreign grandparents sample	0.505 (0.037)	0.361 (0.044)	0.369 (0.075)	1,084	780

*Notes.* 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 panels (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.

The literature suggests an IGE of education of 0.46 (Hertz *et al.*, 2008). As Table 6 shows, I find a much lower IGE parameter for schooling, between 0.206 and 0.264. This suggests that, like mobility measures based on income, educational mobility in the US was higher in the early twentieth century than it is today.<sup>65</sup> Similar to the results presented on income mobility, there is more mobility of education among rural sons than urban sons as well, though these differences are not always statistically significant. Finally, the last two rows of panel (a) in Table 6 are also in agreement with the earnings mobility results: the persistence of educational attainment is much lower among sons with foreign grandparents than sons with native grandparents.

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

When measured with occupation score, mobility was quite high in the early twentieth century, corroborating my findings with income and education.<sup>68</sup> I present the occupation mobility results in panels (b) and (c) of Table 6. I estimate an IGE parameter between 0.184 and 0.265 for the occupational score measure based on income data from 1915 Iowa. This is higher than my IGE estimate for income but lower than my IGE estimate for education. However, when I use the 1950-based occupation score measure, I find significantly larger IGE estimates, indicating less mobility. These measures are higher than any previous IGEs estimated in this article, suggesting only slightly more mobility in the early twentieth century than today, if

<sup>65</sup> 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.

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

<sup>67</sup> In the online 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.

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

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

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

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

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

I find similarly high rates of mobility when using surname-based measures, following the work Greg Clark and others.<sup>70</sup> Rare surnames provide another form of pseudo-

<sup>69</sup> Olivetti and Paserman (2015) note this concordance as well in footnote 29, page 2719.

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

Table 7  
*Intergenerational Mobility Based on First Names*

Sample	Specification			Observations	Clusters
	(1)	(2)	(3)		
Panel (a): log earnings					
<i>First name based father's occupation score</i>					
Full sample	0.300 (0.102)	0.177 (0.103)	0.404 (0.250)	2,942	2,248
Urban sample	0.314 (0.170)	0.333 (0.172)	0.433 (0.470)	1,156	945
Rural sample	0.123 (0.126)	0.094 (0.129)	0.367 (0.309)	1,786	1,303
Excluding farmer sons	0.382 (0.134)	0.309 (0.136)	0.741 (0.416)	1,792	1,457
Only native grandparents sample	0.310 (0.182)	0.260 (0.185)	0.157 (0.510)	950	758
Only foreign grandparents sample	0.279 (0.190)	0.122 (0.190)	0.324 (0.408)	1,000	743
Panel (b): 1950 occupation score mobility					
<i>First name based father's occupation score</i>					
Full sample	0.353 (0.074)	0.178 (0.076)	0.318 (0.148)	3,158	2,372
Urban sample	0.171 (0.106)	0.184 (0.112)	0.120 (0.216)	1,205	977
Rural sample	0.164 (0.097)	0.183 (0.102)	0.430 (0.206)	1,953	1,395
Excluding farmer sons	0.160 (0.083)	0.165 (0.087)	0.190 (0.187)	1,867	1,498
Only native grandparents sample	0.237 (0.121)	0.060 (0.127)	0.155 (0.319)	1,032	804
Only foreign grandparents sample	0.638 (0.123)	0.306 (0.118)	0.546 (0.214)	1,078	787

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

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

linking between generations, because men with a very rare surname in one generation are likely to be the fathers of men with that same rare surname in the next generation.

I start by locating rare, elite surnames in Iowa in the complete count 1920 Federal Census.<sup>71</sup> For surnames held by 10 to 30 adult men, I find the names overrepresented

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

Table 8  
*Occupation Transitions Table for Fathers and Sons, 1915 to 1940*

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

*Notes.* 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.

among doctors and lawyers, the two high status groups studied by Clark in the US. Following Clark, I define the relative representation of a surname  $k$  among doctors as:

$$RR_k^{\text{doctors}} = \frac{\text{doctors}_k \div \text{doctors}_{\text{total}}}{\text{population}_k \div \text{population}_{\text{total}}}, \quad (3)$$

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

How does this persistence of high occupational status compare to the findings from Clark (2014) in the US? Ashkenazi Jewish names had a similarly high prevalence in the first generation from 1920 to 1949 among doctors (between 6.73 to 13.64 depending on assumptions about their share of the population), but still had relative representations of more than 5 three generations later (from 1980 to 2009). Two other elite groups studied by Clark et al. (2012) were the 1920s Rich (high earners whose tax bills were listed in the 1923 and 1924 newspapers) and Ivy League graduates from 1650 to 1850. Neither are as prevalent among doctors in the early twentieth century as the Ashkenazi Jews – relative representation ranging from 2.97 to 5.86 – and both regressed to rates in the same range as the sons of elite Iowa a generation later. My results suggest that there was likely more mobility in Iowa than among the samples Clark examines in the US but an exact comparison is complicated by the differences in sources.

### 3.3. *Mobility According to Occupation Transitions*

Similar to income, educational and occupation score mobility, there appears to be more broad-category occupational mobility during my period of study than there is

<sup>72</sup> Names include 'Alt', 'Appel', 'Willett' and 'Woodhouse'. None of the names in the list are especially noteworthy. This contrasts to the elite names Clark et al. draw from the high income tax payer lists in 1923 and 1924 which includes Vanderbilt, Roosevelt, and Winthrop, but accords with the names Clark et al draw from the Ivy League graduate database from 1650 to 1850 (Clark, 2014, p. 4: 'only a very few have any resonance').

today. I measure occupational mobility using Altham statistics (Altham and Ferrie, 2007). My occupational transition table from 1915 to 1940 is presented in Table 8. Occupations are categorised by linking occupational strings (exactly as entered by the census enumerators) to the 1940 occupational code charts for both the 1915 father and 1940 son samples.<sup>73</sup> Sample sizes are different from previous portions of the analysis because not all occupation strings could be matched to occupation groups.

The distance between the occupation transition table and the identity table,  $I$ , can be thought of as a measure of occupational immobility – the larger the distance, the more likely it is that sons enter the same occupational class as their fathers. Calculating the Altham statistic for Table 8 yields  $d(IA1940, I) = 16.14$ . Long and Ferrie (2007) report  $d(US1880, I) = 12.09$ ,  $d(US1900, I) = 14.58$  and  $d(US1973, I) = 20.76$ . The Altham statistic generated by my linked sample of fathers and sons from 1915 to 1940 is larger – thus indicating less mobility – than the measures presented by Long and Ferrie (2007) for the nineteenth century and statistically significantly different from these historical measures as well. These results are summarised in Figure 4 and in the first column of Table 9. In addition, echoing the results presented previously suggesting more mobility historically than today, there appears to be more mobility between 1915 and 1940 than between 1950 and 1976 – as compared to the more contemporary estimates reported by Long and Ferrie.

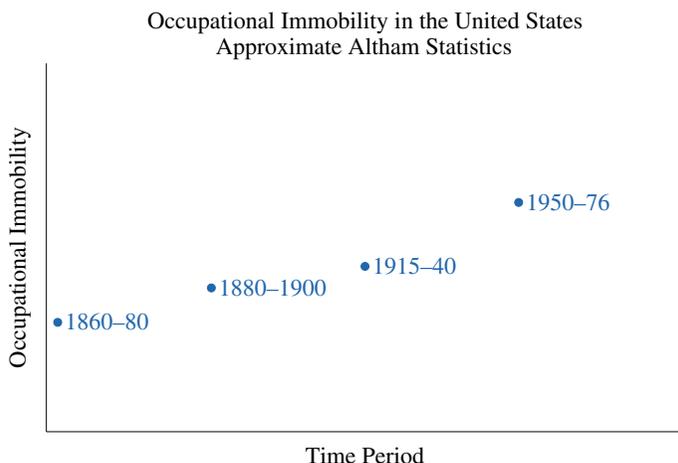


Fig. 4. Occupational Immobility Measured by Altham Statistics from 1860 to 1976

Notes. This graph suggests that occupational immobility rose over time and that there is less mobility in the mid-twentieth century than there was between 1915 and 1940. The Altham statistics for each period presented in the plot are statistically different from one another. See Table 9 for exact values and distance tests. Colour Figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).

<sup>73</sup> 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 9

*Altham Statistic Summary, Iowa 1915 to 1940*

	Sample ( <i>P</i> )		
	All sons	Urban sons	Rural sons
Altham statistic ( $d(P, D)$ )	16.14	10.74	14.34
$\Pr(d(P, US1880) = 0)$	0.000	0.627	0.141
$\Pr(d(P, US1900) = 0)$	0.000	0.017	0.511
$\Pr(d(P, US1973) = 0)$	0.000	0.003	0.004
$\Pr(d(P, IAsons) = 0)$	–	1.000	0.718
$\Pr(d(P, IAurbansons) = 0)$	0.718	–	0.786
$\Pr(d(P, IAurwalsons) = 0)$	0.711	0.786	–

*Notes.* 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).

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

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

#### 4. Conclusion

This article presents estimates of intergenerational mobility for men born in Iowa between 1900 and 1910 across a variety of measures. Mobility based on earnings, education, occupation and the socio-economic content of names all roughly align, suggesting that this was a period of high mobility and that there was less persistence in status for rural sons than urban sons and for the grandsons of the foreign-born than

<sup>74</sup> 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 be an undesirable property of Altham statistics. In theory, an alternative statistic might possess a form of continuity and the intermediate value theorem.

<sup>75</sup> This difference is the reverse of what I found with respect to both income and educational mobility in the previous Section.

grandsons of the native-born. The results also suggest that the mobility rates for this generation were higher than those for men born since 1960. Because many of the challenges of estimating mobility historically such as imperfect census linking and observing only a single year of earnings all serve to reduce the estimated persistence parameter and inflate mobility, any comparison between eras is not conclusive but my results are another piece of evidence in favour of the theory that mobility was higher in the early twentieth century than today, echoing the findings of Long and Ferrie (2013), Parman (2011) and others.

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

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

<sup>76</sup> Parman (2011) shows how education in Iowa might have affected mobility in the previous generation.

non-farmers in both urban and rural Iowa. Variation in the access to transport or differences in inequality across Iowa could also play an important role.<sup>77</sup>

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

*Boston University*

Additional Supporting Information may be found in the online version of this article:

**Appendix A.** Additional Material.

**Data S1.**

## References

- Aaronson, D. and Mazumder, B. (2008). 'Intergenerational economic mobility in the United States, 1940 to 2000', *Journal of Human Resources*, vol. 43(1), pp. 139–72.
- Abramitzky, R., Boustan, L.P. and Eriksson, K. (2012). 'Europe's tired, poor, huddled masses: self-selection and economic outcomes in the age of mass migration', *American Economic Review*, vol. 102(5), pp. 1832–56.
- Abramitzky, R., Boustan, L.P. and Eriksson, K. (2013). 'Have the poor always been less likely to migrate? Evidence from inheritance practices during the age of mass migration', *Journal of Development Economics*, vol. 102, pp. 2–14.
- Ager, P., Boustan, L.P. and Eriksson, K. (2016). 'Inter-generational transmission of wealth shocks: evidence from the US Civil War, Working Paper'.
- Altham, P.M.E. and Ferrie, J.P. (2007). 'Comparing contingency tables tools for analyzing data from two groups cross-classified by two characteristics', *Historical Methods*, vol. 40(1), pp. 3–17.
- Becker, G.S. and Tomes, N. (1979). 'An equilibrium theory of the distribution of income and intergenerational mobility', *Journal of Political Economy*, vol. 87(6), pp. 1153–89.
- Biavaschi, C., Giulietti, C. and Siddique, Z. (forthcoming). 'The economic payoff of name Americanization', *Journal of Labor Economics*, <https://doi.org/10.1086/692531>.
- Bjorklund, A., Roine, J. and Waldenstrom, D. (2012). 'Intergenerational top income mobility in Sweden: capitalist dynasties in the land of equal opportunity?', *Journal of Public Economics*, vol. 96(5–6), pp. 474–84.
- Black, S.E. and Devereux, P.J. (2011). 'Recent developments in intergenerational mobility', in (D. Card and O. Ashenfelter), *Handbook of Labor Economics*, volume 4, pp. 1487–541, North Holland: Elsevier B.V., volume 4b edition.
- Checchi, D., Fiorio, C.V. and Leonardi, M. (2013). 'Intergenerational persistence of educational attainment in Italy', *Economics Letters*, vol. 118(1), pp. 229–32.
- Chetty, R., Hendren, N., Kline, P. and Saez, E. (2014a). 'Where is the land of opportunity? The geography of intergenerational mobility in the United States', *Quarterly Journal of Economics*, vol. 129(4), pp. 1553–623.
- Chetty, R., Hendren, N., Kline, P., Saez, E. and Turner, N. (2014b). 'Is the United States still a land of opportunity? Recent trends in intergenerational mobility', *American Economic Review Papers & Proceedings*, vol. 104(5), pp. 141–7.

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

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

- Clark, G. (2014). *The Son Also Rises: Surnames and the History of Social Mobility*, Princeton, NJ: Princeton University Press.
- Clark, G. and Cummins, N. (2015). 'Intergenerational wealth mobility in England, 1858–2012: surnames and social mobility', *Economic Journal*, vol. 125(582), pp. 61–85.
- Clark, G., Marcin, D., Abu-sneh, F., Chow, W.M., Mo, K., Marek, A.M. and Williams, K.M. (2012). 'Social mobility rates in the USA, 1920–2010: A surname analysis', Working Paper. Available at: <http://old.econ.ucdavis.edu/faculty/gclark/papers/Social%20Mobility%20Rates%20in%20the%20USA.pdf> (last accessed: 24 October 2016).
- Corak, M. (2006). 'Do poor children become poor adults? Lessons from a cross country comparison of generational earnings mobility', in (J. Creedy and G. Kalb eds.), *Dynamics of Inequality and Poverty (Research on Economic Inequality, Volume 13)*, pp. 143–88. Bingley, UK: Emerald Group Publishing Limited.
- Corak, M. and Heisz, A. (1999). 'The intergenerational earnings and income mobility of Canadian men: evidence from longitudinal income tax data', *Journal of Human Resources*, vol. 34(3), pp. 504–33.
- Dahl, M.W. and DeLeire, T. (2008). 'The association between children's earnings and fathers' lifetime earnings: estimates using administrative data', available at: <http://irp.wisc.edu/publications/dps/pdfs/dp134208.pdf> (last accessed: 24 October 2016).
- Duncan, O.D. (1965). 'The trend of occupational mobility in the United States', *American Sociological Review*, vol. 30(4), pp. 491–8.
- Feigenbaum, J.J. (2016a). 'Automated census record linking: a machine learning approach', available at: <http://scholar.harvard.edu/files/jfeigenbaum/files/feigenbaum-censuslink.pdf> (last accessed: 24 October 2016).
- Feigenbaum, J.J. (2016b). 'Intergenerational mobility during the great depression', available at: <http://james-feigenbaum.github.io/research/pdf/jmp.pdf> (last accessed: 24 October 2016).
- Gentzkow, M., Shapiro, J.M. and Sinkinson, M. (2011). 'The effect of newspaper entry and exit on electoral politics', *American Economic Review*, vol. 101(7), pp. 2980–3018.
- Goldin, C. and Katz, L.F. (2000). 'Education and income in the early twentieth century: evidence from the Prairies', *Journal of Economic History*, vol. 60(3), pp. 782–818.
- Goldin, C. and Katz, L.F. (2008). *The Race Between Education and Technology*, Cambridge, MA: The Belknap Press of Harvard University Press.
- Goldin, C. and Katz, L.F. (2011). 'Mass secondary schooling and the state: the role of state compulsion in the high school movement', in (D.L. Costa and N.R. Lamoreaux eds.), *Understanding Long Run Economic Growth*, pp. 275–310, Chicago, IL: University of Chicago Press.
- Goldin, C. and Margo, R.A. (1992). 'The great compression: the wage structure in the United States at mid-century', *Quarterly Journal of Economics*, vol. 107(1), pp. 1–34.
- Grawe, N.D. (2006). 'Lifecycle bias in estimates of intergenerational earnings persistence', *Labour Economics*, vol. 13(5), pp. 551–70.
- Guell, M., Rodriguez Mora, J.V. and Telmer, C.I. (2015). 'The informational content of surnames, the evolution of intergenerational mobility, and assortative mating', *Review of Economic Studies*, vol. 82(2), pp. 693–735.
- Guest, A.M., Landale, N.S. and McCann, J.C. (1989). 'Intergenerational occupational mobility in the late 19th century United States', *Social Forces*, vol. 68(2), pp. 351–78.
- Haidar, S. and Solon, G. (2006). 'Life-cycle variance in the association between current and lifetime earnings', *American Economic Review*, vol. 96(4), pp. 1308–20.
- Hertz, T. (2009). 'Rags, riches, and race: the intergenerational economic mobility of black and white families in the United States', in (S. Bowles, H. Gintis and M.O. Groves, eds.), *Unequal Chances: Family Background and Economic Success*, pp. 165–91, Princeton, NJ: Princeton University Press.
- Hertz, T., Jayasundera, T., Piraino, P., Selcuk, S., Smith, N. and Verashchagina, A. (2008). 'The inheritance of educational inequality: international comparisons and fifty-year trends', *B.E. Journal of Economic Analysis & Policy*, vol. 7(2), <https://doi.org/10.2202/1935-1682.1775>.
- Hornbeck, R. (2012). 'The enduring impact of the American dust bowl: short- and long-run adjustments to environmental catastrophe', *American Economic Review*, vol. 102(4), pp. 1477–507.
- Krueger, A.B. (2012). 'The rise and consequences of inequality', available at: [https://obamawhitehouse.archives.gov/sites/default/files/krueger\\_cap\\_speech\\_final\\_remarks.pdf](https://obamawhitehouse.archives.gov/sites/default/files/krueger_cap_speech_final_remarks.pdf) (last accessed: 24 October 2016).
- Lee, C.I. and Solon, G. (2009). 'Trends in intergenerational income mobility', *Review of Economics and Statistics*, vol. 91(4), pp. 766–72.
- Levine, D.I. and Mazumder, B. (2002). 'Choosing the right parents: changes in the intergenerational transmission of inequality between 1980 and the early 1990s', available at: <http://www.ssrn.com/abstract=323881> (last accessed: 24 October 2016).
- Long, J. and Ferrie, J.P. (2007). 'The path to convergence: intergenerational occupational mobility in Britain and the US in Three Eras', *Economic Journal*, vol. 117(519), pp. C61–C71.
- Long, J. and Ferrie, J.P. (2013). 'Intergenerational occupational mobility in Great Britain and the United States Since 1850', *American Economic Review*, vol. 103(4), pp. 1109–37.

- Mayer, S.E. and Lopoo, L.M. (2005). 'Has the intergenerational transmission of economic status changed?', *Journal of Human Resources*, vol. 40(1), pp. 169–85.
- Mazumder, B. (2009). 'The apple falls even closer to the tree than we thought', in (S. Bowles, H. Gintis and M.O. Groves, eds.), *Unequal Chances: Family Background and Economic Success*, pp. 80–99. Princeton, NJ: Princeton University Press.
- Mazumder, B. (2015). 'Estimating the intergenerational elasticity and rank association in the US: overcoming the current limitations of tax data', available at: [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2647277](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2647277) (last accessed: 24 October 2016).
- McGraw, G., Rehling, J. and Goldstone, R. (1994). 'Letter perception: toward a conceptual approach', in *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, pp. 613–8 Atlanta, GA: The Cognitive Science Society.
- Nybom, M. and Stuhler, J. (2014a). 'Biases in standard measures of intergenerational income dependence', available at: <https://janstuhler.files.wordpress.com/2013/06/lifecyclebiaspt2november2014.pdf> (last accessed: 24 October 2016).
- Nybom, M. and Stuhler, J. (2014b). 'Interpreting trends in intergenerational mobility', available at: [http://www.sofi.su.se/polopoly\\_fs/1.170138.1394463038!/menu/standard/file/WP14no3.pdf](http://www.sofi.su.se/polopoly_fs/1.170138.1394463038!/menu/standard/file/WP14no3.pdf); [http://www.ho.mepages.ucl.ac.uk/uctpjst/stuhlerinterpretingtrends\(JMP\).pdf](http://www.ho.mepages.ucl.ac.uk/uctpjst/stuhlerinterpretingtrends(JMP).pdf) (last accessed: 24 October 2016).
- Olivetti, C. and Paserman, M.D. (2015). 'In the name of the son (and the daughter): intergenerational mobility in the United States, 1850–1940', *American Economic Review*, vol. 105(8), pp. 2695–724.
- Parman, J.M. (2011). 'American mobility and the expansion of public education', *Journal of Economic History*, vol. 71(1), pp. 105–32.
- Perez, S. (2016a). 'Moving to opportunity: railroads, migrations and economic mobility in 19th century Argentina', available at: <http://web.stanford.edu/santip/> (last accessed: 24 October 2016).
- Perez, S. (2016b). 'The (South) American dream: mobility and economic outcomes of first and second generation immigrants in 19th century Argentina', available at: <http://web.stanford.edu/santip/PerezSouthAmerican.pdf> (last accessed: 24 October 2016).
- Rockoff, H. (2008). 'US economy in World War I', available at: <http://eh.net/encyclopedia/u-s-economy-in-world-war-i/> (last accessed: 24 October 2016).
- Simpson, I.C., Mousikou, P. Montoya, J. M. and Defior, S. (2013). 'A letter visual-similarity matrix for Latin-based alphabets', *Behavior Research Methods*, vol. 45(2), pp. 431–9.
- Solon, G. (1999). 'Intergenerational mobility in the labor market', *Handbook of Labor Economics*, vol. 3, chapter 29, Part A, pp. 1761–800.
- Tassava, C. (2008). 'The American economy during World War II', available at: <http://eh.net/encyclopedia/the-american-economy-during-world-war-ii/> (last accessed: 24 October 2016).
- Thernstrom, S. (1964). *Poverty and Progress: Social Mobility in a Nineteenth Century City*, Cambridge, MA: Harvard University Press.
- Thernstrom, S. (1973). *The Other Bostonians: Poverty and Progress in the American Metropolis, 1880–1970*, Cambridge, MA: Harvard University Press.
- Xie, Y. and Killewald, A. (2013). 'Intergenerational occupational mobility in Great Britain and the United States since 1850: comment', *American Economic Review*, vol. 103(5), pp. 2003–20.