2015-08-01

In the name of the son (and the daughter): intergenerational mobility in the United States, 1850-1940

Claudia Olivetti, M Daniele Paserman. 2015. "In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940." AMERICAN ECONOMIC REVIEW, v. 105, Issue 8, pp. 2695 - 2724 (30).

https://hdl.handle.net/2144/26121

In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940.*

Claudia Olivetti and M. Daniele Paserman

Boston University and NBER

April 2015

Abstract

This paper estimates historical intergenerational elasticities between fathers and children of both sexes in the United States using a novel empirical strategy. The key insight of our approach is that the information about socio-economic status conveyed by first names can be used to create pseudo-links across generations. We find that both father-son and father-daughter elasticities were flat during the 19$^{\text{th}}$ Century, increased sharply between 1900 and 1920, and declined slightly thereafter. We discuss the role of regional disparities in economic development, trends in inequality and returns to human capital, and the marriage market in explaining these patterns.

Keywords: Intergenerational Mobility, Marriage.

JEL codes: J62, J11, N31

*We thank Dara Lee, Laura Salisbury and Matthew S. Johnson for excellent research assistance; Robert Margo, Andrew Newman and Marc Rysman for many valuable comments and conversations; Evan Roberts for providing us the crosswalk between the raw 1940 occupational code and the IPUMS occupational score and for his help in understanding the 1940 Restricted Complete Count micro data; and participants at the NBER Cohort Studies, Development of the American Economy and Income Distribution and Macroeconomics workshops, the BU/BC Green Line Labor Meetings, the Society for Labor Economics meetings, the Workshop on Intergenerational Mobility at the University of Copenhagen, and seminar participants at Tufts University, University of Michigan, Northeastern University, NORC at the University of Chicago, the St. Louis Fed, Vanderbilt University, New York University, IIES, ECARES, LSE, Toulouse and UC Davis for many helpful suggestions. The authors declare that they have no relevant material or financial interests that relate to the research described in this paper.

The degree to which economic status is passed along generations is key to understanding differences across societies and over time in the extent of inequality. A low degree of intergenerational mobility can undermine the notion of equality of opportunity and may lead to persistent inequality. Recent research reveals that today intergenerational mobility in the U.S. is lower than in most other developed countries (Corak, 2013). This finding stands in contrast with the national *ethos* of the United States as the land of unlimited opportunity. Was this view ever justified?

In this paper we provide a new perspective on the evolution of intergenerational mobility in the United States in the late 19th and early 20th Centuries. We extend the existing literature by looking at the intergenerational elasticity in economic status between fathers and children of both sexes. Focusing only on father-son correlations may miss part of the picture. Daughters should be included if we want to know how the *average* well-being of a generation correlates with that of their parents. If there is a strong stratification in marriage by social class, assortative mating might magnify individual-level intergenerational persistence. Moreover, to the extent that mothers play a key role in the human capital accumulation of their children, investment in daughters could have important consequences for the transmission of status across multiple generations. Thus, to reach a fuller understanding of the transmission of resources across generations, it is important to focus on daughters as well as sons.

Typically, the estimation of intergenerational elasticities is based on a regression of an individual's economic status at time t on that of his or her own father at time $t-k$. This requires the use of longitudinal data sets that link fathers to their offspring. Historical longitudinal data sets based on Census data make it possible to link fathers and sons by first and last names. However, one cannot link fathers and daughters in this manner because women change last name upon marriage. The contribution of this paper is to develop an empirical strategy that enables us to estimate the intergenerational elasticity between fathers and daughters, as well as between fathers and sons, even when it is not possible to link individuals directly across generations.[1,2] The key insight of our approach is that the information about socio-economic status conveyed by first names can be used to create a *pseudo-link* between fathers and sons, as well as between fathers and daughters.

To illustrate this idea, consider a simple example. Assume that the only possible names

[1]The data does not allow us to calculate the intergenerational elasticity in income, as this information is not available before 1940. Instead in most of our specifications we proxy income using an index of occupational status based on the 1950 income distribution. Somewhat loosely, we refer to our estimates as the intergenerational income elasticity, or simply intergenerational elasticity.

[2]Since married women during this period had low labor force attachment, we measure daughters' economic status by that of their husbands.

in the population are Aaron and Zachary. Moreover, assume that high socioeconomic status parents are more likely to name their child Aaron, while Zachary is more common among low socioeconomic status parents. If adult Aarons are still more likely to be high socioeconomic status than adult Zacharys, then we would infer that the degree of social mobility in this society is relatively low. Importantly, we can easily apply the same idea to girls, and ask whether the young Abigails (born to high socioeconomic status parents) are more likely to marry husbands that are themselves high socioeconomic status than the young Zoës (born to low socioeconomic status parents). It is important to note that this whole exercise will work only if names do in fact carry information about their parents' socioeconomic status. We present evidence that this is indeed the case: between 11 and 17 percent of the total variation in father's socioeconomic status can be explained by the variation *between* names given to their children.

Our empirical strategy amounts to imputing father's income, which is unobserved, using the average income of fathers of children with a given first name. This is essentially a "two-sample two-stage least squares" estimator (TS2SLS, Inoue and Solon, 2010). In the first step, we use the sample of fathers and regress father's log earnings on a full set of children's first name dummies. In the second step, we use the sample of sons, and regress son's log earnings on the cross-sample predicted values from the first step. We sometimes refer to this estimator as a *pseudo-panel* estimator, as it is based on the creation of pseudo-links across generations. It is important to emphasize that our goal is not to uncover the "true" intergenerational elasticity, but rather to provide a specific estimator that can be calculated consistently over time and for both genders, and can identify trends in intergenerational mobility.

We estimate father/son and father/son-in-law intergenerational income elasticities using 1% extracts from the Decennial Censuses of the United States between 1850 and 1940. Our baseline results indicate that the intergenerational elasticity between fathers and sons increased by 24% between 1870 and 1940. This increase is consistent with the findings of Ferrie (2005) and Long and Ferrie (2007, 2013), who document a marked decrease in intergenerational mobility in the United States between the late 19$^{\text{th}}$ Century and the middle of the 20$^{\text{th}}$ Century. The elasticity however does not increase smoothly over time: it is relatively flat throughout the second half of the 19$^{\text{th}}$ Century, then increases sharply between 1900 and 1920, followed by a slight decline between 1920 and 1940.

The main finding of a sharp increase in elasticities between 1900 and 1920 is robust to different methods of treating farmers, imputing income, coding names, and to differential mortality across socioeconomic groups and selection into marriage. Only the apparent dip post-1920 is somewhat sensitive to whether farmers are included in the sample and to the exact imputation of their income.

The intergenerational elasticity between fathers and sons-in-law displays a similar trend between 1870 and 1920 (mostly flat between 1870 and 1900, and a sharp increase between 1900 and 1920), suggesting that there was a substantial degree of assortative mating. There are however some slight differences in timing. The father/son-in-law elasticity is higher than the father/son elasticity in the late part of the 19$^{\text{th}}$ Century, but the two elasticities converge by 1920. After 1920, the trend in father/son-in-law elasticity is sensitive to the exact measure of income used, with some, but not all, estimates pointing to father-daughter mobility in 1940 returning to the levels prevalent in the late 1800s.

We investigate which historical developments may explain the trends and the gender differentials in intergenerational elasticity. We argue that the sharp increase in elasticities between 1900 and 1920 is consistent with patterns of regional disparities in economic development, and with the increase in inequality and returns to human capital. Other mechanisms such as changes in fertility, immigration and internal migration seem less likely to matter. Gender differentials in elasticities are consistent with imbalances in the sex ratio due to maternal and infant mortality, wars, and changes in migratory flows.

Our paper is related to an extensive literature that studies intergenerational mobility using modern panel data (see the comprehensive surveys by Solon, 1999, and Black and Devereux, 2011). The bulk of the literature focuses on father-son intergenerational mobility and finds an intergenerational labor income elasticity hovering around 0.4. Only a limited number of papers in this literature have studied the correlations between father-in-law and son-in-law. Chadwick and Solon (2002) use PSID data to study intergenerational mobility in the daughter's family income. They find that for modern US data the father/son elasticity - estimated to be equal to 0.523 - tends to be somewhat larger than the father/son-in-law elasticity- estimated at 0.360. Raaum et al. (2007) confirm this result for the US, the UK, and three Nordic countries. Associated to the increasing labor force participation of women, recent studies have focused on father-daughter occupational mobility. Jäntti et al. (2006) document that in five of six developed countries, the father/son intergenerational elasticity is higher than the father/daughter one. Hellerstein and Morrill (2011) find that the probability that a woman works in the same occupation as her father has increased over the course of the 20$^{\text{th}}$ Century.

Closely related to our project is the work by Güell, Rodríguez-Mora and Telmer (2007), who use the informative content of surnames to study intergenerational mobility in Spain. They develop a model whose endogenous variable is the joint distribution of surnames and income, and explore the relationship between mobility and the informative content of surnames, allowing for assortative mating to be a determinant of both. They find that the degree of mobility in Spain has substantially decreased over time. Others have instead exploited the distribution of surnames in data sets that are centuries apart to estimate long-run social mobility. Collado, Ortuño Ortín and Romeu (2012), using data from two Spanish regions, find that socioeconomic status at the end of the 20$^{\text{th}}$ Century still depends heavily on the socioeconomic status of one's great-great grandparents. Clark and Cummins (2012) use the distribution of surnames in England, and conclude that there is considerable persistence of status in the UK between 1800 and 2012, higher than that estimated in most modern studies. Clark (2014) shows that this high degree of persistence in economic status is in fact common to many other societies, ranging from Communist China and egalitarian Sweden to caste-based India.

The rest of the paper proceeds as follows. Section I describes the econometric methodology. Section II presents the data and discusses measurement issues. The main results are presented in Section III. Section IV provides robustness checks and Section V explores alternative factors underlying the trends. Section VI concludes.

I Methodology

Consider an individual i who is young at time $t-1$ and adult at time t. Let y_i^S be individual i's *log earnings* at time t, and y_i^F be his father's *log earnings* at time $t-1$. With individually linked data, both y_i^S and y_i^F are observed, and the intergenerational elasticity estimate is obtained by regressing y_i^S on y_i^F. We will call this estimator the *linked* estimator, $\hat{\eta}_{LINKED}$.[3]

Assume instead that we only observe two separate cross-sections and it is impossible to link individuals across the two. This means that y_i^F is unobserved, and it becomes necessary to impute it. Our strategy is to base the imputation on an individual's first name, which is available for both adults and children in each cross-section. It is important to emphasize that the main goal of our strategy is to derive a measure of intergenerational elasticity that is consistent over time and across genders, rather than uncovering the "true" intergenerational

[3]Becker and Tomes (1979, 1986) show that this regression can be derived from a model in which altruistic parents choose how to allocate their lifetime earnings between their own consumption and investment in the earning capacity of their children. Children's earnings are a function of parental investment and of a stochastic "endowment" that is transmitted across generations and follows an AR(1) process.

elasticity. Therefore, even if our estimator in general will not be equal to the one obtained with individually linked data, for the purposes of our analysis we only need that the sources of bias be consistent over time and across genders.

Mechanically, for an adult at time t named j, we replace y_i^F with $(\bar{y}_j^F)'$, the average log earnings of fathers of children named j, obtained from the time $t-1$ cross section (the "prime" indicates that this average is calculated using a different sample). We have thus created a "generated regressor" by using one sample to create a proxy for an unobserved regressor in a second sample. As highlighted by Inoue and Solon (2010), this estimator is essentially a "two-sample two-stage least squares" (TS2SLS) estimator. In the first step, we use the sample of fathers and regress father's log earnings on a full set of children's first name dummies. In the second step, we use the sample of sons, and regress son's log earnings on the cross-sample fitted values from the first stage.[4] We rely on these results to calculate appropriate standard errors for our estimator. Since we do not aim to identify the causal effect of parental income on children's income,[5] we do not require the first name dummies (the instruments in the first stage of the TS2SLS estimator) to satisfy any exclusion restriction. Because we are effectively creating a pseudo-panel of individuals linked by first names, we refer to this estimator as the "pseudo-panel" estimator, and label it $\hat{\eta}_{PSEUDO}$.[6]

We now derive the probability limit of our estimator and compare it with that of the traditional linked estimator, in order to facilitate comparisons to estimates of intergenerational elasticities based on linked data.

Let y_{ij}^S be the log earnings of a son i named j and y_{ij}^F be the log earnings of a father of a son named j. We can write:

$$y_{ij}^S = \beta\, y_{ij}^F + \lambda_j + u_{ij};$$
$$y_{ij}^F = \mu_j + z_{ij}.$$

[4] In fact there are really two levels of instrumenting because we are using occupational income instead of actual income. We come back to this point in section II.

[5] After all, the traditional linked estimator also does not identify a causal parameter because parental income is correlated with the error term in the son's earnings equation.

[6] The second stage has a particularly simple structure because the right hand side variable is constant for every individual with the same first name. Therefore, in the special case of no additional regressors, the TS2SLS estimator is equivalent to a weighted least squares regression of $\bar{y}_j^S$ on $(\bar{y}_j^F)'$, where $\bar{y}_j^S$ is the average log earnings of adults named j at time t, and the weights are equal to the frequency counts of first names in the son's sample. This equivalence highlights the similarity between our approach and the synthetic cohort method pioneered by Browning, Deaton and Irish (1985). In our case, the synthetic cohorts are defined on the basis of both first names and age. Aaronson and Mazumder (2008) use an estimation strategy that is also based on synthetic cohorts. They estimate intergenerational mobility in the US between 1940 and 2000 by imputing father's income using state and year of birth.

In the above, λ_j is a name fixed effect that captures any return in the labor market associated with a given first name above and beyond any direct effect of father's income. This may reflect factors such as ethnicity, religion, state of birth, or any other signal of social status associated with a given first name. On the other hand, μ_j is the conditional expectation of y_{ij}^F given that the father named his son j. By construction, z_{ij} is uncorrelated with μ_j and λ_j. Furthermore, let us decompose u_{ij} into a part that is potentially correlated with μ_j and one that is not, i.e., $u_{ij} = \kappa_j + \tilde{u}_{ij}$, where $\kappa_j \equiv E\left(u_{ij}|\mu_j\right)$ and $\tilde{u}_{ij} \equiv u_{ij} - \kappa_j$.

A positive correlation between κ_j and μ_j could arise if parents engage in "aspirational naming," i.e. if ambitious and motivated parents who assign children high socio-economic status names (high μ_j) also transfer to them their work-ethic and push them to succeed in the labor market (high u_{ij}). A positive correlation between $\tilde{u}_{ij}$ and z_{ij}, instead, represents the correlation between unobserved characteristics of father and son that are not captured in the son's name, i.e., cognitive and physical ability, connections, etc.

The probability limit of the linked estimator is equal to:

$$\begin{aligned} p\lim \hat{\eta}_{LINKED} &= \frac{Cov\left(y_{ij}^S, y_{ij}^F\right)}{V\left(y_{ij}^F\right)} \\ &= \beta + \frac{Cov\left(\lambda_j + \kappa_j, \mu_j\right)}{V\left(\mu_j\right) + V\left(z_{ij}\right)} + \frac{Cov\left(\tilde{u}_{ij}, z_{ij}\right)}{V\left(\mu_j\right) + V\left(z_{ij}\right)}. \end{aligned} \tag{1}$$

As is well known, $\hat{\eta}_{LINKED}$ is not a consistent estimator of the causal effect of an increase in father's income on son's income because of the potential correlation between the unobservables. In the equation above we have decomposed the correlation between the error term and father's earnings into a part that comes from the group-specific component μ_j, and a part that comes from the idiosyncratic component z_{ij}.

The probability limit of the pseudo-panel estimator is:

$$\begin{aligned} p\lim \hat{\eta}_{PSEUDO} &= \frac{Cov\left(y_{ij}^S, \bar{y}_j'^F\right)}{V\left(\bar{y}_j'^F\right)} \\ &= \frac{V\left(\mu_j\right)}{V\left(\mu_j\right) + E(\frac{1}{N_j})V\left(z_{ij}'\right)}\beta + \frac{Cov\left(\lambda_j + \kappa_j, \mu_j\right)}{V\left(\mu_j\right) + E(\frac{1}{N_j})V\left(z_{ij}'\right)}, \end{aligned} \tag{2}$$

where N_j is the number of observations in cell j; and $\bar{y}_j'^F = \mu_j + \bar{z}_j'$, with the "prime" indicating variables drawn from a different sample. We have used the fact that $Cov\left(\mu_j, \bar{z}_j'\right) = 0$ by construction; and the covariance between the idiosyncratic terms drawn from different samples, $Cov\left(\tilde{u}_{ij}, \bar{z}_j'\right)$, is equal to zero.

The first thing to note about this probability limit is that it depends crucially on the between-name variance in father's income, $V(\mu_j)$, being greater than zero (and henceforth, $Cov(\lambda_j + \kappa_j, \mu_j) \neq 0$). This is equivalent to requiring that names are not distributed randomly in the population. If this were the case, the generated regressor would be just noise. In large samples, both the numerator and the denominator in both terms of equation (2) would be equal to zero, making the pseudo-panel estimator asymptotically indeterminate. In finite samples, however, the number of observations per cell is finite, so that the denominator would not vanish even if $V(\mu_j)$ is equal to zero. In this case, the pseudo-panel estimator would converge to zero.

Therefore, a key requirement of our methodology is that first names carry information about socioeconomic status. The higher the informational content of first names, the more accurate is $\bar{y}_j'^F$ as a predictor of y_{ij}^F. There is abundant empirical evidence supporting the assumption that parents choose first names partly to signal their own standing in society, or their cultural and religious beliefs. Bertrand and Mullainathan (2004) document that in a sample of baby names in Massachusetts there is substantial between-name heterogeneity in the social background of mothers; similarly, Fryer and Levitt (2004) show that names provide a strong signal of socioeconomic status for blacks, but also that there are systematic and large differences in name choices by whites with different levels of education. This practice is also widespread in other societies, both today and in the past. Head and Mayer (2008) investigate the social transmission of parental preferences through naming patterns in France. Hacker (1999) and Haan (2005) document a relationship between first names, religiosity and fertility in Canada and the US during the 19th Century. Cook, Logan and Parman (2014) find that distinctively black names were already common in the post-Civil War period.

We can now compare the probability limits of the linked and pseudo-panel estimator. The first term in the pseudo-panel estimator is unambiguously smaller than the corresponding term in the linked estimator. This is the traditional attenuation bias deriving from the fact that we replace true father's income with an imputed value, thus introducing measurement error. The attenuation bias is larger the larger is the variance of z_{ij} relative to the variance of μ_j, indicating that first names carry little information about socioeconomic status; and the smaller is N_j, the number of observations per name, reflecting the fact that averages computed on a smaller sample will measure father's income less precisely.[7]

[7] This point illustrates that using finer cells to impute father's income (such as last names, or first names by state of birth), while possibly achieving higher precision in the imputed values for father's income, can also exacerbate measurement error.

The second term in the pseudo-panel estimator is larger in absolute value than the corresponding term in the linked estimator. If the covariance between $\lambda_j + \kappa_j$ and μ_j is positive, this term will pull up the pseudo-panel estimator relatively to the linked estimator, counteracting the attenuation bias associated with the first term. As this term is distinctively tied to our methodology of imputing income based on first names, we discuss it in detail below.

Finally, the third term in equation (1) vanishes from equation (2). Whether this introduces upward of downward bias depends on the sign of the correlation in motivation, genetic ability, social capital and other unobservables that are not embodied in first names. If these unobservables are positively correlated across generations, as is reasonable to assume, then the pseudo-panel estimator will be pulled down relative to the linked estimator.

As discussed above, the second term in equations (1) and (2) represents both the direct labor market premium (or penalty) potentially associated with a given first name (the covariance between λ_j and μ_j), and the effect due to aspirational naming (the covariance between κ_j and μ_j).[8] Although there are reasons to believe that these covariances are positive, there may also be forces that push in the other direction. For example, the literature on the economic consequences of first names is mixed. Bertrand and Mullainathan (2004) show that distinctively black names decrease the likelihood that a job applicant is called for an interview, while Fryer and Levitt (2004) find no negative causal impact of having distinctively black names on life outcomes. As for "aspirational" naming, parents may believe that by choosing names that are associated with a higher social class they may facilitate their children's social mobility and prevent discrimination. On the other hand, there may be advantages in choosing a name that signals membership in an ethnic, religious or socio-economic group. For example, names that deviate from the group norm may carry a social stigma and lead to a penalty in the labor market or marriage market.[9]

Overall, the pseudo-panel estimator can be either lower or higher than the linked estimator, depending on which of the three effects dominates. In practice, we show in Section III that for samples in which we can calculate both the linked and pseudo-panel estimators, the latter is lower by about 30%.

The discussion above was presented in terms of the intergenerational elasticity between fathers and sons. One of the distinct advantages of this methodology is that it can be easily

[8] In practice, it is not possible empirically to distinguish between the two separate elements of this covariance.

[9] Aspirational naming is likely to be especially widespread among immigrants (Biavaschi, Giulietti and Siddique, 2013). Our results however are not sensitive to controls for immigrant status or to excluding immigrants from the sample altogether. See Section B.

applied to calculate the correlation in economic status between fathers-in-law and sons-in-law, where the daughters' names are used to create the intergenerational link. Our estimator boils down to a regression of son-in-law's income on father-in-law's income, where father-in-law's income for men married to women named j is proxied by the average income of fathers of daughters named j at time $t-1$.

II Data

We now apply our methodology to data from the 1850 to 1940 Decennial Censuses of the United States, which contain information on first names. For 1850 to 1930 we use the 1% IPUMS samples (Ruggles et al., 2010). For 1940 we create a 1% extract of the IPUMS Restricted Complete Count Data (Minnesota Population Center and Ancestry.com, 2013). We restrict all the analysis to whites to avoid issues associated with the almost complete absence of blacks in the pre-Civil War period, and the fact that even in the late cohorts many blacks would have spent a substantial part of their lives as slaves.

Measuring Earnings. The first challenge that generally applies to the computation of historical intergenerational elasticities, is to obtain appropriate quantitative measures of socioeconomic status. Because income and earnings at the individual level are not available before the 1940 Census, we are constrained to use measures of socioeconomic status that are based on individuals' occupations. There is a long tradition in sociology to focus on measures of occupational prestige, and these are believed to be better indicators of long-run income (Duncan, 1966; see also the survey by Erikson and Goldthorpe, 1992). On the other hand, these measures fail to capture the potentially large within-occupation variance in income. In practice, estimates of intergenerational elasticities based on multi-year averages of father's income (as recommended in Solon, 1992) are quite close to estimates based on predicted income by occupation (Björklund and Jäntti, 1997).

One of the advantages of the IPUMS data set is that it contains a harmonized classification of occupations, and several measures of occupational status that are comparable across years. For our benchmark analysis, we choose the OCCSCORE measure of occupational standing.[10] This variable indicates the median total income (in hundreds of dollars) of the persons in each occupation in 1950. We address the sensitivity of our results to alternative measures of occupational standing in Section IV.

[10] A number of other papers have used this same variable to measure occupational standing, among them Abramitsky, Platt-Boustan and Eriksson (2012), Cvrcek (2012), Jones and Tertilt (2008) and Katz and Margo (2014).

Coding of names. The second challenge, specific to our methodology, is how to correctly match first names across censuses. In our benchmark classification of names we ignore middle initials (that is, we treat “William” as equivalent to “William J.”) and we treat nicknames as distinct names (that is, “William” and “Bill” are considered two different names).[11] These choices may not be harmless, since there may be systematic differences in socioeconomic status between individuals with middle initials or nicknames and those without. We assess the sensitivity of our estimates to these choices in Online Appendix A. The results are robust to using different name coding schemes.

The Distribution of Names. We first document some features of the distribution of first names in the sample. Table 1 reports the summary statistics for children's names in the initial year of the pseudo-panel by gender. Both population (column 1) and the number of distinct names (column 2) grow between 1850 and 1920, but the average number of observations per name (column 3) is roughly constant. This pattern is common across genders. In every decade, a large proportion of names appears only once in the sample (column 4). However, as shown in column 5, singleton names only account for 6 to 7% of all names. Furthermore, we can link at least 90% of children's names across Census decades (column 6).

The last two columns of the table present features of the name distribution. Column 7 reports the share of the total population with one of the 50 most popular names. This describes how concentrated the name distribution is. Both male and female names become markedly less concentrated over the sample period, with the decline for girls occurring earlier and being more pronounced. Column 8 reports the R^2 coefficient obtained by regressing log father's occupational income on a set of name indicators. Note that if names were assigned at random, and we had a sufficiently large number of occurrences for every name, the between-name variation would not explain any of the total variation in father's income, and the R^2 coefficient would be equal to zero. The entries in the column show that the between name variation varies by gender: it accounts for 11% to 14% of the total variation in fathers' log earnings for boys and 13% to 17% for girls. Because of the large number of singleton names, we could observe a positive R^2 even if names were assigned completely at random. Based on Monte Carlo simulations, we calculate the probability that the R^2 obtained under this assumption, holding constant the actual frequency distribution of names, is as high as the observed R^2 in the data. In all years and for both genders we can soundly reject the

[11] The only exception to this rule is that we transform obvious abbreviations into their correspondent full name (e.g., “Wm.” becomes “William,” “Geo.” becomes “George,” etc.).

hypothesis that names carry no information about the father's socioeconomic status (p-value < 0.001).

Table 2 reports the 5 most prestigious and least prestigious names based on father's occupational income, separately for each Census year. The shaded entries in the table refer to names that appear more than once within the category of most prestigious names (light gray) and least prestigious names (dark gray). The patterns of shaded areas reveals that there is indeed persistence both in the top 5 and in the bottom 5 names across Census decades for both male children and female children. If names were assigned at random, it would be quite unlikely for a given name to appear more than once in this table.

III Results

Figure 1 and rows 1 and 4 in Table 3 report the results of our benchmark analysis. We report 20-year elasticities in occupational income for both the father-son and the father in law-son in law comparisons.[12]

Between 1870 and 1940, the intergenerational elasticity between fathers and sons increases by 24%, and that between fathers and sons-in-law by 9%. The father/son elasticity is relatively flat throughout the second half of the 19$^{\text{th}}$ Century, increases sharply between 1900 and 1920, and declines slightly between 1920 and 1940. The father/son-in-law elasticity exhibits a first increase between 1870 and 1880 and then a further jump between 1900 and 1920, which coincides with the increase in the father/son elasticity. The two elasticities are almost identical in 1920 but they diverge at the end of the period with the father/son-in-law elasticity declining more sharply and dipping below the father/son elasticity. Overall, the father/son and father/son-in-law elasticities exhibit similar trends, suggesting that there was a high degree of positive marital sorting during the sample period. The ranking of son-in-law and son elasticities is consistent with modern estimates for the US and other developed economies (Chadwick and Solon, 2002, Raaum et al., 2007).[13] We defer to section V for a discussion of the historical developments that can rationalize these findings.

[12]The intergenerational *correlation* may differ from the elasticity if the dispersion of earnings varies substantially across generations. We find that this is not the case: the magnitude and trends of intergenerational correlations are almost identical to the elasticities reported in Table 3.

[13]For 1940 we can also estimate the intergenerational elasticities using actual wage and salary income as the dependent variable, as opposed to the occupational income score. Our estimates are higher than those shown in Table 3. This is consistent with the analysis in Björklund and Jäntti (1997), who show that a regression of actual son's income on predicted father's income (by occupation and education) yields higher estimates than those obtained from actual-actual or predicted-predicted regressions.

The remaining rows in Table 3 show how our benchmark estimates are affected by sample selection issues due to either differences in child mortality by socioeconomic status, or to differences in the age distribution and marital status of sons and sons-in-law. In the second row of each panel we present estimates where we restrict the sample to children who were aged 5-15 in the earlier census. The incidence of child mortality was still very high during much of the sample period (Preston and Haines, 1991), so that it is likely that a non-negligible fraction of children did not survive into adulthood. If child mortality differs by socioeconomic status, or if healthier children are also more likely to be employed as adults in high-income occupations, this would lead to a standard sample selection problem and potentially biased coefficients. Since most child mortality occurred before age 5, restricting the sample to include only older children should alleviate this problem. The estimated coefficients for sons are somewhat lower than the benchmark, but the trends in elasticities are mostly unaffected. The father/son-in-law elasticities are not sensitive to the exclusion of younger daughters.

In all societies men marry later in life than women and the gender differential in age at first marriage tends to be largest in more traditional societies. The 19th Century US is no exception. As documented in Ferrie and Rolf, (2008) and Fitch and Ruggles (2000), the male-female differential in median age at first marriage was quite large in the 19th Century, peaked in 1900 at more than 4 years, and then declined to about 2 years at the beginning of the 20th Century. In our samples this implies that sons-in-law are, on average, older than sons (especially at the beginning of the period) and that a fraction of the sons are unmarried. Failing to control for differences in the age distribution has the potential to affect the comparison of father/son-in-law and father/son elasticities. In particular, if the wage-age profile is concave, and sons are systematically younger than their brothers-in-law, we would systematically overestimate the father/son-in-law elasticity relative to the father/son elasticity. In the third and sixth rows of Table 3 we attempt to make the son and son-in-law samples more comparable in terms of their demographic characteristics. In the third row, we restrict the sons sample to married individuals. In the sixth row, we only include individuals aged 20 to 35 in the sample of sons-in-law. There is some variation in the point estimates, but on the whole the results are very similar to the benchmark.

As a further robustness check, we also estimated our model with controls for a quadratic function in father's and son/son-in-law's age, with no discernible effects on our estimates.[14] Intergenerational elasticities at 30-year intervals exhibit a similar pattern, confirming the fact that in the period under examination estimates are not very sensitive to the age at

[14] See Olivetti and Paserman (2013), Table 6, for detailed results.

which income is measured (see Online Appendix B). The insensitivity of our estimates to the timing of income measurement can also be explained by the shape of the age-income profile during this period, which displayed an earlier peak and less concavity relative to modern times (Sutch, 2011). The fact that we use occupational income further alleviates concerns about life-cycle effects.

The bottom panel of Table 3 presents estimates of the father/son elasticities for the two 20-year comparisons for which individually linked data are available.[15] The first coefficient is the least squares estimate and yields intergenerational elasticities that are 13 to 15 log points higher than those obtained with the pseudo-panel estimator. Remarkably, both the linked estimates and the pseudo-panel estimates are fairly stable across the two data points. Similarly, for the two 30-year comparisons that are available, we find that the pseudo-panel estimates are lower than the corresponding linked data estimates, and both series exhibit an upward trend (see Online Appendix B).

To understand where the difference between the two estimators comes from, we report in the bottom row of the table the coefficients from a regression that also includes name fixed effects. Even though the name fixed effects are jointly statistically significant (F = 1.23, p-value < 0.001), the estimated intergenerational elasticity drops by a only a tiny amount.[16] In terms of equations (1) and (2), this shows that $Cov(\lambda_j + \kappa_j, \mu_j)$ is quite small, implying that the aspirational naming/direct labor market effect of first names is relatively unimportant. In the pseudo-panel estimator the term that depends on this covariance is inflated by $\frac{V(\mu_j)+V(z_{ij})}{V(\mu_j)+E(\frac{1}{N_j})V(z_{ij})}$; at the same time, the term in β is attenuated by a factor of $\frac{V(\mu_j)}{V(\mu_j)+E(\frac{1}{N_j})V(z_{ij})}$. It turns out that, for values of β between 0.3 and 0.4,[17] and estimates of the variance components from the linked sample, these two forces are on the order of magnitude of 1 to 4 log points and tend to exactly offset each other.[18] Therefore, it appears that the main source of attenuation in the pseudo-panel estimator comes from the fact that, because father's income is computed from a different sample, the third term in equation (1) vanishes.

[15]Source: IPUMS Linked Representative Sample (Ruggles et al., 2010). Since the linking is done using information on first and last names, no linked data on married women is available. Therefore, we can only compute father-son elasticities.

[16]This is true also for the 30-year elasticities presented in Online Appendix B.

[17]Given positive covariance between the unobservables in fathers' and sons' income, the true value of β must be lower than the OLS estimate.

[18]For example, in the 1880-1900 linked sample, we estimate $Cov\left(\lambda_j + \kappa_j, \mu_j\right) = 0.0014$; $V\left(\mu_j\right) + V\left(z_{ij}\right) = 0.188$; $V\left(\mu_j\right) + E(\frac{1}{N_j})V\left(z_{ij}\right) = 0.040$; and $\frac{V(\mu_j)}{V(\mu_j)+E(\frac{1}{N_j})V(z_{ij})} = 0.895$. Assuming $\beta = 0.35$, these estimates imply that the difference in the first term of equations (1) and (2) is about -0.04, and the difference in the second term is about 0.03.

Assuming that the degree of attenuation in the pseudo-panel estimator is constant over time, our estimates for 1940 imply an intergenerational elasticities of about 0.60. This value is close to the estimate of Aaronson and Mazumder (2008) for 1940, the earliest estimate available using modern Census data; and they are in the lower range of the estimates of Clark et al. (2012) for the United States.

For the two years in which linked data is available, it does indeed appear that the degree of attenuation caused by our methodology is fairly constant over time. However, there may still be some uncertainty as to whether this is also true in later years, and potentially account for the trends in elasticities. We believe that this is unlikely to be the case, for a number of reasons. First, we have shown that the aspirational naming/direct labor market effect of first names is relatively small. Even though we cannot rule out that the importance of this channel grew over time, it strikes us as implausible that the increase could have been large enough to explain the whole 15 log point increase in the intergenerational elasticity. Second, we have conducted a numerical exercise to study how the pseudo-panel estimator responds to changes in the name distribution.[19] The estimated intergenerational elasticity is not sensitive to the degree of concentration of names. Moreover, in order to generate the observed increase in intergenerational elasticities from the beginning to the end of the period, the informational content of names should have increased by a large amount. However, as we document in Table 1, the informational content of names has remained remarkably stable over time, with, if anything, a slight uptick in the 1920 cohort. This stands in contrast with the slight *decline* in the estimated elasticities after 1920. We conclude that the observed trends in elasticities are caused by fundamental changes in the degree of transmission of economic status, and are not an artifact of our methodology.

Finally, one may wonder why we choose to use first names to link the two data sets, rather than other variables that carry information about socioeconomic status and are available in both the son's and the father's samples. Two such candidates are family names (even though they would preclude us from studying daughters) and place of birth. It turns out that these alternative methods produce estimates that are substantially more distant from the estimates based on the linked data. Using family names, the intergenerational elasticity is estimated to be 0.08 in 1860-1880 and 0.14 in 1880-1900. Using place of birth, the corresponding numbers are 1.24 and 0.90. These results are not surprising in light of equations (1) and (2). When we use family names, cell sizes are very small, and therefore the attenuation of the first term is substantial. Intuitively, when we only use few individuals to estimate

[19] Olivetti and Paserman (2013), Section 7.1

father's occupational income, the estimate will be more noisy. On the other hand, using place of birth yields cell sizes that are very large, meaning that the attenuation of the first term is almost zero, while the inflation of the second term is substantial. This is because of a combination of two factors: first, the denominator of the second terms is small because cell sizes are very large; second, place of birth has a large direct effect on son's incomes: in regressions using the linked data where we directly control for place of birth, the estimated elasticity drops by 5 to 6 log points, and the implied covariance between $\lambda_j + \kappa_j$ and μ_j is an order of magnitude larger than the one obtained using first names. Intuitively, we are no longer estimating an intergenerational elasticity, but rather the degree of persistence in incomes across birth places.[20]

IV Robustness to the Measurement of Income

As is well known, the 1950 income distribution was relatively compressed (Goldin and Margo, 1992). Moreover, the 1950 occupational classification may not reflect accurately the relative standing of occupations that were common during the late 19$^{\text{th}}$ Century and early 20$^{\text{th}}$ Century. This issue is important from our standpoint as "farmers" represent a large part of our sample and farming occupations and farm ownership were associated with higher socioeconomic status during our sample period than in 1950.[21] As pointed out by Xie and Killewald (2011) measures of occupational mobility during this period of structural transformation can be sensitive to the treatment of farmers. We address these concerns by studying whether our estimates are sensitive to alternative imputations of occupational income, paying special attention to the imputation of farmers' income. The results are reported in Table 4. The first row of each panel reproduces the benchmark estimates from Table 3.

We start by imputing income using the 1900 occupational-earnings distribution obtained from the tabulations in Preston and Haines (1991). These tabulations are based on the 1901 Cost of Living Survey, which was designed to investigate the cost of living of families in industrial locales in the United States. The main advantage of using the 1900 occupational income distribution is that the list of occupational categories matches more closely the list,

[20]A third alternative would be to link individuals by both first name and state or region of birth. When we do this, we probably improve the accuracy of the imputation of father's income, but we also must rely on smaller cells, thus leading to greater attenuation of the estimates.

[21]The proportion of children whose father is a farmer is as high as 57 percent in 1850, and even though it declines steadily over the sample period, it is always above 30 percent.

types and ranking of occupations that were common during much of the sample period. This categorization, however, suffers from two limitations. First, the 1901 survey collected data for the "typical" urban family, meaning that by construction the resulting income distribution is more compressed than what one would obtain in a representative sample. Second, while Preston and Haines do impute income for some agricultural occupations, they explicitly refrain from imputing an average income for generic farm owners and farm tenants. We experiment with two different methods for imputing farmers' income, which are described in detail in Online Appendix C.

The intergenerational elasticities estimates based on the different imputation methods are reported in the second and third rows of each panel in Table 4. The father/son intergenerational elasticity is not very sensitive to using the 1900 occupational income distribution. The estimate of the father/son-in-law intergenerational elasticity is very similar to the benchmark under the first imputation method but the results differ somewhat under the second method. In this case, the estimated elasticity for 1870 to 1900 is 8 to 9 points lower relatively to the benchmark, but the difference becomes smaller in the following periods.

The next two rows of Table 4 show the estimated elasticities if we completely remove farmers from the analysis, using either the 1950 or the 1900 occupational income distribution. Both the son and son-in-law intergenerational elasticities are substantially lower than those in the benchmark analysis. This reflects the unsurprising fact that farming status is highly correlated across generations so that excluding farmers altogether raises intergenerational mobility.

The overall pattern in elasticity – flat between 1870 and 1900 and then sharply increasing between 1900 and 1920 – is robust to all imputation methods in Table 4. The evidence on the decline in elasticities post-1920 is slightly more mixed, with some of the measures indicating a more moderate decline, and some no decline whatsoever. The standard errors are such that it is difficult to state with a high degree of confidence whether there is an effective trend reversal after 1920. Nevertheless, it seems safe to conclude that over the whole sample period intergenerational mobility declined, no matter how we treat farmers' income. Therefore, the decline in intergenerational mobility does not seem to be driven by the structural transformation of the U.S. economy, from agricultural to industrial, over this period.

In Table 5 we assess the robustness of our results to additional alternative measures of occupational income. In the second row of each panel we replace occupational income with an individual's percentile rank in the distribution. The rationale for using rank is that it does not depend on the potentially noisy imputed level of occupational income. We find an attenuated trend for the father-son elasticity while the father/son-in-law elasticity mirrors the baseline fairly closely. We then exploit the information on wealth available in the 1850-1870 Censuses to derive an occupational hierarchy more appropriate for the beginning of the period (see Online Appendix D for details on the imputation procedure). The estimated elasticities track quite closely the benchmark estimates both in levels and in trends. In particular, the sharp increase between 1900 and 1920 is robust to the different imputation methods, as is the slight decline after 1920.

In the next row we re-estimate the model using average occupational incomes in 1990. The 1990 distribution has the advantage of being substantially more dispersed than the 1950 distribution, and therefore allows us to assess whether our measures of intergenerational mobility are affected by the variance of measured earnings. The estimated elasticities are lower than the benchmark estimates, but again the trends are broadly similar. Not surprisingly, the estimates are especially attenuated in the beginning of the sample period. Using an occupational distribution that is more distant in time from the actual period of analysis is likely to introduce more noise and attenuate the results. Finally, the last row in the table reports the estimates obtained using the Duncan socioeconomic index (SEI), a well-known measure of occupational prestige that combines occupational education and occupational income. As in the benchmark, there is a sizable increase in the father/son elasticity between 1900 and 1920 but it appears to plateau thereafter. On the other hand, the father/son-in-law elasticity does a exhibit a slight decline after 1920, albeit not as pronounced as in the benchmark case.

To conclude, our robustness analysis confirms the overall decline in mobility over the whole sample period, and especially the sharp increase in elasticities between 1900 and 1920, for both sons and sons-in-law. The sign of the trend in father/son elasticity after 1920 appears to be somewhat sensitive to the measure of occupational income used.

V What factors can explain the trends?

A Changes in Fertility

The total fertility rate of white women gradually dropped from 5.42 in 1850 to 2.22 in 1940 (Haines, 2008). The drop in fertility is likely to have affected the ability of parents to invest in their children's human capital: a larger family size is associated with lower human capital investment per child. The impact of this change on the intergenerational elasticity is not clear-cut and it will depend on how the income-fertility gradient changes over time. The observed elasticity would increase if the fertility decline occurs earlier for the high income group than for the low income group. In this case. the resources of high income parents would be split among fewer children, giving each one an even stronger initial advantage relative to children from lower income families. Jones and Tertilt (2008) document that the fertility transition did in fact occur earlier for high socio-economic status women. The fertility gap across socioeconomic groups was widest for the cohorts born between 1860 and 1900, implying that it should have been these cohorts to experience the largest increase in intergenerational elasticity. However, according to our estimates the jump occurs for cohorts born at the beginning of the 20$^{\text{th}}$ century. Thus changes in fertility do not seem to be able to match up with the timing of the observed trends.

We further assess this point by directly controlling for fertility in our baseline regressions. Ideally, to account for changes in fertility and for potential asymmetries in the allocation of family resources across children, we could control for the number of siblings and birth order. However, information on these variables is not available in the adult sample. Therefore, we control for the average number of siblings and the distribution of birth orders by first name in the children's sample. The results are reported in Table 6. For sons, the differences relative to the baseline results are minimal, with the possible exception of the first two cohorts. For sons-in-law, there is a consistent pattern of slightly lower estimated elasticities when controlling for family size, but the overall pattern of coefficients over time is unchanged.

B Migration

The sample period that we analyze was characterized by dramatic migratory flows from outside of the US. The very notion of the "American Dream" is based on the belief that migration serves as one of the main engines of social mobility. According to this view, immigrants with very few resources were quickly able to rise through the social ranks and take advantage of the opportunities available in the New World. It follows that mobility should be positively correlated with the size of the migration flows.

While this hypothesis is appealing at first glance, it appears to be inconsistent with the evolution over time in our intergenerational elasticity estimate. Immigration to the US had an early peak in the 1880s and then a second, larger peak between 1900 and 1915.[22] If immigration plays a major role in driving the overall level of mobility, and, in particular, the

[22]U.S. Department of Homeland Security, *Yearbook of Immigration Statistics* (various years).

children of immigrants are the ones who are able to climb up the social ladder most rapidly, then we should observe a large *drop* in intergenerational elasticity for the cohorts that came of age after the turn of the Century. This stands in stark contrast to the large increase in elasticity that we actually observe for the 1900 and 1910 cohorts.

Of course, it is possible that immigration contributed to attenuate what would have otherwise been an even larger decrease in intergenerational mobility. This would be the case if immigrant fathers tend to be employed in low-paying occupations, but their children quickly rise through the social ranks. To assess this possibility, we control for the immigrant status of sons and fathers (when possible) in our baseline regression. For the son-in-law specification we control for the immigrant status of both spouses and their fathers.

The results are presented in the second and third row of Panels A (sons) and B (sons-in-law) in Table 7. Both father/son and father/son-in-law elasticities are somewhat lower for the first three cohorts, but are then almost identical to the benchmark estimates for the latter three cohorts. These results arise because in the early part of the sample period, immigrants (both fathers and sons) were substantially less likely to be employed in farming occupations, and hence tended to have higher occupational income, than natives. Thus controlling for immigrant status has only a very modest effect on our estimates and, if anything, the adjusted estimates go in the "wrong" direction. We conclude that the trends in intergenerational elasticity are unlikely to be driven by changes in immigration over the sample period.[23,24]

C Internal mobility

This historical period was also characterized by dramatic migratory flows within the US. Long and Ferrie (2013) argue that residential mobility, either across state or county lines, is a prime candidate to explain the high level of intergenerational mobility in the US in the

[23] One important caveat to this conclusion: our estimates can only capture the degree of intergenerational mobility in *occupational* status. We cannot rule out that there was substantial intergenerational mobility *within* occupations (e.g., an immigrant father starts out setting up a small construction firm, and the son goes on to build a large empire in the construction industry) and that this might explain the trends.

[24] For immigrants, or children of immigrants, there may potentially be a problem in that their names are not stable over time. For example, a child named "Giuseppe" may become "Joseph" as an adult. The main consequence of this phenomenon would likely be to exacerbate the attenuation bias of our estimator, especially for cohorts with a high incidence of immigrants. We are reassured by the fact that controlling for immigrant status, or restricting the whole analysis to exclude all immigrants, or even all children with immigrant fathers (results available upon request), leaves the results essentially unchanged. We should also note that coding names using the Soundex algorithm (see Online Appendix Table A1) in parts take care of the Giuseppe/Joseph ambiguity and delivers very similar results.

19$^{\text{th}}$ Century, both relative to Britain during the same time period and relative to the US a century later. The argument is that residential mobility is itself a form of investment, which can improve a child's chances for occupational mobility in the same way as a human capital investment. Moreover, the 19$^{\text{th}}$ Century US was characterized by large opportunities for locational arbitrage, as the degree of regional specialization was at its peak (Kim, 1998).

Prima facie, there is some support for the notion that the trends in our estimates can be explained by patterns of internal mobility. The fraction of individuals aged 20-35 living in a state different from their state of birth decreased between 1850 and 1900 from 37% to 28%, but then remained at that level between 1900 and 1940.[25] Therefore, the trends in mobility across states are broadly consistent with the trends in intergenerational elasticity: the elasticity was low when mobility was high, and vice versa.

If much of intergenerational mobility is driven by children of low socioeconomic status "moving to opportunity" by crossing state lines, controlling for internal mobility should explain the trends. To further investigate this hypothesis, in the remaining rows of Table 7, we add to our basic specification controls for internal migrant status of both generations. We define internal migrants as individuals living in a different state than their state of birth. Contrary to our conjecture, the inclusion of these controls has essentially no effect on the intergenerational elasticity estimates. If anything, as with migration, the adjusted estimates for the first three cohorts decline suggesting an even larger jump in intergenerational elasticity than that implied by our baseline estimates.

Based on this evidence it does not appear that inter-state mobility explains much of the trend in the intergenerational elasticity estimates. However, as documented in Long and Ferrie (2013), of greater importance for geographic mobility were the movements across counties within a state. They estimate that between 1870 and 1880 the fraction of 20-29 years old white, native born males who changed county was twice as large as the fraction who changed states (50% vs. 26%). Because of lack of data, unfortunately, we cannot control for internal mobility at the county level. Therefore, we cannot rule out that the decline in intergenerational mobility can be explained by the decline in inter-county mobility.

D Regional Differences

The trends in intergenerational elasticity could also be explained by geographic differences in the degree of economic development. The industrial revolution did not spread uniformly across the United States. Regional income diverged significantly in the second half of the

[25] Source: our own calculations from the IPUMS samples.

19th Century. Income per capita in the South fell sharply during the Civil War, absolutely and relatively to other regions, and recovered at a slow pace. By 1900 income per capita in the South was barely half of the national average (Kim and Margo, 2004, p. 2991). If fathers and sons tend to live in the same region (and we have seen that geographic mobility was on the decline in the first part of the 20th Century), large economic disparities across regions will translate into a high correlation between father's and son's income.[26]

To assess this possibility, we include in our basic regressions controls for state of residence. The results are presented in the second row of Table 8. In all years, controlling for state of residence substantially lowers the estimated intergenerational elasticity. In the individually linked data controlling for state of residence also reduces the estimated elasticity, albeit by a smaller amount.[27] Our methodology may in part be responsible for the sensitivity of the estimates to geographic controls. If first names exhibit distinctive geographic patterns, the raw pseudo-panel estimator may already reflect regional differences in economic development. However, it is also true that one of the reasons for the transmission of economic status across generations is that fathers and sons tend to be located in the same geographic region and therefore their economic outcomes will be correlated. Which one is the better measure of mobility (the national or controlling for state or region) is a matter of interpretation. Nevertheless, even after controlling for state of residence, we still observe a sizable change in mobility between 1900 and 1920.[28] Therefore, it appears that regional differences in economic development can explain some, but not all, of the decline in mobility in this period.

To further understand the role of regional differences, we conduct our analysis separately for each region of birth. Specifically, for every individual born in a specific region in Census year t, we proxy his father's income by the average income of fathers of children with that first name in Census year $t-20$, and who lived in the same region. The results are presented in the second and fourth panels of Table 8. The region-specific intergenerational elasticity is almost always lower than the national elasticity, providing further support for the notion that part of the national estimate is accounted for by regional differences in development. There is also a fairly stable ranking of regions in terms of elasticity after 1880, with the

[26] A similar argument is made by Page and Solon (2003), who show that much of the correlation in adult earnings of neighboring boys can be explained by the large earnings differential between urban and non-urban areas combined with the strength with which urban status in childhood predicts urban status in adulthood.

[27] In years in which both are available, $\hat{\eta}_{PSEUDO}$ drops by about 12 log points while $\hat{\eta}_{LINKED}$ declines by about 5 log points.

[28] Notably, the percentage change in elasticity between 1900 and 1920 with controls for state of residence is quite sizable: about a 30 percent increase for men, and a 27 percent increase for women, with the latter number actually larger than the corresponding increase in the specification without controls.

Northeast being the most mobile, the South the least mobile, and the Midwest somewhere in between.[29] Similar patterns are observed for the father/son-in-law elasticities.

The ranking across regions of the elasticity estimates could also be interpreted in light of regional differences in compulsory schooling and investments in public education. The intergenerational elasticity is lowest in the Northeast, where all states had compulsory schooling in 1900 (Lingwall, 2010); is highest in the South, where only 3 of 16 states (plus the District of Columbia) had compulsory schooling; and is somewhere in the middle in the Midwest, where 10 of 12 states had introduced compulsory schooling by 1900. These results are consistent with Solon (2004), who shows, based on modern data, that intergenerational income elasticity decreases with the progressivity of public investment in human capital. This conclusion, however, is not without caveats: it rests on the assumption of substitutability between public and private investment in education. Parman (2011), however, argues that in the early 20th Century, the expansion of public schooling may actually have led to a decline in intergenerational mobility because the wealthy were better able to take advantage of it. While Parman's explanation does not fit well the cross-regional comparison, it is actually consistent with our finding of a national decline in mobility between 1900 and 1920.

E Human Capital and Inequality

What additional factors might contribute to explain the trends in elasticities after accounting for regional differences in economic development and geographic mobility?

In the modern context, there has been an upsurge of interest in the link between inequality and intergenerational mobility. Solon (2004) illustrates the tight link between the return to human capital and the intergenerational elasticity on a theoretical level. Corak (2013) documents a strong cross-sectional correlation between Gini coefficients and intergenerational elasticities in a sample of 22 countries. Therefore, the increase in intergenerational elasticities between 1900 and 1920 could be explained by an increase in inequality or an increase in the return to human capital.

There are a number of pieces of evidence showing that inequality did increase between the late 19th and early 20th centuries peaking in the 1920s. Piketty (2014) documents that the top decile share of wealth in the U.S. increased substantially between 1870 and 1920, and then dropped in the following two decades. A similar pattern is observed for the top decile share of income which drops in 1940. Cvrcek (2012) shows that men's career prospects, measured

[29] The 1940 estimate for the Midwest is remarkably similar to the one obtained by Feigenbaum (2014) based on matching fathers from the 1915 Iowa Census to their sons in the 1940 Federal Census.

by occupational upgrading, improved substantially between 1880 and 1930. Katz and Margo (2014) document a substantial increase in the share of white collar jobs in the overall economy between 1850 and 1910 (from 6.9% to 19.7%), and a contemporaneous upward trend in relative wages of white collar workers relative to common laborers and artisans. Margo (1999) provides further evidence of a long-term rise in the returns to educated labor beginning before the Civil War and continuing until the turn of the 20$^{\text{th}}$ Century. This was followed by a decline in the returns to education associated with the massive expansion of secondary schooling dating to the 1910s (Goldin, 1999, and Goldin and Katz, 2008).

Changes in labor market returns to human capital can also explain the trend in father/son-in-law elasticity. In the framework of matching tournament models of marriage with pre-marital investment (Peters and Siow, 2002; Bhaskar and Hopkins, 2011), the rate of return to female human capital is determined endogenously as a function of male returns to human capital and marriage market conditions. In a society where women do not work, the incentive to invest for girls increases with the labor market returns of boys. Thus the improvement in men's labor market outcomes would be consistent with our finding that the father/son and father/son-in-law elasticity share a common trend over the period of interest.

However, there are periods where the two elasticities diverge. For example, the father/son-in-law elasticity is greater than the father/son elasticity between 1880 and 1920 and then dips below it in 1930 and 1940. This divergence could be driven by gender differences in the informational content of first names across time periods. For example, Table 2 shows that names typical of recently arrived eastern European Jews circa 1900 (such as Abraham, Max and Nathan) rose to the top of the prestige scale in 1910 and 1920. No similar pattern emerges for female names. However, as shown in the last column of Table 1, for the overall population there is no evidence that the informational content of male names, relative to female names, exhibited an abnormal increase in these years.

Alternatively, differences between male and female elasticities may be driven by changes in the sex ratio (defined as the ratio of men to women), which affects the relative position of women in the marriage market. As women become scarce, even lowest quality women become desirable and can fetch a high quality mate. This would push the return to female human capital down.[30] A similar but opposite argument holds if there is a decline in the sex

[30] On the other hand, the increased competition on the male side of the market leads to male over-investment in human capital and, as a result, an increase in the variance of the quality of potential husbands. This, in turn, increases women's incentives to invest in human capital, pushing up the returns. Bhaskar and Hopkins (2011) show that the net effect on female returns to human capital is negative. The sex ratio imbalance induces a greater investment by the abundant sex.

ratio.

It follows that historical episodes and trends in fecundity and immigration that affected the sex ratio may help explain differences in the evolution of the father-son and father-daughter elasticity over the sample period.

Differential fecundity by gender implies that marriageable women are scarce and this affects their relative power in the marriage market (Siow, 1998). The scarcity of fecund women is especially important when infant and maternal mortality are high and people have more children. Both infant mortality and fertility were very high in 1850 but plummeted by the early decades of the 20$^{\text{th}}$ Century (Haines, 2008).[31] Maternal mortality declined from 850 deaths per 100,000 births in 1900 to 660 by 1917 (Loudon, 1992). These developments would lead to an increase in the number of eligible women and thus to a decline in the sex ratio, and, consequently, a higher return to female investment.

The large imbalance in the sex ratio induced by the Civil War, especially in the South, can rationalize the divergence between father/son-in-law and father/son elasticities around 1880 (Table 8). While the father/son elasticity in the South collapses between 1870 and 1880, the corresponding decline in father/son-in-law elasticity is much more modest. This is consistent with women becoming more abundant in the South, therefore gaining a stronger position in the marriage market and benefiting more from their human capital investment.[32]

The large migratory flows during this period may also have generated an imbalance in the sex ratio and increased heterogeneity of the pool of marriageable men. Haines (1996) shows that immigration to the US peaked in the opening decades of the 20$^{\text{th}}$ Century and was heavily skewed towards white males. Bandiera, Rasul and Viarengo (2012) show that the ratio of male to female immigrants spiked after the 1917 Immigration Act, which led to relatively higher barriers to entry for women. By raising the sex ratio this development may have lowered women's return to investment. This is consistent with the dip in the father/son-in-law elasticity in 1930 and 1940.[33]

[31] The infant mortality rate was 216.8 per 100,000 births in 1850, 110.8 in 1900 and 60 in 1930.

[32] For the importance of a war-induced imbalance in the sex-ratio on women's marriage outcomes, see also Abramitzky, Delavande and Vasconcelos (2011) study of post-World War I France.

[33] This decline in the father/son-in-law elasticity may also be related to the increase in married women's labor force participation rate during this period, from less than 10% up to 1920 to 17% in 1940. The increase in labor market opportunities for women is likely to have dampened the marriage market returns to human capital investment.

VI Conclusion

In this paper we have provided a consistent and continuous estimate of intergenerational elasticity for both sons and sons-in-law between 1870 and 1940. We find that the father/son elasticity was relatively flat throughout the second half of the 19^{th} Century, increased sharply between 1900 and 1920, and declined slightly between 1920 and 1940. Overall there was a marked increased in elasticity between the beginning and the end of the period. The father/son-in-law elasticity broadly follows the same trend, with some differences in timing, and drops below the father/son elasticity at the end of the sample period.

Our analysis offers a new perspective on intergenerational mobility in the United States in the late 19^{th} and early 20^{th} centuries, by allowing us to calculate the degree of social mobility for both genders, and at several points in time. We are therefore able to identify more precisely the key inflection point in the evolution of mobility. An exploration of historical, demographic and economic trends suggests that regional differences in economic development and fluctuations in income and wealth inequality were the main factors driving the trends.

The methodology developed in this paper can also be applied to other settings: intergenerational mobility across multiple generations (Olivetti, Paserman and Salisbury, 2014), as well as intra-generational mobility, assortative mating, and life-cycle patterns of occupational status and fertility. Of course, as long as information on first names is available, this methodology can be equally applied to other countries.

References

[1] Aaronson, Daniel and Bhashkar Mazumder (2008). "Intergenerational Economic Mobility in the United States, 1940 to 2000," *Journal of Human Resources*, 43(1), pp. 139-172.

[2] Abramitzky, Ran; Adeline Delavande, and Luis Vasconcelos (2011). "Marrying Up: The Role of Sex Ratio in Assortative Matching." *American Economic Journal: Applied Economics*, 3(3), pp. 124-157.

[3] Bandiera, Oriana; Imran Rasul and Martina Viarengo (2012). "The Making of Modern America: Migratory Flows in the Age of Mass Migration," *Journal of Development Economics*, http://dx.doi.org/10.1016/j.jdeveco.2012.11.005.

[4] Becker, Gary S. and Nigel Tomes (1979). "An equilibrium theory of the distribution of income and intergenerational mobility," *Journal of Political Economy*, 87(6), pp. 1153-89.

[5] Becker, Gary S. and Nigel Tomes (1986). "Human Capital and the Rise and Fall of Families," *Journal of Labor Economics*, 4(3), pp. S1-39.

[6] Bertrand, Marianne and Sendhil Mullainathan (2004)."Are Emily and Greg more Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination, " *American Economic Review*, 94(4), pp. 991-1013.

[7] Bhaskar, V. and Ed Hopkins (2011) "Marriage as Rat Race: Noisy Pre-Marital Investments with Assortative Matching," CEPR Discussion Paper No. 8529.

[8] Biavaschi, Costanza; Corrado Giulietti, and Zahra Siddique (2013). "The Economic Payoff of Name Americanization," IZA Discussion Paper No. 7725.

[9] Björklund, Anders and Markus Jäntti (1997). "Intergenerational Income Mobility in Sweden Compared to the United States." *American Economic Review*, 87(5), pp. 1009-1018.

[10] Black, Sandra E. and Paul J. Devereux (2011). "Recent Developments in Intergenerational Mobility." in (O. Ashenfelter and D. Card, eds.), *Handbook of Labor Economics*, Volume 4B, pp. 1487-1541, Amsterdam: North-Holland.

[11] Browning, Martin; Angus Deaton and Margaret Irish (1985). "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle," *Econometrica*, 53(3), pp. 503-544.

[12] Chadwick, Laura and Gary Solon (2002). "Intergenerational Income Mobility among Daughters," *American Economic Review*, vol. 92(1), pp. 335-44.

[13] Clark, Gregory (2014) *The Son Also Rises: Surnames and the History of Social Mobility.* Princeton: Princeton University Press.

[14] Clark, Gregory and Neil Cummins (2012) "What is the True Rate of Social Mobility? Surnames and Social Mobility in England, 1800-2012," manuscript, UC Davis.

[15] Clark, Gregory; Daniel Marcin; Firas Abu-Sneneh; Wilfred M. Chow; Kuk Mo Jung; Ariel M. Marek and Kevin M. Williams (2012). "Social Mobility Rates in the USA, 1920-2010: A Surname Analysis." Manuscript, UC Davis.

[16] Collado, M. Dolores; Ignacio Ortuño Ortín and Andrés Romeu (2012). "Long-run Intergenerational Social Mobility and the Distribution of Surnames," manuscript, Universidad de Alicante.

[17] Cook, Lisa D.; Trevon D. Logan and John M. Parman (2014) "Distinctively Black Names in the American Past," *Explorations in Economic History*, 53 (July), pp. 64-82.

[18] Corak, Miles (2013). "Inequality from Generation to Generation," in (R. Rycroft, ed.) *The Economics of Inequality, Poverty, and Discrimination in the 21st Century*, ABC-CLIO.

[19] Cvrcek, Tomas (2012) "America's Settling Down: How Better Jobs and Falling Immigration led to a Rise in Marriage, 1880-1930" *Explorations in Economic History*, 49(3), pp. 335-351.

[20] Duncan, Otis D. (1966). "Methodological Issues in the Analysis of Social Mobility," in (N. J. Smelser, S. M. Lipset, eds.) *Social Structure and Mobility in Economic Development*, pp. 51-97, Chicago: Aldine.

[21] Erikson, Robert and John H. Goldthorpe (1992). *The Constant Flux: A Study of Class Mobility in Industrial Societies.* Oxford: Clarendon Press.

[22] Feigenbaum, James J. (2014). "A New Old Measure of Intergenerational Mobility: Iowa 1915 to 1940." Mimeo., Harvard University.

[23] Ferrie, Joseph P. (2005). "History Lessons: The End of American Exceptionalism? Mobility in the United States since 1850," *Journal of Economic Perspectives*, 19(3), pp. 199-215.

[24] Ferrie, Joseph P. and Karen Rolf (2008). "The May-December Relationship since 1850: Age Homogamy in the United States", manuscript, Northwestern University.

[25] Fitch, Catherine A. and Steven Ruggles (2000). "Historical trends in marriage formation: The United States 1850-1990," in (L.J. Waite, C. Bachrach, M. Hindin, E. Thomson and A.Thornton, eds.) *The Ties that Bind: Perspectives on Marriage and Cohabitation*, pp. 59-88, Hawthorne: Aldine de Gruyter.

[26] Fryer, Roland G. and Steven D. Levitt (2004). "The Causes and Consequences of Distinctively Black Names." *Quarterly Journal of Economics*, 119 (3), pp.767-805.

[27] Goldin, Claudia (1999). "Egalitarianism and the Returns to Education during the Great Transformation of American Education," *Journal of Political Economy* 107 (December): S65-S94.

[28] Goldin, Claudia and Lawrence Katz (2008). *The Race between Education and Technology*. Cambridge, MA: The Belknap Press of Harvard University Press.

[29] Goldin, Claudia and Robert A. Margo (1992). "The Great Compression: The Wage Structure in the United States at Mid-century," *Quarterly Journal of Economics*, 107 (1), pp. 1-34.

[30] Güell, Maia; José V. Rodríguez-Mora and Chris Telmer (2007). "Intergenerational Mobility and the Informative Content of Surnames," manuscript, University of Edinburgh.

[31] Haan, Michael (2005). "Studying the Impact of Religion on Fertility in Nineteenth-Century Canada: The Use of Direct Measures and Proxy Variables," *Social Science History*, 29 (3), pp. 373-411.

[32] Hacker, J. David (1999). "Child Naming, Religion, and the Decline of Marital Fertility in Nineteenth-Century America," *The History of the Family: An International Quarterly*, 4 (3): pp. 339-65.

[33] Haines, Michael R. (1996). "Long-Term Marriage Paterns in the United States from Colonial Times to the Present," *The History of the Family*, 1 (1): pp. 15–39.

[34] Haines, Michael (2008). "Fertility and Mortality in the United States," in (R. Whaples ed.) EH.Net Encyclopedia, http://eh.net/encyclopedia/fertility-and-mortality-in-the-united-states/.

[35] Head, Keith and Thierry Mayer (2008). "Detection Of Local Interactions From The Spatial Pattern Of Names In France," *Journal of Regional Science*, 48(1): pp. 67-95.

[36] Hellerstein, Judith K. and Melinda S. Morrill (2011). "Dads and Daughters: The Changing Impact of Fathers on Women"s Occupational Choices," *Journal of Human Resources*, 46(2): pp. 333-372.

[37] Inoue, Atsushi and Gary Solon (2010). “Two-sample Instrumental Variables Estimators,” *Review of Economics and Statistics*, 92(3), pp. 557-561.

[38] Jäntti, Markus; Brent Bratsberg, Røed Knut; Oddbjørn Raaum; Robin A. Naylor; Eva Österbacka; Anders Björklund and Tor Erkisson (2006). “American Exceptionalism in a New Light: A Comparison of Intergenerational Earnings Mobility in the Nordic Countries, the United Kingdom and the United States,” Discussion Paper no. 1938, Institute for the Study of Labor (IZA), Bonn.

[39] Jones, Larry E. and Michèle Tertilt (2008). “An Economic History of Fertility in the United States: 18261960,” in (Peter Rupert ed.) *Frontiers of Family Economics,* Volume 1, pp. 165-230, Emerald Press.

[40] Katz, Lawrence F. and Robert A. Margo (2014). “Technical Change and the Relative Demand for Skilled Labor: The United States in Historical Perspective,” in *Human Capital in History: The American Record*, edited by L. Platt Boustan, C. Frydman and R.A. Margo, University of Chicago Press.

[41] Kim, Sukko (1998). “Economic Integration and Convergence: U.S. Regions, 1840-1990,” *Journal of Economic History*, 58 (3): pp. 659-683.

[42] Kim, Sukko and Robert A. Margo (2004). “Historical Perspectives on U.S. Economic Geography,” in (V. Henderson and J. Thisse, eds.) *Handbook of Regional and Urban Economics*, Volume 4: Cities and Geography, pp. 2982-3019, Amsterdam: North-Holland.

[43] Lingwall, Jeff (2010). “Compulsory Schooling, the Family, and the ‘Foreign Element’ in the United States, 1880-1900, ”manuscript, Carnegie Mellon University, http://www.heinz.cmu.edu/research/372full.pdf.

[44] Long, Jason and Joseph P. Ferrie (2007). “The Path to Convergence: Intergenerational Occupational Mobility in Britain and the US in Three Eras,” *Economic Journal*, 117 (519), pp. C61-C71.

[45] Long, Jason and Joseph P. Ferrie (2013). “Intergenerational Occupational Mobility in Britain and the U.S. Since 1850,” *American Economic Review*, 103(4), pp. 1109-1137.

[46] Loudon, Irvine (1992). *Death in Childbirth: An International Study of Maternal Care and Maternal Mortality 1800-1950*, Oxford: Clarendon Press.

[47] Margo, Robert A. (1999). "The History of Wage Inequality in America, 1920 to 1970," Jerome Levy Economic Institute Working Paper No. 286.

[48] Minnesota Population Center and Ancestry.com (2013). IPUMS Restricted Complete Count Data: Version 1.0 [Machine-readable database]. Minneapolis: University of Minnesota.

[49] Olivetti, Claudia; M. Daniele Paserman and Laura Salisbury (2014). "Three-generation Mobility in the United States, 1850-1940: The Role of Maternal and Paternal Grandfathers" Mimeo., Boston University.

[50] Olivetti, Claudia; M. Daniele Paserman (2013). "In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1930." NBER WP 18822.

[51] Page, Marianne E. and Gary Solon (2003). "Correlations between Brothers and Neighboring Boys in Their Adult Earnings: The Importance of Being Urban," *Journal of Labor Economics*, 21(4), pp. 831-855.

[52] Parman, John M. (2011). "American Mobility and the Expansion of Public Education," *Journal of Economic History*, 71(1) 105-132.

[53] Peters, Michael and Aloysius Siow (2002). "Competing Premarital Investments." *Journal of Political Economy*, 110(3), pp. 592-608.

[54] Piketty, Thomas (2014). *Capital in the 21st Century*. Cambridge: Harvard University Press.

[55] Preston, Samuel H. and Michael R. Haines (1991). *Fatal Years: Child Mortality in Late Nineteenth Century America*, Princeton, NJ: Princeton University Press.

[56] Raaum, Oddbjørn; Bernt Bratsberg, Knut Røed, Eva Österbacka, Tor Eriksson, Markus Jäntti and Robin A. Naylor (2007). "Marital Sorting, Household Labor Supply, and Intergenerational Earnings Mobility across Countries," *The B.E. Journal of Economic Analysis & Policy*, Vol. 7: Iss. 2 (Advances), Article 7.

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

[58] Siow, Aloysius (1998). “Differential Fecundity, Markets, and Gender Roles,” *Journal of Political Economy*, vol. 106(2), pp. 334-354.

[59] Solon, Gary R. (1992). “Intergenerational Income Mobility in the United States,” *American Economic Review*, 82 (3), pp. 393-408.

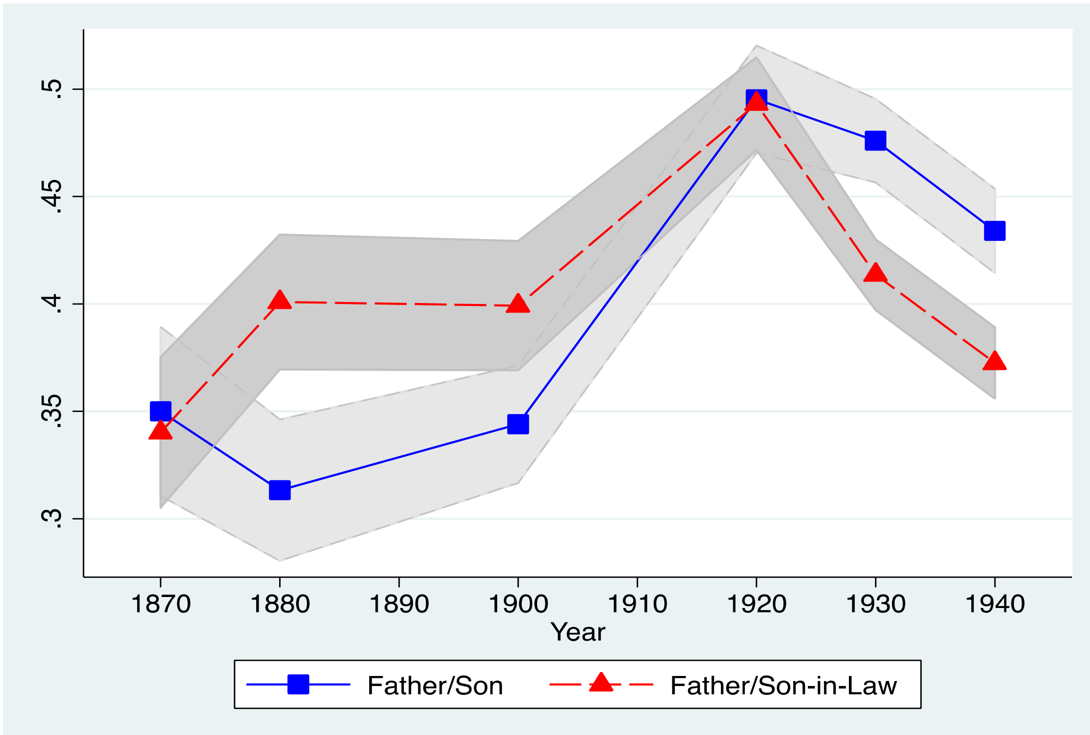
[60] Solon, Gary R. (1999). “Intergenerational Mobility in the Labour Market,” in (O. Ashenfelter and D. Card, eds.) *Handbook of Labor Economics*, Volume 3A, pp. 1761-800, Amsterdam: North-Holland.

[61] Solon, Gary R. (2004). “A Model of Intergenerational Mobility Variation over Time and Place,” in (M. Corak, ed.) *Generational Income Mobility in North America and Europe*, Cambridge: Cambridge University Press.

[62] Sutch, Richard (2011). “Hard Work, Nonemployment, and the Wealth-Age Profile: Evidence of a Life-Cycle Strategy in the United States during the Nineteenth Century,” available at http://users.nber.org/confer/2011/DAEs11/Sutch.pdf.

[63] Xie, Yu and Alexandra Killewald (2011). “Has Social Mobility in America Declined? Data, Methods, and Farming,” manuscript, University of Michigan.

Figure 1: Father/Son and Father/Son in Law Elasticities in Occupational Income: 1870-1940



Note: The figure presents point estimates and 90% confidence intervals for the father/son and father-son-in-law intergenerational elasticities. The values on the horizontal axes represent the year from which the son's (son-in-law's) sample are drawn. The elasticities are obtained from a regression of son (son-in-law) log occupational income on imputed father's (father-in-law's) log occupational income. See text for details of the imputation procedure. Occupational income is based on average earnings in the occupation

Table 1. Summary Statistics for Children's Names: 1850-1920

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Number of children ages 0-15	Number of distinct names	Mean number of observations per name	Percent of names that are singletons	Percent of children with unique names	Percent of children with names linked 20 years later	Share with top-50 name	Share of total variation in log earnings explained by between name variation
Year				*Males*				
1850	35,597	3,524	10.1	71.9	7.1	92.6	0.692	0.134
1860	48,114	4,083	11.8	70.5	6.0	93.7	0.695	0.111
1870	58,039	4,582	12.7	69.4	5.5	-	0.698	0.105
1880	75,004	6,589	11.4	69.4	6.1	92.9	0.653	0.112
1900	103,817	9,696	10.7	71.0	6.6	92.8	0.564	0.126
1910	117,612	9,818	12.0	69.5	5.8	94.1	0.534	0.126
1920	139,109	12,272	11.3	71.4	6.3	92.5	0.519	0.136
				Females				
1850	34,272	3,442	10.0	71.9	7.2	92.4	0.698	0.136
1860	46,874	4,488	10.4	70.7	6.8	92.8	0.657	0.132
1870	55,739	5,206	10.7	71.1	6.6	-	0.619	0.136
1880	72,160	7,161	10.1	69.0	6.8	92.0	0.548	0.133
1900	101,516	10,081	10.1	70.9	7.0	92.3	0.474	0.153
1910	114,074	10,103	11.3	69.3	6.1	93.5	0.473	0.154
1920	134,418	12,895	10.4	71.1	6.8	89.9	0.466	0.166

Note: Column (7) shows the share of children that have one of the 50 most popular names, by gender. Column (8) shows the R^2 from a regression of father's log occupational income on a full set of name dummies. Unless noted otherwise, the source for this and all following Tables are the 1850 to 1920 Integrated Public Use Micro Samples of the US decennial population censuses (Ruggles et al., 2010).

Table 2: Common Names Given to Children, Ranked by Mean Father's Occupational Income, 1850-1920.

	1850	1860	1870	1880	1900	1910	1920
Males							
Rank:	*Most Prestigious*						
1	Edward	Walter	Harry	Paul	Donald	Abraham	Jerome
2	Frederick	Frank	Walter	Harry	Kenneth	Max	Irving
3	Edwin	Willie	Herbert	Frederick	Harold	Nathan	Jack
4	Charles	Louis	Theodore	Ralph	Morris	Vincent	Nathan
5	Franklin	Fred	Edward	Philip	Max	Edmund	Abraham
	Least Prestigious						
1	Jesse	Levi	Jesse	Luther	Luther	Jessie	Willie
2	Hiram	Isaac	Franklin	Ira	Dewey	Otis	Loyd
3	Isaac	Benjamin	Isaac	Isaac	Perry	Luther	Luther
4	Daniel	Andrew	Hiram	Willis	Virgil	Eddie	Jessie
5	David	Jacob	Martin	Charley	Ira	Charley	Otis
Females							
Rank:	*Most Prestigious*						
1	Emma	Ada	Bertha	Bessie	Dorothy	Eleanor	Betty
2	Alice	Kate	Jessie	Mabel	Marion	Marian	Jean
3	Anna	Lizzie	Grace	Helen	Helen	Dorothy	Jane
4	Isabella	Clara	Carrie	Ethel	Louise	Marion	Kathryn
5	Josephine	Fanny	Helen	Blanche	Marie	Virginia	Muriel
	Least Prestigious						
1	Sally	Amanda	Nancy	Nancy	Nancy	Sallie	Lela
2	Nancy	Nancy	Lucinda	Viola	Ollie	Addie	Maggie
3	Lucinda	Rachel	Rebecca	Martha	Nannie	Ollie	Ollie
4	Martha	Lucinda	Amanda	Rachel	Sallie	Mattie	Effie
5	Lydia	Martha	Martha	Amanda	Alta	Iva	Eula

(light grey) Exact name, nickname or alternative spelling appears more than once (most prestigious).
(dark grey) Exact name, nickname or alternative spelling appears more than once (least prestigious).

Notes: Entries in the table represent the five children names with the highest and lowest average father occupational score, by gender and Census year. Only names that appear at least 100 times are considered for the ranking.

Table 3. Intergenerational Elasticities in Occupational Income, 1870-1940.

	(1)	(2)	(3)	(4)	(5)	(6)
	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930	1920-1940
Sample:						
Sons: baseline	0.3500	0.3133	0.3440	0.4953	0.4756	0.4340
	(0.0240)	(0.0201)	(0.0166)	(0.0153)	(0.0119)	(0.0119)
	[37077, 1182]	[50847, 1478]	[80255, 2234]	[109079, 3253]	[122469, 3721]	[119406, 3866]
Son's Age 5-15	0.3406	0.2735	0.3174	0.4043	0.3890	0.3995
	(0.0301)	(0.0234)	(0.0197)	(0.0162)	(0.0131)	(0.0129)
	[24116, 853]	[32376, 1072]	[53156, 1581]	[75765, 2401]	[83051, 2787]	[82129, 2963]
Married Sons	0.2868	0.3433	0.3805	0.4715	0.4423	0.3765
	(0.0312)	(0.0261)	(0.0224)	(0.0178)	(0.0133)	(0.0125)
	[17912, 891]	[24510, 1155]	[36521, 1641]	[57570, 2586]	[67138, 3052]	[70751, 3175]
Sons in law: baseline	0.3402	0.4009	0.3992	0.4932	0.4143	0.3725
	(0.0214)	(0.0194)	(0.0185)	(0.0134)	(0.0102)	(0.0101)
	[23280, 976]	[30081, 1376]	[45804, 2063]	[68439, 2888]	[79319, 3328]	[77001, 3320]
Daughter's Age 5-15	0.3543	0.3563	0.3209	0.4489	0.3976	0.3546
	(0.0285)	(0.0218)	(0.0194)	(0.0146)	(0.0115)	(0.0113)
	[16650, 726]	[21774, 1027]	[34370, 1597]	[52532, 2264]	[60577, 2577]	[58967, 2555]
Sons in law 20-35	0.3283	0.4394	0.3860	0.4889	0.4151	0.3691
	(0.0251)	(0.0226)	(0.0220)	(0.0154)	(0.0118)	(0.0116)
	[15404, 840]	[20383, 1197]	[30533, 1712]	[46762, 2479]	[54600, 2885]	[54131, 2843]
Sons: Individually linked data						
OLS		0.4654	0.4743			
		(0.0175)	(0.0119)			
First name fixed effects		0.4628	0.4665			
		(0.0206)	(0.0133)			
		[3947]	[9076]			

Notes: Entries in the rows 1-6 represent OLS coefficients from a regression of son's (son-in-law's) log occupational income on imputed father's (father-in-law's) log occupational income. Standard errors in parentheses. In brackets, the number of observations used in each regression, and the number of distinct first names used to impute father's (father-in-law's) income. Row 7 represent the OLS coefficients from a regression of son's occupational income on father's occupational income using the IPUMS Linked Representative Samples, 1860-1880 and 1880-1900. Row 8 adds name fixed effects. Standard errors in parentheses, number of observations in brackets.

Table 4. Intergenerational Elasticities 1870-1940.
1900 Income Distribution and Farmers' Income.

	(1)	(2)	(3)	(4)	(5)	(6)
	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930	1920-1940
Log occupational income in:	**A: Fathers-Sons**					
1950	0.3500	0.3133	0.3440	0.4953	0.4756	0.4340
	(0.0240)	(0.0201)	(0.0166)	(0.0153)	(0.0119)	(0.0119)
1900	0.3502	0.3542	0.3823	0.4471	0.4432	0.4316
	(0.0222)	(0.0189)	(0.0155)	(0.0121)	(0.0101)	(0.0104)
1900, imputed farmer wage	0.3467	0.2879	0.3634	0.4660	0.4696	0.4779
	(0.0284)	(0.0229)	(0.0196)	(0.0150)	(0.0127)	(0.0137)
1950 excluding farmers	0.1899	0.1561	0.1463	0.2540	0.2919	0.2939
	(0.0298)	(0.0221)	(0.0171)	(0.0165)	(0.0128)	(0.0146)
1900 excluding farmers	0.2487	0.2075	0.2320	0.2992	0.2954	0.3333
	(0.0238)	(0.0196)	(0.0156)	(0.0122)	(0.0107)	(0.0124)
N, no. of names: 1950	[37077, 1182]	[50847, 1478]	[80255, 2234]	[109079, 3253]	[122469, 3721]	[119406, 3866]
N, no. of names: 1950 ex. Farmers	[26988, 741]	[36460, 943]	[65726, 1529]	[92664, 2337]	[109832, 2847]	[108086, 2947]
	B: Fathers-Sons in Law					
1950	0.3402	0.4009	0.3992	0.4932	0.4143	0.3725
	(0.0214)	(0.0194)	(0.0185)	(0.0134)	(0.0102)	(0.0101)
1900	0.3115	0.4229	0.4120	0.4900	0.4387	0.4139
	(0.0203)	(0.0192)	(0.0182)	(0.0126)	(0.010)	(0.0103)
1900, imputed farmer wage	0.2509	0.3161	0.3166	0.4415	0.4221	0.4269
	(0.0242)	(0.0205)	(0.0208)	(0.0146)	(0.0120)	(0.0128)
1950 excluding Farmers	0.2150	0.2003	0.1802	0.3270	0.3220	0.3496
	(0.0287)	(0.0182)	(0.0183)	(0.0158)	(0.0121)	(0.0128)
1900 excluding Farmers	0.1986	0.2290	0.2224	0.3490	0.3744	0.4016
	(0.0245)	(0.0182)	(0.0182)	(0.0151)	(0.0126)	(0.0135)
N, no. of names: 1950	[23280, 976]	[30081, 1376]	[45804, 2063]	[68439, 2888]	[79319, 3328]	[77001, 3320]
N, no. of names: 1950 ex. Farmers	[22586, 697]	[29344, 1004]	[44917, 1547]	[67488, 2313]	[78032, 2727]	[76028, 2757]

Notes: Entries in the table represent OLS coefficients from a regression of son's (son-in-law's) log occupational income on imputed father's (father-in-law's) log occupational income. Different rows use different measures of occupational income. See text for details of the 1900 occupational income measure, and the imputation procedure for farmer's income. Standard errors in parentheses. The number of observations used in each regression, and the number of distinct first names used to impute father's (father-in-law's) income are reported in brackets at the bottom of each panel.

Table 5. Intergenerational Elasticities 1870-1940.
Alternative Measures of Occupational Income.

	(1)	(2)	(3)	(4)	(5)	(6)
	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930	1920-1940
	A: Fathers-Sons					
1950 (baseline)	0.3500	0.3133	0.3440	0.4953	0.4756	0.4340
	(0.0240)	(0.0201)	(0.0166)	(0.0153)	(0.0119)	(0.0119)
1950 rank	0.2163	0.3556	0.3862	0.4562	0.4348	0.3782
	(0.0151)	(0.0196)	(0.0166)	(0.0134)	(0.0106)	(0.0104)
1850-1870 wealth	0.2967	0.3263	0.3386	0.4117	0.4127	0.3882
	(0.0218)	(0.0189)	(0.0164)	(0.0146)	(0.0124)	(0.0124)
1990	0.2571	0.2069	0.2388	0.3585	0.4156	0.3631
	(0.0255)	(0.0215)	(0.0185)	(0.0157)	(0.0134)	(0.0130)
SEI	0.2695	0.2979	0.3062	0.4599	0.4674	0.4418
	(0.0210)	(0.0196)	(0.0165)	(0.0148)	(0.0129)	(0.0125)
N, no. of names	[37077, 1182]	[50847, 1478]	[80255, 2234]	[109079, 3253]	[122469, 3721]	[119406, 3866]
	B: Fathers-Sons in Law					
1950 (baseline)	0.3402	0.4009	0.3992	0.4932	0.4143	0.3725
	(0.0214)	(0.0194)	(0.0185)	(0.0134)	(0.0102)	(0.0101)
1950 rank	0.2663	0.4285	0.4362	0.5101	0.4289	0.3883
	(0.0162)	(0.0188)	(0.0184)	(0.0138)	(0.0105)	(0.0105)
1850-1870 wealth	0.2589	0.379	0.3437	0.436	0.4101	0.3736
	(0.0204)	(0.0199)	(0.0201)	(0.0154)	(0.0128)	(0.0129)
1990	0.2137	0.2685	0.2586	0.4418	0.4009	0.3848
	(0.0228)	(0.0210)	(0.0216)	(0.0159)	(0.0127)	(0.0125)
SEI	0.1887	0.3243	0.3244	0.5097	0.4889	0.4488
	(0.0202)	(0.0208)	(0.0218)	(0.0157)	(0.0132)	(0.0125)
N, no. of names	[23280, 976]	[30081, 1376]	[45804, 2063]	[68439, 2888]	[79319, 3328]	[77001, 3320]

Notes: Entries in the table represent OLS coefficients from a regression of son's (son-in-law's) log occupational income on imputed father's (father-in-law's) log occupational income. Different rows use different measures of occupational income (see text for details). SEI is a constructed measure that assigns a Duncan Socioeconomic Index (SEI) score to each occupation using the 1950 occupational classification. Standard errors in parentheses. The number of observations used in each regression, and the number of distinct first names used to impute father's (father-in-law's) income are reported in brackets at the bottom of each panel.

Table 6. Fertility and Birth Order

	(1)	(2)	(3)	(4)	(5)	(6)
	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930	1920-1940
	A: Fathers-Sons					
Baseline	0.3500	0.3133	0.3440	0.4953	0.4756	0.4340
	(0.0240)	(0.0201)	(0.0166)	(0.0153)	(0.0119)	(0.0119)
Control for number of sibling	0.2836	0.2735	0.3444	0.5024	0.4738	0.4178
	(0.0255)	(0.0215)	(0.0169)	(0.0157)	(0.0121)	(0.0124)
Control for birth order	0.3277	0.2860	0.3433	0.4964	0.4644	0.4154
	(0.0247)	(0.0208)	(0.0167)	(0.0154)	(0.0120)	(0.0122)
N, no. names (baseline)	[37077, 1182]	[50847, 1478]	[80255, 2234]	[109079, 3253]	[122469, 3721]	[119406, 3866]
	B: Fathers-Sons in Law					
Baseline	0.3402	0.4009	0.3992	0.4932	0.4143	0.3725
	(0.0214)	(0.0194)	(0.0185)	(0.0134)	(0.0102)	(0.0101)
Control for number of sibling	0.292	0.3044	0.3949	0.4651	0.3821	0.3298
	(0.0241)	(0.0212)	(0.0192)	(0.0143)	(0.0111)	(0.0114)
Control for birth order	0.3289	0.3659	0.3962	0.4734	0.3961	0.3472
	(0.0217)	(0.0199)	(0.0186)	(0.0136)	(0.0106)	(0.0105)
N, no. names (baseline)	[23280, 976]	[30081, 1376]	[45804, 2063]	[68439, 2888]	[79319, 3328]	[77001, 3320]

Notes: Entries in the table represent OLS coefficients from a regression of son's (son-in-law's) log occupational income on imputed father's (father-in-law's) log occupational income. Number of siblings is imputed using the average number of siblings for individuals with a given first name. Controls for birth order are the share of individuals with a given first name that are first-born, second-born, and higher order. Standard errors in parentheses. The number of observations used in each regression, and the number of distinct first names used to impute father's (father-in-law's) income are reported in brackets at the bottom of each panel.

Table 7. Immigration and Internal Migration

	(1)	(2)	(3)	(4)	(5)	(6)
	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930	1920-1940
			A: Fathers-Sons			
Baseline	0.3500	0.3133	0.3440	0.4953	0.4756	0.4340
	(0.0240)	(0.0201)	(0.0166)	(0.0153)	(0.0119)	(0.0119)
Control for immigrant status:						
Son	0.2992	0.2769	0.3247	0.4705	0.4655	0.4247
	(0.0236)	(0.0198)	(0.0165)	(0.0152)	(0.0118)	(0.0120)
Son and father		0.2367	0.2883	0.4420	0.4365	
		(0.0195)	(0.0163)	(0.0151)	(0.0118)	
Control for internal migrant status:						
Son	0.2984	0.2766	0.3249	0.4708	0.4664	0.4256
	(0.0236)	(0.0198)	(0.0165)	(0.0151)	(0.0118)	(0.0120)
Son and father		0.2328	0.2862	0.4387	0.4339	
		(0.0195)	(0.0163)	(0.0150)	(0.0117)	
N, no. names (baseline)	[37077, 1182]	[50847, 1478]	[80255, 2234]	[109079, 3253]	[122469, 3721]	[119406, 3866]
			B: Fathers-Sons in Law			
Baseline	0.3402	0.4009	0.3992	0.4932	0.4143	0.3725
	(0.0214)	(0.0194)	(0.0185)	(0.0134)	(0.0102)	(0.0101)
Control for immigrant status:						
Son-in-law, daughter	0.2720	0.3625	0.3676	0.4773	0.4093	0.3687
	(0.0212)	(0.0192)	(0.0184)	(0.0134)	(0.0102)	(0.0102)
Son-in-law, daughter and fathers		0.3254	0.3122	0.4433	0.3821	
		(0.0190)	(0.0182)	(0.0133)	(0.0102)	
Control for internal migrant status:						
Son-in-law, daughter	0.2722	0.3619	0.3640	0.4733	0.4050	0.3629
	(0.0212)	(0.0192)	(0.0184)	(0.0133)	(0.0102)	(0.0101)
Son-in-law, daughter and fathers		0.3215	0.3051	0.4372	0.3750	
		(0.0189)	(0.0181)	(0.0132)	(0.0101)	
N, no. names	[23280, 976]	[30081, 1376]	[45804, 2063]	[68439, 2888]	[79319, 3328]	[77001, 3320]

Notes: Entries in the table represent OLS coefficients from a regression of son's (son-in-law's) log occupational income on imputed father's (father-in-law's) log occupational income. Immigrants are defined to be all those born outside of the United States. Internal migrants are those who live in a state different from their state of birth. Immigrant status and internal migrant status are taken from the individual level data. Father's immigrant status was not available in the 1870 Census and it is only available for a subset of observations in the 1940 Census. Standard errors in parentheses. The number of observations used in each regression, and the number of distinct first names used to impute father's (father-in-law's) income are reported in brackets at the bottom of each panel.

Table 8. Intergenerational Elasticities by Region of Birth.

	(1)	(2)	(3)	(4)	(5)	(6)
	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930	1920-1940
	A: Fathers-Sons					
All	0.3500	0.3133	0.3440	0.4953	0.4756	0.4340
	(0.0240)	(0.0201)	(0.0166)	(0.0153)	(0.0119)	(0.0119)
Control for state of residenc	0.2765	0.1943	0.2108	0.2746	0.2799	0.2539
	(0.0229)	(0.0189)	(0.0156)	(0.0142)	(0.0111)	(0.0116)
Northeast	0.2948	0.2539	0.1677	0.2187	0.1911	0.1639
	(0.0384)	(0.0337)	(0.0311)	(0.0279)	(0.0224)	(0.0248)
Midwest	0.1499	0.2521	0.2677	0.2771	0.2702	0.3481
	(0.0468)	(0.0369)	(0.0315)	(0.0279)	(0.0230)	(0.0230)
South	0.4593	0.1591	0.2878	0.3081	0.3631	0.2738
	(0.0564)	(0.0337)	(0.0312)	(0.0293)	(0.0229)	(0.0207)
N, no. names (All)	[37077, 1182]	[50847, 1478]	[80255, 2234]	[109079, 3253]	[122469, 3721]	[119406, 3866]
N, no. names (Northeast)	[11461, 580]	[14846, 672]	[19327, 727]	[23818, 891]	[29959, 1040]	[29883, 1053]
N, no. names (Midwest)	[7091, 442]	[12713, 629]	[25372, 1039]	[35418, 1406]	[38069, 1589]	[38897, 1524]
N, no. names (South)	[7709, 474]	[11481, 607]	[16570, 973]	[23490, 1558]	[30306, 1966]	[33909, 2035]
	B: Fathers-Sons in Law					
All	0.3402	0.4009	0.3992	0.4932	0.4143	0.3725
	(0.0214)	(0.0194)	(0.0185)	(0.0134)	(0.0102)	(0.0101)
Control of state of residence	0.2474	0.2947	0.2509	0.3199	0.2606	0.2577
	(0.0206)	(0.0183)	(0.0176)	(0.0129)	(0.0100)	(0.0100)
Northeast	0.2014	0.2221	0.3111	0.2743	0.2095	0.1698
	(0.0381)	(0.0384)	(0.0413)	(0.0336)	(0.0262)	(0.0267)
Midwest	0.3471	0.3811	0.3289	0.3371	0.3013	0.2768
	(0.0525)	(0.0356)	(0.0339)	(0.0239)	(0.0185)	(0.0181)
South	0.3975	0.3303	0.3192	0.4649	0.3787	0.3591
	(0.0483)	(0.0289)	(0.0308)	(0.0256)	(0.0180)	(0.0171)
N, no. names (All)	[23280, 976]	[30081, 1376]	[45804, 2063]	[68439, 2888]	[79319, 3328]	[77001, 3320]
N, no. names (Northeast)	[6602, 448]	[8102, 559]	[9741, 602]	[12819, 769]	[16866, 924]	[16562, 910]
N, no. names (Midwest)	[4877, 354]	[7883, 586]	[14957, 964]	[22529, 1340]	[24913, 1458]	[25170, 1367]
N, no. names (South)	[5337, 408]	[7200, 587]	[10413, 926]	[16556, 1335]	[21104, 1625]	[23275, 1559]

Notes: Entries in the table represent OLS coefficients from a regression of son's (son-in-law's) log occupational income on imputed father's (father-in-law's) log occupational income. Standard errors in parentheses. The region-specific elasticities are obtained by imputing father's income as the average income of fathers of children with a given first name who lived in the same region. At the bottom of each panel, the number of observations used in each regression, and the number of distinct first names used to impute father's (father-in-law's) income.