In the Name of the Father: Marriage and Intergenerational Mobility in the United States, 1850-1930.*

Claudia Olivetti and M. Daniele Paserman
Boston University and NBER

November 2012

Abstract

This paper provides a new perspective on intergenerational mobility in the United States in the late 19th and early 20th centuries. We develop a new methodology based on synthetic cohorts of individuals with the same first name. This methodology allows to calculate intergenerational elasticities not only between fathers and sons, but also between fathers-in-law and sons-in-law, something that is typically not possible with historical data. Thus, the paper sheds light on the role of marriage in the intergenerational transmission of economic status from a historical perspective.

We find that the father-son correlation in economic status grows throughout the sample period. The trend in father/son-in-law correlation is broadly similar, but it rises more markedly in the latter part of the 19th Century, and falls below the father/son correlation toward the end of the sample period. We argue that most of the increase in the intergenerational elasticity estimate in the early part of the 20th Century can be accounted for by the vast regional disparities in economic development.

Keywords: Intergenerational Mobility, Marriage, Assortative Mating
JEL codes: J62, J12

*We thank Dara Lee and Laura Salisbury for excellent research assistance; Robert Margo, Andrew Newman and Marc Rysman for many valuable comments and conversations; 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 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 and ECARES for many helpful suggestions.

1 Introduction

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. This paper constructs a long time series of intergenerational correlations in economic status, and highlights the complex interrelationship between intergenerational mobility and marriage institutions. Marriage, through the sorting of individuals into households, plays a critical role in the transmission of human capital and wealth across generations. In turn, the degree to which parents are able to pass on their endowments to their children also has consequences for the degree of assortative matching in the economy.

Previous research has found that in the United States the degree of intergenerational mobility declined markedly between the end of the 19th and the middle of the 20th century when it reached levels comparable to those observed in the United Kingdom.

This paper builds and improves on the previous research by constructing a consistent and continuous measure of intergenerational mobility. The key innovation of this paper is to develop a methodology that makes it possible to study the link between fathers-in-law and sons-in law as well as the link between fathers and sons (as is done in almost all the literature). It is important to add this dimension to reach a fuller understanding of the transmission of socioeconomic status across generations. To see this, consider two societies with the same degree of intergenerational correlation between fathers and sons, but that differ in the degree of assortative matching. Clearly, the correlation between the *average* well-being of the children's generation and the parent's generation will be higher in societies in which there is strong stratification in marriages by social class. By looking only the correlation between fathers and sons, one may be missing some important aspects of the transmission of socioeconomic status across generations.

The calculation of intergenerational correlations typically require the use of longitudinal data sets that link sons and daughters to their fathers. Such data are not consistently available over extended periods of time. Recently, historical longitudinal data sets based on Census data have been able to link fathers and sons by first and last names. However, it is impossible to link fathers and daughters in this manner because women change last name upon marriage.

Our methodology is based on a simple idea: in a random sample from any population, the set of individuals with a given first name is itself a random sample of the population with that first name. Hence, intergenerational correlations can be calculated as the correlation between the average outcomes of individuals with a given first name in year t and the average outcomes of the fathers of individuals with that name in year $t-k$. This same idea can be applied to link the average outcomes of fathers to those of their daughters' husbands (i.e., their sons in law).

To illustrate this idea, consider a simple example. Assume that the only possible names 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 10 and 15 percent of the total variation in father's socioeconomic status can be explained by the variation *between* names given to their children.

Our methodology is a modified version of a synthetic cohort approach where the cohorts are constructed by grouping individuals both by age and by first name. We assess its validity by conducting two exercises, where we take advantage of the 1850-1930 IPUMS Linked Representative Samples, which link individual fathers and sons.

First, we conduct a numerical exercise to study the behavior of our pseudo-panel estimator. We augment a standard model of intergenerational transmission of income with a process that assigns names based on socioeconomic status. The model parameters are estimated by minimizing the distance between key moments in the 1860-1880 data and their counterparts in the simulated data. Two important parameters in the model are the degree of concentration of names, and the extent to which names carry economic content. We show that the estimated intergenerational elasticity is not sensitive to these parameters. Therefore, observed movements in our estimator capture changes in the fundamentals of the income generating process as opposed to changes in features of the name distribution.

Second, we use data from the linked samples to compare estimates obtained with our methodology to measures of intergenerational mobility obtained using conventional methods. We find that our methodology yields remarkably similar intergenerational father-son elasticities, both in terms of levels and in terms of time trends.

The core of our analysis is the estimation of intergenerational elasticities for both sons and daughters. We use the full 1% IPUMS samples between 1850 and 1930 to compute five intergenerational coefficients at 20-year intervals, and four intergenerational coefficients at 30-year intervals. Our results indicate that: a) the intergenerational elasticity between fathers and sons at 20-year and 30-year intervals increases between 1850 and 1910, with most of the increase occurring after the turn of the century. These results are in accord with the findings of Ferrie (2005; 2007), who documents a marked decrease in intergenerational mobility in the United States between the late 19th century and the middle of the 20th century; b) the relationship in economic status between fathers in law and sons in law does not change substantially over the sample period; c) the father in law-son in law elasticity is always higher than the corresponding father-son elasticity. The gap declines over time, and almost disappears for the generation of children aged 0-15 in 1910. We argue that these results are consistent with an increase in returns to human capital in the labor market for men and a fairly stable balance of power between the sexes in the marriage market.

The main findings are robust to different methods of imputing income and coding names, to the use of different outcome variables, and to controlling for immigrant status, differential mortality across socioeconomic groups and selection into marriage. Intergenerational elasticities are considerably lower if we exclude farmers and within geographic regions. This suggests that part of the high degree of intergenerational mobility can be explained by the persistence of farming status and by regional differences in economic development.

The rest of the paper proceeds as follows. The next section discusses the relevant literatures on intergenerational mobility, on marital sorting, and on the informative content of names. Section 3 presents a simple model of marriage and intergenerational mobility that defines the parameters of interest and facilitates the interpretation of our empirical results. Section 4 describes the econometric methodology, and Section 8.1 verifies its validity. Section 5 presents the data and discusses measurement issues. The main results are presented in Section 6. Section **??** provides robustness checks and additional analysis. Section 10 concludes.

2 Related Literature

This paper lies at the intersection of several literatures. The main literature of reference studies intergenerational mobility, both historically and in modern times, for the US and across countries.

Intergenerational Mobility: Historical Data. The main contributions to the historical literature are those of Ferrie (1995, 2004), who uses the 1880 U.S. Census and extracts from the 1850-1910 public use Census samples to construct a large, nationally representative longitudinal data set and studies occupational mobility between fathers and sons. Based on this data, Ferrie (2005) shows that in the United States the degree of intergenerational mobility declined markedly between the end of the 19th and the middle of the 20th century ("the end of American exceptionalism"). Using a comparable data set for the United Kingdom, Long and Ferrie (2007 and forthcoming) show that although in the late 19th-century intergenerational occupational mobility was higher in the US than in the UK, the two countries converged over time.[1] Sacerdote (2005) also looks at the intergenerational transmission of economic status in 19th century and early 20th century, focusing on the comparison between the descendants of slaves and those of free blacks.

Intergenerational Mobility: Modern Data. There is an extensive literature that studies

[1] On the other hand, Xie and Killewald (2011) claim that the Long and Ferrie findings are due to a methodological artifact, as their main measure of mobility is ill-suited for measuring mobility of farmers, a group whose proportion in the population declined dramatically during the period of analysis.

intergenerational mobility using modern panel data sets. These data sets typically include information on education and labor market outcomes of individuals and their spouses, as well as background information on both sets of parents (see Solon, 1999, and references therein, for a comprehensive survey, and Black and Devereux, 2010, for a more recent assessment of the literature). The estimates of father-son intergenerational mobility obtained for the U.S. range from 0.13 to 0.54 with the median estimate of the elasticity between father's log labor income and son's log labor income hovering around 0.4. Fewer studies provide estimates of the father-daughter labor income elasticity. These estimates range from 0.11 to 0.54, with a median estimate of 0.31.

Only a very 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 in law - son in law elasticity- estimated at 0.360. This is in contrast with previous studies for the US that reported estimates of father/son-in-law income elasticities to be similar to the corresponding father/son elasticities (see Solon, 1999, for references). Lam and Schoeni (1993, 1994) compute correlations between son's income and the background characteristics (such as education) of father and father-in-law in Brazil. They find that the effect of father-in-law's schooling on wages is larger than the effect of father's schooling in Brazil, while the opposite is observed in the United States. Finally, Hellerstein and Morrill (2009) study trends in fathers-daughters occupational mobility across cohorts of US women, and document that the probability that a woman works in the same occupation as her father has increased over the course of the 20th Century.

Sorting, Intergenerational Mobility and Inequality. Very few papers have studied the link between sorting and intergenerational mobility. One exception is the paper by Ermisch, Francesconi and Siedler (2006) that shows that positive assortative mating can explain 40 to 50 percent of the covariance between parents' and own family income both in Germany and in the UK. Also, Raaum et al. (2007) compare the contribution of gender and marital status to intergenerational mobility in Denmark, Finland and Norway and in the US and the UK. They find that for married women mobility is higher in Nordic countries than in the US or the UK. More recently, Charles, Hurst and Killewald (2011) use PSID data to study the degree to which spouses sort in the marriage market on the basis of parental wealth. They find that in the US men and women marry spouses from similar parental wealth background and that this cannot be explained only on the basis of sorting by educational attainment.

The paper is also related to a large literature in economics and other fields on assortative mating. Since the pioneering work by Becker (1991) and Lam (1988) a large literature has developed that studies how and why individuals sort by education, income, ethnicity, religion, or nationality. Papers have looked at how sorting affects intergenerational transmission of cultural and religious traits (Bisin and Verdier, 2000); inequality and economic growth (Kremer, 1997;

Fernández and Rogerson, 2001; and Fernández, Guner and Knowles, 2005); and the transmission of genetic traits and its effect on inequality and growth (Galor and Moav, 2002).

Names. Finally, the paper is related to recent studies that have investigated the economic content and consequences of names. Fryer and Levitt (2004) and Bertrand and Mullainathan (2004) study the labor market effects of distinctively black names, while Aura and Hess (2010) examine the effects of a broader set of first name features on several lifetime outcomes. Goldin and Shim (2004) analyze the patterns of maiden name retention among married women. Head and Mayer (2008) investigate the social transmission of parental preferences through naming patterns. Hacker (1999) and Haan (2005) use biblical first names as a measure of religiosity and study its impact on fertility in 19th Century Canada and US, respectively.

Closely related to our project is the work by Guell et al. (2007), who use the informative content of family names 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. Similarly, Clark (2010) uses the distribution of surnames as a measure of long run social mobility in England, and concludes that the English society from the heart of the Middle Ages to today was a society without persistent social classes.

3 An Illustrative Model of Marriage and Mobility

We derive intergenerational links between son's income and father's income and between daughter's family income and father's income using utility-maximizing behavior by parents in the spirit of the model by Becker and Tomes (1979, 1986). A family containing two parents and two children, one male, one female, must allocate the parents' lifetime earnings between the parents own consumption and investment in the earning capacity of the children.

Formally, parental utility is defined over own consumption (c_{t-1}) and over the expected family income of their son ($Y_{M,t}$) and daughter when they are adults ($Y_{F,t}$). We assume consensus parental preferences. Parents choose how much of their resources to allocate to household consumption and how much to invest in their children's human capital. The human capital investment affects both earnings on the labor market and spouse's earnings through the marriage market. We assume that only men work in this economy. Consequently, the investment in the children's human capital affects the son's labor income directly and the daughter's family income indirectly through her spouse. Parental preferences are described by: $\beta_1 \log c_{t-1} + \beta_2 E\left[\log\left(Y_{M,t}\right)\right] + \beta_3 E\left[\log\left(Y_{F,t}\right)\right]$, where β_2 and β_3 measure parental altruism towards their son and daughter, respectively. Parents choose c_{t-1}, the investment in human capital of their son, $H_{M,t}$, and daughter $H_{F,t}$, to maximize utility subject to the budget constraint: $c_{t-1} + p_H\left(H_{M,t} + H_{F,t}\right) \leq y_{t-1}$ where p_H is the monetary cost of the investment in human capital and y_{t-1} is the father's labor income.

Labor Market:

Men's labor income depends on human capital according to the following expression:

$$y_t = H_{M,t}^{\gamma_1} \exp\left(E_{M,t}\right),$$

where $E_{M,t}$ represents the combined effect of all determinants, other than human capital, of a man's lifetime earnings. As in Becker and Tomes (1979, 1986) this term can be decomposed in two components:

$$E_{M,t} = e_t + u_t,$$

where e_t is the child's "endowment" of earning capacity and the iid stochastic term u_t, with zero mean and variance σ_u^2, is the son's luck on the labor market assumed to be independent of y_{t-1} and e_t. The child's endowment e_t follows the first-order autoregressive process:

$$e_t = \lambda e_{t-1} + v_t,$$

where $0 \leq \lambda < 1$ measures the persistence of the family endowment and v_t is serially uncorrelated with variance σ_v^2. The key parameter in this equation is γ_1, the rate of return to human capital on the labor market. Stationarity of the labor income process requires that $\gamma_1 \in (0, 1)$.

Marriage market

Women's spousal income depends on human capital according to the following functional form:

$$Y_{F,t} = y_{SIL,t} = H_{F,t}^{\alpha_1} \exp\left(E_{F,t}\right),$$

where $E_{F,t}$ represents the combined effect of all determinants, other than human capital, of a woman's spouse earnings.

According to this function a higher investment in daughter's human capital combined with a higher family endowment would "earn" a higher income husband (independently on whether he comes from a high socioeconomic background or he is a "self-made" man). Similarly to the income generating process on the labor market we assume that:

$$E_{F,t} = \theta e_t + \mu_t,$$

where e_t is the daughter's endowment of 'earning' capacity on the marriage market and the iid stochastic term μ_t, with zero mean and variance σ_μ^2, is the daughter's luck on the marriage market assumed to be independent on y_{t-1} and e_t.

The key parameters in these equations are α_1, the rate of return to (female) human capital on the marriage market and θ the parameter that measure the relative importance of a daughter's family endowment on the marriage market. As is common in the literature on intergenerational mobility (see Lam and Schoeni, 1993, 1994; Chadwick and Solon, 2002; and Ermisch, Francesconi and Siedler, 2006), we assume the existence of positive assortative mating in the marriage market, meaning that $\alpha_1 > 0$. We do not make any assumption about the parameter θ. If $\theta < 1$ family endowment has a greater weight in the labor market than in the marriage market and vice versa if $\theta > 1$. The higher are α_1 and θ the greater the degree of assortative matching in the marriage market.

Under these assumptions about the labor and marriage market opprtunities, the optimal parental human capital investment is obtained by substituting (3) and (3) into the parents' maximization problem. Due to the assumption of Cobb-Douglas preferences the optimal investment in children's human capital is proportional to the father's income y_{t-1}. The factor of proportionality is a function of parents' gender preferences and of the rates of return to human capital on the labor market and on the marriage market. It follows that the son's log earnings equation in equilibrium is given by:

$$\log y_t = \gamma_1 \log y_{t-1} + e_t + u_t. \tag{1}$$

In addition the model also delivers an equilibrium earnings equation for the son-in-law. This is given by:

$$\log y_{SIL,t} = \alpha_1 \log y_{t-1} + \theta e_t + \mu_t. \tag{2}$$

Equations (1) and (2) form the basis of our econometric specification. The goal of the econometric analysis is to estimate the relationship between son and father log income, and between son-in-law and father-in-law log income, and how these relationships evolve over time. Since the child's endowment, e_t, follows a first-order autoregressive process, the least squares regression of $\log y_t$ on $\log y_{t-1}$ does not yield consistent estimates of γ_1. Assuming stationarity, one can show that the probability limit of the least squares coefficient, which we will refer to as the father-son *intergenerational elasticity,* is given by:

$$\eta_{SON} \equiv p\lim \frac{\widehat{Cov(y_t, y_{t-1})}}{\widehat{Var(y_{t-1})}} = \gamma_1 + \frac{\lambda\,(1-\gamma_1^2)}{(1+\gamma_1\lambda) + (1-\gamma_1\lambda)\,(\sigma_u^2/\sigma_e^2)} \tag{3}$$

The formula shows that the probability limit of the simple OLS coefficient is equal to γ_1 plus a term that depends on λ, the degree of persistence in the endowment process, and on σ_u^2/σ_e^2, the ratio between the variance of labor market "luck" and the variance in the endowment. The intuition for these results is straightforward: the more persistent the endowment process and the larger the variance of the endowment relative to that of the idiosyncratic shock to labor market earnings, the more likely it is that any differences in earnings between sons are due to differences in their initial endowment rather than to differences in investment. Similarly, the least squares regression of $\log y_{SIL,t}$ on $\log y_{t-1}$ also gives inconsistent estimates of α_1. Given equation (2) the expression for the father/son-in-law intergenerational elasticity is given by:

$$\eta_{SIL} \equiv p\lim \frac{\widehat{Cov(y_{SIL,t}}, y_{t-1})}{\widehat{Var(y_{t-1})}} = \alpha_1 + \theta \left(\frac{\lambda\left(1-\gamma_1^2\right)}{\left(1+\gamma_1\lambda\right) + \left(1-\gamma_1\lambda\right)\left(\sigma_u^2/\sigma_e^2\right)} \right) \tag{4}$$

The formula shows how the relationship between $\log y_{SIL,t}$ and $\log y_{t-1}$ is influenced by the same determinants of the father-son intergenerational elasticity, η_{SON}. In addition, η_{SIL} is higher the higher is the rate of return to (female) human capital on the marriage market, α_1, and the relative importance of family endowment for the daughter's marriage market outcomes.

The literature on intergenerational mobility is usually interested in the estimation of equation (1) based on individual level data that is linked across generations. In an historical perspective, the work by Ferrie (1995) has spanned a literature that studies father-son occupational mobility based on longitudinal data sets constructed by linking fathers and sons across Census surveys. However, it is impossible to apply exactly the same procedure to fathers and daughters because daughters change last name upon marriage. In the next section, we discuss a methodology that circumvents this problem and allows us to estimate equation (2) over the 1850-1930 period.

4 Methodology

Consider an individual i who is young at time $t-1$ and adult at time t With slight abuse of notation, let y_{it} be individual i's *log earnings* at time t, and y_{it-1} be his father's *log earnings* at time $t-1$. With individually linked data, both y_{it} and y_{it-1} are observed, and the intergenerational elasticity estimate is obtained by regressing y_{it} on y_{it-1}. We will call this estimator the *linked* estimator, $\hat{\eta}_{LINKED}$.

Assume instead that we only observe two separate cross-sections, at times t and $t-1$, and it is impossible to link individuals across the two. This means that y_{it-1} 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. In other words, for an adult at time t named j, we replace y_{it-1} with the average log earnings of fathers of children named j at time $t-1$. We are therefore in a "generated regressor" situation, in which we use one sample to create a proxy for an unobserved regressor in a second sample. The econometric properties of this two-step estimator are well known (Murphy and Topel, 1985). As highlighted by Inoue and Solon (2010), this estimator is essentially a "two-sample two-stage least squares" (TS2SLS) estimator.[2] 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. We rely on these results to calculate appropriate standard errors for our estimator.

[2] The TS2SLS estimator is in itself a special case of a Two-Sample IV estimator (Angrist and Krueger, 1992).

Because of the particularly simple structure of our second stage, where the right hand side variable is constant for every individual with the same first name, we can further simplify the analysis by estimating a weighted least squares regression of $\tilde{y}_{jt}$ on $\tilde{y}_{j,t-1}$, where $\tilde{y}_{jt}$ is the average log earnings of adults named j at time t, and $\tilde{y}_{j,t-1}$ is the average log earnings of fathers of children named j at time $t-1$. This "means-on-means" regression is numerically equivalent to the TS2SLS estimator when there are no additional regressors and the weights are equal to the frequency counts of first names in the son's sample. We refer to this estimator as the *pseudo-panel* estimator of the intergenerational elasticity, or $\hat{\eta}_{PSEUDO}$.[3]

The key requirement for our strategy is that first names carry information about socioeconomic status. The higher the informational content of first names, the more accurate is $y_{j,\tilde{t}-1}$ as a predictor y_{it-1}. In the limit, if names are distributed randomly in the population, then the generated regressor would be just noise, and the pseudo-panel estimator would be asymptotically equal to zero. Alternatively, in the TS2SLS interpretation, names must carry economic content for there to be a valid first stage.

There is abundant empirical evidence supporting the assumption that parents choose names partly to signal their own standing in society, or their cultural and religious beliefs. For example, 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, Levitt and Fryer (2003) 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 was widespread also in the past: for example, Hacker (1999) and Haan (2005) document a relationship between first names, religiosity and fertility in Canada and the US during the 19$^{\text{th}}$ Century. Cook et al. (2012) find that distinctively black names were already common in the post-Civil War period, and were associated with lower mortality rates.

One potential concerns with our strategy is that parents choose 'aspirational' names for their children. Parents may believe that by choosing names that are associated with a higher social class they might facilitate their children social mobility. Alternatively, parents might refrain from using ethnic names to prevent discrimination. This practices would make names a more noisy indicator of parental socio-economic status. This would make our pseudo-panel estimator of the intergenerational elasticity more susceptible to attenuation bias.

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 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. In this case, $\tilde{y}'_{jt}$ is the average

[3] The "means-on-means" regression also highlights the similarity between our approach and the synthetic cohort method pioneered by Browning et al. (1985) and Attanasio and Weber (1995). In our case, the synthetic cohorts are defined on the basis of both first names and age.

income of men married to women named j in generation t, and $\tilde{y}'_{jt-1}$ is the average income of fathers of daughters named j. Then, a regression of $\tilde{y}'_{jt}$ on $\tilde{y}'_{jt-1}$ sheds light on the extent to which economic status is transmitted across generations from fathers to daughters.

5 Data

We now move to the main analysis of the paper, where we apply our methodology to data from the 1850 to 1930 US Census 1% samples from IPUMS. The availability of first names allows the creation of pseudo-panels linked by first name for both men and women.

We also take advantage of the availability of data that links individual fathers and sons to compute the *linked* estimator and to verify our methodology. The IPUMS Linked Representative Sample (Ruggles et al., 2010) links cases from the 1880 census to 1% samples of all other censuses between 1850 and 1930. Since the linking is done using information on first and last names, no linked data on married women is available, as women change their names upon marriage. Therefore, we can only analyze father-son elasticities.

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 advocated by Solon, 1992, to minimize the impact of measurement error and temporary fluctuations in income) tend to be 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.[4] 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 (7) [5]

Coding of names. The second challenge, specific to our methodology, is how to correctly

[4]A number of other papers have used this same variable to measure occupational standing, among them Abramitsky and Platt-Boustan (2012), Cverk (2012), ???

[5]For a full description of the construction of harmonized occupational codes in IPUMS and the occupational standing variables, see http://usa.ipums.org/usa/chapter4/chapter4.shtml#occscore.

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).[6] These choices may raise some issues, since there may be systematic differences in socioeconomic status between individuals with middle initials or nicknames and those without. We address the sensitivity of our estimates to these choices in Section (**??**).

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 name 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 1910, but the average number of observations per names (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 (see column 4). However, as shown in column 5, singleton names only account for 6 to 7% of all names. Furthermore, we can link more than 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 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 13% of the total variation in fathers' log earnings for boys and 13% to 15% for girls. We will discuss later the implications of this finding for our estimates of the intergenerational elasticities.

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. Therefore, this evidence confirms that names are informative about economic status.

[6]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.).

6 Benchmark 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.

The intergenerational elasticity between fathers and sons inceases by 30% between 1870 and 1930, with most of the increase occurring after the turn of the century. These results are in accord with the findings of Ferrie (2005) and Long and Ferrie (2007, forthcoming), who documented a marked decrease in intergenerational mobility in the United States between the late 19th Century and the middle of the 20th Century. The relationship in economic status between fathers in law and sons in law displays a similar trend, although it appears to be slightly higher than the father/son elasticity in the latter part of the 19th Century, and then dips below it at the end of the sample period. Interestingly, the ranking of son-in-law and son elasticities is consistent with modern estimates for the US based on the PSID (Chadwick and Solon, 2002).

The remaining top 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 and fifth rows, 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 aged 0-15 did not survive into adulthood. If child mortality differs by socioeconomic status, or if healthier children are also more likely to be employed 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 sample selection problem. The estimated coefficients for sons are somewhat lower than the benchmark, but the trends in elasticities are unaffected. The main difference is that in this case father/son-in-law elasticities tend to be larger than father/son elasticities for all the cohorts born before the turn of the 20th century. However, the gap eventually closes by the end of the sample period.

In all societies men marry later in life than women and the gender differential in age at first marriage tend 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 much larger in the 19th century (4+ years) than in 20th century (2 years). The gap peaked around 1900. 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. Omitting 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, by comparing fathers to sons that are systematically younger (or at an earlier stage of the life-cycle) we would be systematically over-estimating father/son-in-law intergenerational elasticities (even though in Tables 8 and 9 we show that our results are not particularly sensitive to the point in the lifecycle at which socioeconomic status is measured). Thus, 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. Our results are robust to both sample restrictions. The most notable difference is that the father/son-in-law elasticity is now higher than the father/son elasticity both in 1870 and in 1880. The differential, however, vanishes by 1900 and, as in the benchmark results, it reverse itself by 1930.

The difference between father/son and father/son-in-law elasticities could be driven by gender differences in the the extent to which names carry economic content. In fact, the last two columns of Table 1 showed that names are a stronger indicator of father's socioeconomic status for females than for males. This difference could affect the comparison of father-son and father/son-in-law intergenerational elasticities. However, it must be noted that while the female-male differential in the economic content of names increases over time, the daughter-son differential in earning elasticities declines. It follows that while gender differences in the economic content of names might explain part of the level differences in the estimated elasticities, they are unlikely account for the trend differences.

Finally, the last row of Table 3 presents estimates of the father/son elasticities for the two 20-year comparisons for which individually linked data are available. Not surprisingly, our pseudo-panel estimator of the intergenerational elasticity is lower than the individually-linked estimator, by about 28-33%. This is because of the attenuation bias induced by measurement error in father's and son's earnings when we take averages by first names. Thus our estimate should be interpreted as a lower bound to the actual intergenerational elasticities.

7 Robustness

In this section we explore the sensitivity of our results to the measurement assumptions used in the benchmark specification.

Measuring Income

We start by investigating whether our results are driven by our choice of the 1950 occupational scores. One issue is that the 1950 income distribution is relatively compressed. Moreover, one may be concerned that the 1950 occupational classification does not reflect accurately the distribution and the relative standing of occupations that were common during the late 19th Century and early 20th Century. This issue is particularly important from our standpoint as 'farmers' represent a large part of our sample [7] and farming occupations and farm ownership were associated with higher socioeconomic status during our sample period than in 1950.

We explore this issues in Table 4. The first row of each panel reproduces the benchmark estimates from Table 3.

We then obtain estimates of the intergenerational elasticity using the 1900 occupational-earnings distribution in Appendix Table 2. This income distribution is based on the tabulations by Preston and Haines (1991), who collect data from a number of different sources to impute the average income by occupation at around the turn of the 20th Century. The main advantage of using the 1900 occupational income distribution is obviously that the list of occupational categories matches more closely the list, types and ranking of occupations that were common during much of the sample period. The main limitation is that, 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.[8] The 1900 estimates are obtained by imputing owner occupier farmer income as the average income for all farming occupations present in the Preston-Haines classification and coded as "farmer" in the 1950 coding scheme.[9] The estimates based on the 1900 occupational income distribution are reported in the second row of each panel in Table 7. We obtain very similar results to the benchmark estimates, in terms of the levels of the elasticities, how they change over time, and how they differ between sons and sons-in-law.

As an alternative, we also use the 1900 income distribution, but attempt to impute an income for farmers based on the methodology originally described by Mitchell et al. (1922) and recently used by Abramitzky et al. (2010). For owner-occupier farmers, we calculated income as the difference between the value of farm products (augmented by the value of rent and food consumed by the family) and the total expenditures on labor, fertilizer, feed, seeds, threshing, taxes and maintenance. For farm tenants, we imputed an income of \$334, which is the income for specialized farm workers (stock raisers, fruit growers, etc.) in the Preston-Haines tabulations. The intergenerational elasticity estimates based on this imputation are presented in the third row

[7] 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, see Appendix Table 1.

[8] Preston-Haines imputations 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, and as such, the survey includes only information for the urban population. Moreover, the 1901 survey collected data for the "typical" urban family, meaning that by construction the resulting income distribution would be more compressed than what one would obtain in a representative sample.

[9] Specifically, we record all the occupations in the 1910 Census that were coded as farmers in the 1950 occupational classification. We then calculate the average income (weighted by the sample frequencies in 1910) for the occupations with nonmissing income data based on the Preston-Haines tabulations, and assign this value to all farmers (see Appendix Table 1 for the details of the calculation).

of each panel in Table 4. The father/son intergenerational elasticity is not very sensitive to the imputation of farmers' income. However, the father/son-in- law intergenerational elasticity does seem to depend on the imputation method for farmer income. The estimated elasticity is 5 to 9 points lower in the first part of the sample period (1850 to 1880), but in the latter part of the sample the two estimates converge.

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 the estimates of intergenerational mobility. The trends are similar to the benchmark but now the son-in-law elasticity becomes higher than the son elasticity in the last two periods.[10]

In Table 5 we assess the robustness of our results to alternative measures of occupational income. In the second row of each panel we assign to each occupation its percentile rank in the 1950 income distribution and regress the percentile rank of the average income of the son (son-in-law) occupation on that of the father (father-in-law). The rationale for using the rank that the measure does not depend on the, potentially noisy, imputed level of occupational income. We find a more attenuated trend for the father-son elasticity while the father/son-in-law elasticity mirrors the baseline fairly closely. In the next row we re-estimate our 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 especially in the first part of the sample period. This is probably because attempting to match 19th century occupations to those of the late 20th century introduces a large(r) amount of noise, which attenuates the results. The remaining rows in the table report the estimates obtained for two additional labor market outcomes, also based on the recoded 1950 occupational categories. ERSCOR50 assigns the percentile rank of each occupation's median income based on contemporaneous earnings data. The Duncan socioeconomic index (SEI) is a well-known measure of occupational prestige which combines occupational education and occupational income. The results are qualitatively similar to those of the benchmark.

[10]In addition, we have also calculated the father/son elasticities based on the linked sample, using both the 1900 and 1950 occupational distributions, with and without farmers (see the last row of the first panel). For the two pair of years in which we can calculate 20-year correlations (1860-1880 and 1880-1900), we find that the intergenerational correlation increased over time.

Age

We also investigate whether our results may be driven by changes in the age-occupational income profile over time.

A consistent pattern that emerges from the modern literature is that the estimated elasticities tend to be lower when son's earnings are measured early in their careers. In our context, this issue may be somewhat less of a concern. Because investment in formal schooling was much lower than what it is today, it is reasonable to expect that the age-income profile peaks at an earlier age during our sample period, so that sons' occupational income during their twenties would be more reflective of long-run status than it is today.[11] To strengthen this conclusion, we also reestimate our model with controls for a quadratic function in father's and son/son-in-law's age.[12]

The results are presented in Table 6. For each year, we present side by side the baseline estimate (without age controls), and the estimate with age controls. The estimated elasticities are almost completely unaffected by the controls for age, for both sons and sons-in-law. Interestingly, the controls for son's age enter the regressions with the expected signs and are always highly significant. The coefficients indicate that the age-occupational earnings profile reaches a peak at around 30 years of age in most specifications. A similar pattern is found for the age-occupational earnings profile of sons-in-law, even though the estimates are not always as precise and are smaller than for sons. On the other hand, father's age and age squared exhibit a mixed pattern of signs and typically come in not significant. The fact that our basic estimates are not sensitive to the inclusion of age control suggests that the age at which occupational income is measured does not matter much for the estimated elasticities. This result is confirmed when we estimate intergenerational elasticities at 30 year intervals (see Appendix Table 4).

Appendix A and Appendix Table 5 further investigate the robustness of our results to: a) Separating names with and without middle middle initials; b) grouping the main root of a name with its most common nicknames; c) using the Soundex phonetic algorithm to deal with potential misspelling of names and d) adjusting for gender differences in the distribution of names. All this different name coding schemes yield the same pattern of results as in the baseline.

[11] In fact, Sutch (2011) collects data on wages from a number of industries and states in the 1890s, and documents that the age earnings profile peaks as early as 25, and stays relatively flat thereafter.

[12] In practice, this means adding controls for average age by name and its square for both fathers and sons/sons in law, in our benchmark regressions.

8 What factors can explain the trends?

8.1 Name distribution

The previous section established that the father-son intergenerational elasticity increased markedly from 1870 to 1930. Does this increase reflect real changes in the underlying income generating process or could this be an artifact of changing features of the name distribution? For example, we have seen in section 4 that the pseudo-elasticity depends critically on the extent to which names carry information about socio-economic status. Thus, if names become more socially stratified this could translate into a higher elasticity even if the underlying transmission process is unchanged.

To answer this question we conduct a series of numerical simulations. The goal of this analysis is to evaluate the sensitivity of the estimator to different assumptions about the distribution of names in the population, and how they are correlated with socioeconomic status. Our strategy is to generate simulated data based on the income generating process described in section (3) and a specific assumption about the name assignment process. We then find the values of the parameters that minimize the distance between a set of simulated moments and their empirical counterparts in 1860 and 1880. The moments that we attempt to match are the following: the intergenerational elasticity of income and the variance of income based on the individually-linked data ($Cov(y_t, y_{t-1})/V(y_{t-1})$ and $V(y_{t-1})$); the pseudo-panel analogs of these two moments ($Cov(\tilde{y}_{jt}, \tilde{y}_{jt-1})/V(\tilde{y}_{jt-1})$ and $V(\tilde{y}_{jt-1})$); the fraction of the population having one of the 50 most common names, as a measure of the concentration of names; and the R-squared from a regression of father's income on a full set of children name dummies, which we use as our measure of the economic content of names.

The simulated data. We generate a population of N families. For each family the income generating process is given by:

$$\begin{aligned} y_t &= \gamma_1 y_{t-1} + e_t + u_t \\ e_t &= \lambda e_{t-1} + v_t, \end{aligned}$$

with u_t and v_t iid normal with variances σ_u^2 and σ_v^2 respectively.

For the name assignment process we assume that parents of generation $t-1$ choose their children's first name as a function of the family's earning endowment e_{t-1}. The dependence of the naming process on e_{t-1} rather than on actual earnings y_{t-1} reflects the fact that name choices are more likely to be affected by the more permanent component of earnings, whereas y_{t-1} can be affected by transitory shocks. The probability of choosing name j out of a finite set $\{1, 2, ..., J\}$ is given by:

$$P(j|e_{t-1}) = \frac{\exp(\delta_{CON,j} + \delta_{SES,j} e_{t-1})}{\sum_{j'=1}^{J} \exp(\delta_{CON,j'} + \delta_{SES,j'} e_{t-1})} \tag{5}$$

We assume that $\delta_{CON,j}$ and $\delta_{SES,j}$ are normally distributed with mean zero and variance σ_{CON}^2

and σ^2_{SES}, respectively. Furthermore, they are independent of each other as well as of all other variables in the model. The parameter σ^2_{CON} determines the concentration of names in the population: the higher σ^2_{CON} the more likely it is that some names will appear frequently while others are very rare. The parameter σ^2_{SES} instead determines the sensitivity of names to socio-economic status. In the extreme case of $\sigma^2_{SES} = 0$ names are assigned randomly. The larger σ^2_{SES}, the more indicative are names of a family's social standing.

Given the income and name-generating processes, we create a sample of individually linked fathers and sons, and a pseudo-panel of fathers and sons linked by the son's first name. We replicate this process R times, and compute the simulated moments as the average value of the moments across replications. The benchmark vector of unknown parameters $\theta = (\gamma_1, \lambda, \sigma^2_u, \sigma^2_v, \sigma^2_{CON}, \sigma^2_{SES})$ is obtained by minimizing the distance between the simulated moments and their data counterparts. For simplicity, we use equal weighted minimum distance, and we set $R = 15$. Further details about the implementation of this method are given in Appendix B.

Estimation results. The resulting estimates are presented in Table 7. Since the model is just identified, we are able to exactly match all the moments. Interestingly, we find that the returns to human capital investment (γ_1) play a larger role in the transmission of economic status than the autoregressive component of the endowment, λ. The variance of shocks to labor income is about three times as large as that of the endowment. These parameters imply that a large fraction of the overall variance in income is due to labor market "luck." Finally, the estimated values of σ^2_{CON} and σ^2_{SES} indicate that the distribution of first names is fairly concentrated, and names do carry economic content.

Sensitivity to name distribution. We now fix $(\gamma_1, \lambda, \sigma^2_u, \sigma^2_v)$ at the values reported in the bottom panel of Table 7, and show how the pseudo-panel estimator and other moments vary over a grid of values for σ^2_{CON} and σ^2_{SES}. The results are reported in Table 8. The first entry in each cell represents the estimated pseudo-elasticity η in our simulated samples. The second and third entries in the cells represent, respectively, the estimated top-50 share and the R^2 from a regression of log father's income on a full set of name dummies. The cell corresponding to the SMM estimates is highlighted.

Going down the columns, we note that the estimator is generally not very sensitive to the parameter determining the concentration of the name distribution σ^2_{CON}. The estimator tends to increase as the distribution of names becomes more concentrated, but this increase is quite modest given the range of variation in σ^2_{CON}, especially if compared to the range of variation of the other moments. For example, the top-50 share increases from 0.34 to about 0.84, while the R^2 falls by about 40-60% as σ^2_{CON} rises from 2.5 to 15.

On the other hand, the estimated pseudo-elasticity is strongly affected by the parameter σ^2_{SES}. In particular, the estimated value of η is very close to zero when $\sigma^2_{SES} = 0$, i.e. names carry no information about a family's socio-economic status, as discussed in section 4. The table shows that σ^2_{SES} should increase by an order of magnitude (from 5.9 to about 30) in order to generate the increase in the intergenerational elasticity coefficient that is observed in the data (from 0.313 in 1880 to 0.48-0.50 in 1920-30 – see Table 3). This large increase in σ^2_{SES} would be associated with an increase in the R-squared from 0.105 to 0.18. This is much larger than the observed change in the R-squared over the period (from 0.1108 to 0.1256 – see Table 1, column 8).

Table 9 explores whether variation in the main parameters governing the income process can rationalize the observed increase in η. The structure of the tables is analogous to Table 8, with the SMM estimates highlighted. Both γ_1 and λ have fairly large effects on the estimated father-son elasticity. For example, we can generate the observed increase in η with an increase in λ for 0.19 to 0.4 and an increase in γ_1 from 0.42 to 0.5. Despite these changes in γ_1 and λ, the R-squared and the top-50 concentration parameter hardly move at all.

We conclude that the increase in father-son elasticity cannot be explained by changes in the degree to which names carry economic content. Instead reasonable changes in the parameters of the income process are likely to be responsible for the observed increase.

8.2 Other Factors

[Preliminary].

In this section we explore economic and demographic trend that can potentially explain the observe changes in intergenerational elasticity.

Changes in fertility The total fertility rate dropped from 5.5 in 1850 to about 2.5 in 1930. *exact numbers? source?* 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 a 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. It could be possible to observe an increase in observed elasticity if the fertility decline occurs earlier for the high income group than for the low income group.[13] Jones and Tertilt (2008) document that the fertility-income gradient was negative already for the generation of women born in 1828 and that the fertility transition that occurred between this cohort and the cohort of women born in the late 1890s did not occur evenly across socio-economic groups. The total fertility rate for lower-income women hovered around 6 between the 1828 and the 1853 cohorts. It then dropped sharply reaching 3.3 children by the 1898 cohort. In contrast, the fertility decline for high socio-economic status

[13]For example, suppose that both high and low income families have 6 children. The high income family can invest a total of 60 in their children's human capital, the low income family can invest 6. Therefore, each high income child starts with an income of 10 and every low income child starts with 1. In the next generation the fertility of the high income family declines to 3, whereas that of the low income family is unchanged. Now each of the high income children with start off with a higher income and is therefore more like to maintain his income status, intergenerational mobility drops.

women (from around 5 to around 2.5) was smoother. This pattern of change in fertility would imply an increase in intergenerational elasticity between the 1860 and 1900 cohorts. In fact, the jump in intergenerational intergenerational coefficient occurs for cohort born at the beginning of the 20th century. Thus changes in fertility do not seem to be able to explain the observed trends in intergenerational elasticity.

We can further assess this point by directly controlling for fertility in our baseline regression. The results are reported in Table 10. In the second row of the table, we control for the average number of siblings for children with a given first name. An alternative way of accounting for differences in fertility over time and potential asymmetries in the allocation of family resources across children, would be to look separately at children with a given birth order. This is not possible with our methodology, as we do not observe birth order in the adult sample, but we can still control for the distribution of birth orders by first name (third row of Table 10). Both rows exhibit only minimal differences with the baseline results, with the possible exception of the first two cohorts. if anything, accounting for fertility makes the trend in intergenerational mobility even more pronounced.

Farmers [To be completed]

Migration The sample period that we analyze was characterized by dramatic migratory flows, both from outside of the US and internally. The national *ethos* of the "American dream" is very much based on the belief that migration can serve 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 inconsisent with the evolution over time in the intergenerational elasticity estimate. Immigration to the US had an early peak in the 1880s and then a second, larger peak between 1900 and 1915 [14]; if immigration plays a major role in driving the overall level of mobility, and, in particular, the 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 1920 and 1930 cohorts.

Of course, it is also possible that immigration contributed to attenuate what would have otherwise been an even larger decrease in intergenerational mobility. To assess this possibility, we atempt to assess directly the effect of immigration on our basic estimates. These could be downward biased if immigrant fathers tend to be employed in low-paying occupations, but their children quickly rise through the social ranks. It is easy to address these issues by simply controlling for

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

the immigrant status of both fathers and sons.[15,16]

The results are presented in the second row of Panels A (sons) and B (sons-in-law) in Table 11. 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 two cohorts. [17]. Overall, 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 upward trend in intergenerational elasticity is unlikely to be driven by changes in immigration over the sample period. [18]

Internal mobility Long and Ferrie (forthcoming) argue that residential mobility is a prime candidate to explain the high level of intergenerational mobility in the US in the 19th 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 19th 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 1930.[19] Therefore, the trends in mobility are broadly consistent with the trends in intergenerational elasticity: 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, elasticity estimates that do not account for internal mobility would be biased downwards. To further investigate this hypothesis, we add to our basic specification a control for internal migrant status.[20] The results are presented in the third row

[15]In practice, for every first name we control for the percentage of immigrants with that name among both fathers and sons; for the son-in-law specifications, we control for the percentage of immigrants among the father in laws, daughters, and husbands)

[16]Alternatively, one could restrict the whole analysis to exclude all immigrants, or even all children with immigrant fathers. The results, available upon request, are almost identical.

[17]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. This induces an upward bias in the estimates of the intergenerational correlations when one does not control for immigrant status

[18]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.

[19]Source: our own calculations from the IPUMS samples

[20]In practice, for every first name we control for the percentage of internal migrants – defined as individuals

of Table 11. Contrary to our conjecture, the inclusion of these controls has essentially no effect on the intergenerational elasticity estimates. If anything, the estimates in the first part of the sample period seem to be slightly *upward* biased. We can conclude that inter-state mobility does not appear to explain much of the trend in the intergenerational elasticity estimates. Because of lack of data, however, we cannot rule out that mobility across geographic areas within the same state may have played a more significant role.

Regional Differences. [To be completed]

Investment in public education [To be completed]

9 Discussion

What factors might contribute to explain the trends in η_{SON} and η_{SIL} between 1850 and 1930?

The most plausible explanation for the increase in η is the improvement in men's labor market outcomes during this period. Margo (2000) documents a long-term rise in the returns to educated labor beginning before the Civil War and continuing until the turn of the 20th century.[21] In addition, Cverk (2010) shows that men's career prospects improved substantially between 1880 and 1930. These developments in the labor market can be interpreted as an increase in γ_1 in our model, and thus an increase in η_{SON}.

Moving to the marriage market, given the assumption of positive assortative mating in human capital ($\alpha_1 > 0$), increases in γ_1 should also lead to an increase in the return to *female* human capital on the marriage market, everything else being equal.[22] Thus the improvement in men's labor market outcomes would be consistent with our finding that the father-son and the father-daughter elasticity share a common trend over the period of interest.

Other factors might have had an impact on women's incentives to marry, thus affecting η_{SIL}. As discussed in Haines (1996) the marriage rates declined up to 1900 and then reverse direction. At the same time women's age at first marriage increased substantially - from 21 in 1850 to around 23 in1900 (Haines, 1996, Figure 1 and 2). One argument is that, increasing men's opportunity on the labor market contributed to make marriage more appealing to women who, at the same time, were

living in a different state than their state of birth – with that name among both fathers and sons; for the son-in-law specifications, we control for the percentage of internal migrants among the father in laws, daughters, and husbands).

[21]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)). See Margo and Villaflor (1987) for an in-depth analysis of wage growth between 1820 and 1865.

[22]Note that Equation (2) can be interpreted as the reduced form of the equilibrium matching function in a model with non-transferable utility, sorting and stochastic pre-marital investment in human capital (see Bhaskar and Hopkins, 2011).

becoming increasingly selective in their partner choice due to their own increasing opportunities in the labor market (Cverk, 2010). Another possibility is the large imbalance in the sex ratio induced by immigration. Haines(1996) shows that immigration to the US peaked in the opening decades of the 20th Century and was heavily skewed towards white males.[23] . Immigration could affect the observed elasticity because of the increasing option value of marriage due the expansion in the size and heterogeneity of the pool of marriageable men.

10 Conclusion

In this paper we have provided a new perspective on intergenerational mobility in the United States in the late 19th and early 20th centuries. We developed a new methodology that links cohorts across Census years on the basis of first names, and allows us to calculate intergenerational elasticities for both sons and sons in law. We find that the father/son elasticity increased markedly between 1850 and 1930, consistent with previous studies. The father/son-in-law elasticity was higher than the father/son elasticity at the beginning of the period, but has remained fairly stable over time. Consequently, for cohorts born after the turn of the century, the two measures had converged. Our findings indicate that estimates of intergenerational mobility for the 19^{th} Century that ignored the link between fathers and daughters may have understated the true extent of persistence in socioeconomic status across generations.

[23] Cverk (2010) shows that this had an impact on women's marriage opportunities.

References

[1] Abramitzky, Ran, Platt Boustan, Leah and Katherine Eriksson (2010). "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration," NBER Working Paper 15684, January.

[2] Attanasio, Orazio P. and Guglielmo Weber (1995). "Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey." *Journal of Political Economy*, 103(6), pp. 1121-1157.

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

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

[5] Becker, Gary S. (1991). *A Treatise of the Family*, Cambridge, Mass.: Cambridge University Press.

[6] Bergstrom, Ted and Mark Bagnoli (1993). "Courtship as a Waiting Game," *Journal of Political Economy*, 101, 185-202.

[7] Bergstrom, Theodore C. (1997). "A Survey of Theories of the Family." *Handbook of Population and Family Economics*, Chapter 2.

[8] 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.

[9] Bhaskar, V and Ed Hopkins (2011) "Marriage as Rat Race: Noisy Pre-Marital Investments with Assortative Matching',' manuscript, UCL.

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

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

[12] Bisin, Alberto and Tierry Verdier (2000). "Beyond the Melting Pot: Cultural Transmission, Marriage,and the Evolution of Ethnic and Religious Traits." *Quarterly Journal of Economics*, CXV, pp. 955–987.

[13] Black, Sandra E. and Paul J. Devereux (2010). "Recent Developments in Intergenerational Mobility." NBER Working Paper 15889.

[14] Chadwick, Laura and Gary Solon (2002). "Intergenerational income mobility among daughters," *American Economic Review*, vol. 92(1), pp. 335-44.

[15] Chamberlain, Gary. "Panel Data," in *Handbook of Econometrics* Volume II (Z. Griliches and Michael D. Intriligator, eds.), pp. 1248-1318, Amsterdam: North-Holland.

[16] Charles, Kerwin K., Hurst, Eric and Alexandra Killewald (2011). "Marital Sorting and Parental Wealth" NBER Working Paper 16748.

[17] Clark, Gregory (2010) "Was there ever a Ruling Class? Surnames and Social Mobility in England, 1200-2009." Manuscript, UC Davis.

[18] Cvrcek, Tomas (2010) "America's settling down: How Better Jobs and Falling Immigration led to a Rise in Marriage, 1880-1930," NBER Working Paper No. 16161, July.

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

[20] Epstein, E. and Guttman, R. (1984). "Mate selection in man: Evidence, theory, and outcome," Social Biology, vol. 31(3-4), pp. 243-78.

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

[22] Ermisch, John, Francesconi, Marco and Thomas Siedler (2006). "Intergenerational Mobility and Marital Sorting," *Economic Journal*, vol. 116(513), pages 659-679.

[23] Fernández, Raquel, N. Guner, and J. Knowles (2005). "Love and Money: A Theoretical and Empirical Analysis of Household Sorting and Inequality." *Quarterly Journal of Economics*, 120 (1), pp. 273-344

[24] Fernández, Raquel and R. Rogerson (2001). "Sorting and Long-Run Inequality" *Quarterly Journal of Economics*, 116, pp. 1305–1341.

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

[26] 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.

[27] 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.

[28] Galor, Oded and Omer Moav (2002). "Natural Selection and the Origin of Economic Growth" *QuarterlyJournal of Economics*, 117 (4), pp. 1133-1191.

[29] Goldin, Claudia (1990). *Understanding the Gender Wage Gap: An Economic History of American Women.* Oxford University Press.

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

[31] Goldin, Claudia, and Maria Shim (2004). "Making a Name: Women's Surnames at Marriage and Beyond," *Journal of Economic Perspectives*, 18(2): 143–160.

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

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

[34] 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*, Vol. 29, No. 3 (Fall). pp. 373-411.

[35] 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* 43: 339-65.

[36] 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): 15–39.

[37] Haines, Michael (2008) "Fertility and Mortality in the United States." EH.Net Encyclopedia, edited by Robert Whaples. March 19. URL http://eh.net/encyclopedia/article/haines.demography

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

[39] 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): 333-372.

[40] Kalmijn, M. (1994). "Assortative mating by cultural and economic occupational status," *American Journal of Sociology*, vol. 100(2), pp. 422-52.

[41] Kremer, Michael (1997). "How Much Does Sorting Increase Inequality?" *Quarterly Journal of Economics*, vol. 112(1), pp 115-39.

[42] Lam, David (1988). "Marriage markets and assortative mating with household public goods: theoretical results and empirical implications," *Journal of Human Resources*, 23 (4), pp. 1235-58.

[43] Lam, David A. and Robert F. Schoeni (1993). "Effects of family background on earnings and returns to schooling: evidence from Brazil," *Journal of Political Economy*, vol. 101(4), pp. 710-40.

[44] Lam, David A. and Robert F. Schoeni (1994). "Family ties and labor market in the United States and Brazil," *Journal of Human Resources*, vol. 29(4), pp. 1235-58.

[45] Lieberson, Stanley (2000). *A Matter of Taste: How Names, Fashions, and Culture Change.* New Haven: CT Yale University Press.

[46] 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.

[47] Long, Jason and Joseph P. Ferrie (forthcoming). "Intergenerational Occupational Mobility in Britain and the U.S. Since 1850." *American Economic Review.*

[48] Mare, Robert D. (1991). "Five decades of educational assortative mating," *American Sociological Review*, vol. 56(1), pp. 15-32.

[49] Robert A. Margo and Georgia C. Villaflor (1987). "The Growth of Wages in Antebellum America: New Evidence," *The Journal of Economic History* , Vol. 47 (4): 873-895.

[50] Robert A. Margo (2000). "The History of Wage Inequality in America, 1920 to 1970," *Macroeconomics* 0004035, EconWPA.

[51] Mitchell, Wesley, Willford I. King, Frederick R. Macaulay and Oswald W. Knauth (1922). *Income in the United States, Its Amount and Distribution, 1909-1919.* (New York: National Bureau of Economic Research).

[52] Mulligan, Casey B. (1999). "Galton vs. the human capital approach to inheritance," *Journal of Political Economy*, vol. 107(6), Part 2, pp. S184-S224.

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

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

[55] Raaum, Oddbjørn, Bratsberg, Bernt, Røed, Knut; Österbacka, Eva; Eriksson, Tor; Jäntti, Markus and Naylor, Robin A. (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.

[56] Steven Ruggles, 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.

[57] Sacerdote, Bruce (2005). "Slavery and the Intergenerational Transmission of Human Capital." *Review of Economics and Statistics*, 87(2), pp. 217-234.

[58] Schwartz, Christine R., and Robert D. Mare (2005). "Trends in Educational Assortative Marriage from 1940 to 2003." *Demography* 42: 621-46.

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

[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] United Nations (1990). "Patterns of First Marriage: Timing and Prevalence." United Nations publication, ST/ESA/SER.R/111.

[62] van Leeuwen, Marco H.D., Ineke Maas and Andrew Miles (2002). *HISCO: Historical International Standard Classification of Occupations*. Leuven: Leuven University Press.

[63] van Leeuwen, Marco H.D., and Ineke Maas (2005). "A short note on HISCLASS." Working Paper, available at http://historyofwork.iisg.nl/docs/hisclass-brief.doc.

[64] Van Poppel, Frans and Daniel Scott Smith (1999). "Introduction," *History of the Family*, 4(3), pp. 229-237.

[65] Weiss, Yoram (1997). "The Formation and Dissolution of Families: Why Marry? Who Marries Whom? And What Happens upon Divorce?" Chapter 3 in *Handbook of Population and Family Economics*, edited by M. R. Rosenzweig and O. Stark. New York: Elsevier Science.

[66] World Bank (2000). *Social Indicators of Development*. Published by World Bank. The Johns Hopkins University Press. Baltimore and London.

[67] Xie, Yu and Alexandra Killewald (2011). "Has Social Mobility in America Declined? Data, Methods, and Farming." Manuscript, University of Michigan.

Appendices

A Name coding

In this Appendix we assess the robustness of our results to different name coding schemes. In the benchmark specification, for the purpose off aggregating results by name, we treated each individual occurrence of a first name as a separate name, ignoring middle names or middle initials and common abbreviations. In the first part of Appendix Table 5 we relax this restriction.

The first row of each panel in the table reproduces the benchmark estimates from Table 3. The second row presents the intergenerational elasticities when we take into account middle initials as well as the first name (so, for example, "William,"."William J." and "William H." are treated as three separate names). The third row presents results when we group together the main root of a name with its most common nicknames (so, for example, "William," "Bill," "Billy" and "Willie" are all treated as separate instances of the same name). In both of these robustness tests, we obtain results that are broadly consistent with our benchmark estimates.

In the next row, we deal with the potential misspelling of names by using the Soundex algorithm.[24] Coding names this way results in a substantial reduction in the number of names, and an increase in the average number of occurrences per name. This can have two opposite effects on the estimated elasticity. On one hand, it reduces the occurrence of uncommon names and therefore, by the law of large numbers, the average income is a more accurate measure of actual father's income. This tend to reduce the attenuation bias and raise the estimate of the elasticity. One the other hand, the Soundex algorithm groups together names that may belong to very different socioeconomic groups (for example, Michael and Miguel) which may exacerbate measurement error. In practice, the father-son elasticities increase by 6 to 13 percentage points but still exhibits and increasing trend over the sample period. The father-son-in-law elasticities increase by as much as 25 percentage points with the largest increase at the beginning of the sample period. As a result, there is no evidence of an increase in father-son-elasticity.

Finally, as shown in Table 1, the distribution of names for male children is somewhat more concentrated than the distribution for female children, and this could potentially impact the estimates of intergenerational elasticity. We therefore reestimate our model using the frequency distribution of female names as weights for the father-son elasticities, and the distribution of male names as the weights for the father/son-in-law elasticities. The resulting estimates are almost unaffected by this modification, implying that the observed difference between the two elasticities is not due to differences in the name distribution between the male and female samples.

[24] The Soundex is a phonetic algorithm that indexes names by sound, and is specifically designed to assign the same numeric code to similar sounding names (NARA 2007).

B Numerical Simulations

This section describes the implementation of the SMM estimator.

We generate a population of N families. We generate incomes for each family for T periods based on the income transmission process:

$$\begin{aligned} y_t &= \gamma_1 y_{t-1} + e_t + u_t \\ e_t &= \lambda e_{t-1} + v_t, \end{aligned}$$

with u_t and v_t iid normal with variances σ_u^2 and σ_v^2 respectively. We draw the initial values of income and family endowment for each family based on the long run distributions of y_t and e_t. We keep only the last two generations, in order for the observed distribution of income not to be affected by the initial conditions.[25]

The names of generation t children are assigned on the basis of a probabilistic process. The probability of choosing name j out of a finite set $\{1, 2, ..., J\}$ is given by:

$$P(j|e_{t-1}) = \frac{\exp(\delta_{CON,j} + \delta_{SES,j} e_{t-1})}{\sum_{j'=1}^{J} \exp(\delta_{CON,j'} + \delta_{SES,j'} e_{t-1})} \tag{6}$$

We assume that $\delta_{CON,j}$ and $\delta_{SES,j}$ are normally distributed with mean zero and variance σ_{CON}^2 and σ_{SES}^2, respectively.

We then extract two samples: The individually linked sample is a 10% extract from this population; the pseudo-panels are obtained by taking two independent 10% extracts, one from the father's generation and one from the son's generation.

We use these two samples to calculate the six simulated moments that will be matched to the data. Starting from an initial guess for the parameter vector $\theta = (\gamma_1, \lambda, \sigma_u^2, \sigma_v^2, \sigma_{CON}^2, \sigma_{SES}^2)$ we replicate this process $R = 15$ times, and compute the simulated moments as the average value of the moments across replications. We then iterate this process until convergence.

The original population size N is set to $500,000$. This value is chosen to approximately match the number of white males aged 0-15 in the 1860 1% Census sample.[26] The number of distinct names in the census is 4350. However, a careful examination of the data reveals that many of the distinct names are typos (e.g. "???") or slight spelling variations of the same root name ("Michaal" or "Micheal"). More than 3000 names appear only once in the data, and only 800 appear three

[25] In practice, preliminary simulations showed that $T = 14$ was sufficient to guarantee that the initial conditions had no effect on the dsitributions.

[26] Ideally, we would have wanted to generate a population equivalent in size to the population of the United States in 1860, and then draw a 1% sample to make our simulated data exactly analogous to the Census data. Because of computational limitations, we instead generated a smaller population and drew a 10% sample. The results are not sensitive to small modifications in the percentage drawn from the original population.

times or more. Our solution is to start with a pool of $J = 1500$ distinct root names, and then to artificially misspell each name with probability p. To calibrate the misspelling probability, we first group names by their Soundex code, and then calculate the fraction of names within each Soundex code that are not equal to the most common spelling of the name. Averaging across all Soundex codes, we obtain $p = 0.09$.

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

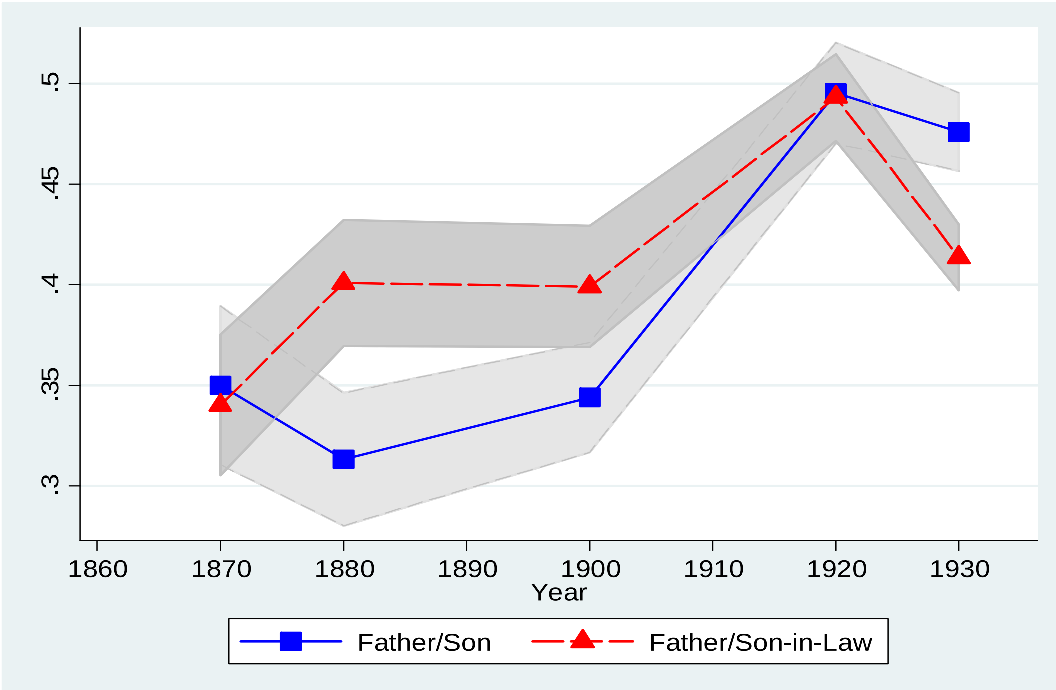

Table 1. Summary Statistics for Children's Names: 1850-1910

	(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	0.6919	0.1343
1860	48,114	4,083	11.8	70.5	6.0	93.7	0.6946	0.1108
1870	58,039	4,582	12.7	69.4	5.5		0.6978	0.1053
1880	75,004	6,589	11.4	69.4	6.1	92.9	0.6529	0.1119
1900	103,817	9,696	10.7	71.0	6.6	92.8	0.5638	0.1265
1910	117,612	9,818	12.0	69.5	5.8	94.1	0.5342	0.1256
				Females				
1850	34,272	3,442	10.0	71.9	7.2	92.4	0.6984	0.1357
1860	46,874	4,488	10.4	70.7	6.8	92.8	0.6573	0.1320
1870	55,739	5,206	10.7	71.1	6.6		0.6193	0.1356
1880	72,160	7,161	10.1	69.0	6.8	92.0	0.5475	0.1331
1900	101,516	10,081	10.1	70.9	7.0	92.3	0.4744	0.1526
1910	114,074	10,103	11.3	69.3	6.1	93.5	0.4726	0.1545

Table 2: Common Names Given to Children, Ranked by Mean Father's Occupational Income 1850-1930.

	1850	1860	1870	1880	1900	1910	1920	1930
Males								
Rank:	*Most Prestigious*							
1	Edward	Walter	Harry	Paul	Donald	Abraham	Jerome	Irving
2	Frederick	Frank	Walter	Harry	Kenneth	Max	Irving	Frederick
3	Edwin	Willie	Herbert	Frederick	Harold	Nathan	Jack	Richard
4	Charles	Louis	Theodore	Ralph	Morris	Vincent	Nathan	Roger
5	Franklin	Fred	Edward	Philip	Max	Edmund	Abraham	Robert
	Least Prestigious							
1	Jesse	Levi	Jesse	Luther	Luther	Jessie	Willie	Jose
2	Hiram	Isaac	Franklin	Ira	Dewey	Otis	Loyd	Loyd
3	Isaac	Benjamin	Isaac	Isaac	Perry	Luther	Luther	Willie
4	Daniel	Andrew	Hiram	Willis	Virgil	Eddie	Jessie	Ervin
5	David	Jacob	Martin	Charley	Ira	Charley	Otis	Archie
Females								
Rank:	*Most Prestigious*							
1	Emma	Ada	Bertha	Bessie	Dorothy	Eleanor	Betty	Jeanne
2	Alice	Kate	Jessie	Mabel	Marion	Marian	Jean	Jane
3	Anna	Lizzie	Grace	Helen	Helen	Dorothy	Jane	Carolyn
4	Isabella	Clara	Carrie	Ethel	Louise	Marion	Kathryn	Ann
5	Josephine	Fanny	Helen	Blanche	Marie	Virginia	Muriel	Joan
	Least Prestigious							
1	Sally	Amanda	Nancy	Nancy	Nancy	Sallie	Lela	Eula
2	Nancy	Nancy	Lucinda	Viola	Ollie	Addie	Maggie	Lorene
3	Lucinda	Rachel	Rebecca	Martha	Nannie	Ollie	Ollie	Dortha
4	Martha	Lucinda	Amanda	Rachel	Sallie	Mattie	Effie	Willie
5	Lydia	Martha	Martha	Amanda	Alta	Iva	Eula	Opal

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

Table 3. Intergenerational Elasticities in Occupational Income, 1850-1930.

	(1)	(2)	(3)	(4)	(5)
	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930
Sample:					
Sons: baseline	0.3500	0.3133	0.3440	0.4953	0.4760
	(0.0239)	(0.0200)	(0.0166)	(0.0152)	(0.0118)
	[37077, 1182]	[50847, 1478]	[80255, 2234]	[109079, 3253]	[122468, 3720]
Son's Age 5-15	0.3286	0.3050	0.3574	0.4527	0.4199
	(0.0293)	(0.0243)	(0.0203)	(0.0173)	(0.0134)
	[24336, 984]	[32657, 1257]	[53629, 1860]	[76365, 2782]	[83920, 3257]
Married Sons	0.2868	0.3433	0.3805	0.4715	0.4428
	(0.0312)	(0.0260)	(0.0223)	(0.0178)	(0.0133)
	[17912, 891]	[24510, 1155]	[36521, 1641]	[57570, 2586]	[67137, 3051]
Sons in law: baseline	0.3402	0.4009	0.3992	0.4932	0.4136
	(0.0213)	(0.0191)	(0.0183)	(0.0131)	(0.0100)
	[23280, 976]	[30081, 1376]	[45804, 2063]	[68439, 2888]	[79314, 3326]
Daughter's Age 5-15	0.3440	0.3991	0.3918	0.5013	0.4186
	(0.0256)	(0.0232)	(0.0214)	(0.0152)	(0.0116)
	[17019, 839]	[22037, 1203]	[34712, 1825]	[52967, 2565]	[61308, 2979]
Sons in law 20-35	0.3283	0.4394	0.3860	0.4889	0.4143
	(0.0250)	(0.0224)	(0.0218)	(0.0151)	(0.0116)
	[15404, 840]	[20383, 1197]	[30533, 1712]	[46762, 2479]	[54600, 2885]
Sons: Individually linked data		0.4654	0.4751		
		(0.0175)	(0.0120)		
		3947	8847		

Table 4. Intergenerational Elasticities 1850-1930.
Sensitivity to Farmers' Income Imputations.

	(1)	(2)	(3)	(4)	(5)
	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930
Log occupational income in:			**A: Fathers-Sons**		
1950	0.3500	0.3133	0.3440	0.4953	0.4760
	(0.0239)	(0.0200)	(0.0166)	(0.0152)	(0.0118)
1900	0.3502	0.3542	0.3823	0.4471	0.4436
	(0.0222)	(0.0189)	(0.0155)	(0.0121)	(0.0101)
1900, imputed farmer wage	0.3467	0.2879	0.3634	0.4660	0.4701
	(0.0284)	(0.0229)	(0.0196)	(0.0150)	(0.0127)
1950 ex. farmers	0.1899	0.1561	0.1463	0.2540	0.2922
	(0.0476)	(0.0359)	(0.0280)	(0.0322)	(0.0277)
1900 ex. farmers	0.2487	0.2075	0.2320	0.2992	0.2954
	(0.0460)	(0.0374)	(0.0329)	(0.0312)	(0.0259)
1950 ex. farmers (linked sample)		0.2860	0.3266		
		(0.0495)	(0.0340)		
N, no. of names: 1950	[37077, 1182]	[50847, 1478	[80255, 2234	109079, 3253	[122468, 3720]
N, no. of names: 1950 ex. Farmers	[26988, 741]	[36460, 943]	[65726, 1529	[92664, 2337]	[109830, 2845]
			B: Fathers-Sons in Law		
1950	0.3402	0.4009	0.3992	0.4932	0.4136
	(0.0213)	(0.0191)	(0.0183)	(0.0131)	(0.0100)
1900	0.3115	0.4229	0.4120	0.4900	0.4387
	(0.0203)	(0.0192)	(0.0182)	(0.0126)	(0.0100)
1900, imputed farmer wage	0.2509	0.3161	0.3166	0.4415	0.4221
	(0.0242)	(0.0205)	(0.0208)	(0.0146)	(0.0120)
1950 ex. Farmers	0.2150	0.2003	0.1802	0.3270	0.3220
	(0.0465)	(0.0303)	(0.0284)	(0.0288)	(0.0227)
1900 ex. Farmers	0.1986	0.2290	0.2224	0.3490	0.3744
	(0.0403)	(0.0316)	(0.0297)	(0.0289)	(0.0248)
N, no. of names: 1950	[23280, 976]	[30081, 1376	[45804, 2063	[68439, 2888]	[79314, 3326]
N, no. of names: 1950 ex. Farmers	[22586, 697]	[29344, 1004	[44917, 1547	[67488, 2313]	[78026, 2724]

Table 5. Intergenerational Elasticities 1850-1930.
Alternative Measures of Log Occupational Income.

	(1)	(2)	(3)	(4)	(5)
	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930
			A: Fathers-Sons		
1950	0.3500	0.3133	0.3440	0.4953	0.4760
	(0.0239)	(0.0200)	(0.0166)	(0.0152)	(0.0118)
Rank regression	0.2896	0.3001	0.2879	0.3384	0.3510
(rank sample only)	(0.0152)	(0.0137)	(0.0112)	(0.0092)	(0.0080)
Rank regression	0.3161	0.3637	0.3621	0.4250	0.4033
(rank all working age males	(0.0165)	(0.0167)	(0.0137)	(0.0110)	(0.0088)
1990	0.2571	0.2069	0.2388	0.3585	0.4159
	(0.0260)	(0.0217)	(0.0187)	(0.0163)	(0.0140)
ERSCOR50	0.2870	0.3584	0.3427	0.4154	0.4005
	(0.0197)	(0.0203)	(0.0142)	(0.0115)	(0.0091)
SEI	0.2695	0.2979	0.3062	0.4597	0.4684
	(0.0204)	(0.0189)	(0.0157)	(0.0135)	(0.0118)
N, no. of names	[37077, 1182]	[50847, 1478]	[80255, 2234]	[109079, 3253]	[122468, 3720]
			B: Fathers-Sons in Law		
1950	0.3402	0.4009	0.3992	0.4932	0.4136
	(0.0213)	(0.0191)	(0.0183)	(0.0131)	(0.0100)
Rank regression	0.3301	0.4405	0.3975	0.4275	0.3700
(rank sample only)	(0.0163)	(0.0165)	(0.0143)	(0.0102)	(0.0085)
Rank regression	0.3087	0.4429	0.4266	0.4902	0.4074
(rank all working age males	(0.0157)	(0.0171)	(0.0160)	(0.0118)	(0.0092)
1990	0.2137	0.2685	0.2586	0.4418	0.3997
	(0.0229)	(0.0211)	(0.0218)	(0.0161)	(0.0128)
ERSCOR50	0.3031	0.4746	0.4228	0.4934	0.4105
	(0.0196)	(0.0218)	(0.0175)	(0.0123)	(0.0096)
SEI	0.1887	0.3243	0.3244	0.5097	0.4879
	(0.0200)	(0.0203)	(0.0213)	(0.0147)	(0.0124)
N, no. of names	[23280, 976]	[30081, 1376]	[45804, 2063]	[68439, 2888]	[79314, 3326]

Table 6. Intergenerational Elasticities 1850-1930.
Age Controls.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	1850-1870		1860-1880		1880-1900		1900-1920		1910-1930	
Variable:					**A: Fathers-Sons**					
Father's Income	0.3500	0.3523	0.3133	0.3307	0.3440	0.3466	0.4953	0.4855	0.4760	0.4605
	(0.0239)	(0.0240)	(0.0200)	(0.0199)	(0.0166)	(0.0164)	(0.0152)	(0.0151)	(0.0118)	(0.0117)
Father's age		0.0096		0.0009		0.0289		0.0196		0.0183
		(0.0093)		(0.0080)		(0.0060)		(0.0055)		(0.0043)
Father's age squared		-0.0001		-0.0001		-0.0004		-0.0002		-0.0002
		(0.0001)		(0.0001)		(0.0001)		(0.0001)		(0.0001)
Son's age		0.1075		0.0879		0.1014		0.0907		0.1174
		(0.0069)		(0.0058)		(0.0048)		(0.0044)		(0.0039)
Son's age squared		-0.0017		-0.0013		-0.0015		-0.0014		-0.0018
		(0.0001)		(0.0001)		(0.0001)		(0.0001)		(0.0001)
N, no. of names	[37077, 1182]		[50847, 1478]		[80255, 2234]		[109079, 3253]		[122468, 3720]	
					B: Fathers-Sons in Law					
Father's Income	0.3402	0.3330	0.4009	0.3873	0.3992	0.3987	0.4932	0.4869	0.4136	0.4077
	(0.0213)	(0.0219)	(0.0191)	(0.0192)	(0.0183)	(0.0183)	(0.0131)	(0.0134)	(0.0100)	(0.0102)
Father's age		0.0062		0.0106		0.0016		0.0093		0.0046
		(0.0100)		(0.0085)		(0.0073)		(0.0059)		(0.0040)
Father's age squared		-0.0001		-0.0002		-0.0000		-0.0001		-0.0001
		(0.0001)		(0.0001)		(0.0001)		(0.0001)		(0.0000)
Son's age		0.0447		0.0328		0.0282		0.0179		0.0249
		(0.0029)		(0.0020)		(0.0018)		(0.0013)		(0.0013)
Son's age squared		-0.0006		-0.0004		-0.0004		-0.0002		-0.0003
		(0.0000)		(0.0000)		(0.0000)		(0.0000)		(0.0000)
N, no. of names	[23280, 976]		[30081, 1376]		[45804, 2063]		[68439, 2888]		[79314, 3326]	

Table 7. Moments and Parameters Used in the Simulations

Moments	Simulation	Data	Source
$Cov(y_t, y_{t-1})/V(y_{t-1})$	0.464	0.465	1860-1880 Linked sample
$V(y_{t-1})$	0.158	0.160	1860-1880 Linked sample
$Cov_{PS}(y_t, y_{t-1})/V_{PS}(y_{t-1})$	0.314	0.313	1860 and 1880 1% samples
$V_{PS}(y_{t-1})$	0.011	0.011	1860 1% sample
Share of top 50 names	0.695	0.695	1860 1% sample
R-squared	0.105	0.111	1860 1% sample

Distance minimizing parameters

γ_1	λ	σ^2_u	σ^2_v	σ^2_{CON}	σ^2_{SES}
0.421	0.191	0.092	0.031	7.833	5.958

Table 8. The Effects of the Features of the Name Distribution on Estimated Elasticities
Simulation Results.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Concetration of the name distribution (σ^2_{con})	Socio-economic content of names (σ^2_{ses})						
	0	1	3	5.958	10	20	30
2.5	η=0.0345	0.1131	0.2301	0.3107	0.3735	0.4343	0.4662
	[share50= 0.3444]	[0.344]	[0.3437]	[0.3452]	[0.3468]	[0.3542]	[0.3651]
	(R^2=0.1078)	(0.1139)	(0.1269)	(0.1421)	(0.1592)	(0.1897)	(0.209)
5	0.0275	0.1073	0.2203	0.3087	0.3757	0.4385	0.4616
	[0.5526]	[0.5524]	[0.5521]	[0.5517]	[0.552]	[0.5542]	[0.5584]
	(0.0894)	(0.0967)	(0.1084)	(0.1232)	(0.1406)	(0.1718)	(0.1901)
7.833	0.0139	0.1160	0.2246	**0.3144**	0.3794	0.4494	0.4746
	[0.6976]	[0.6965]	[0.6958]	**[0.6952]**	[0.6949]	[0.6947]	[0.6972]
	(0.0713)	(0.0774)	(0.0898)	**(0.1053)**	(0.1215)	(0.1519)	(0.1716)
10	0.0146	0.1169	0.2324	0.3148	0.3890	0.457	0.48
	[0.7638]	[0.7638]	[0.7636]	[0.7623]	[0.7615]	[0.7609]	[0.761]
	(0.0605)	(0.0666)	(0.0774)	(0.0922)	(0.1098)	(0.138)	(0.1596)
15	0.0122	0.1209	0.2419	0.3385	0.4009	0.4703	0.4892
	[0.8444]	[0.8447]	[0.8438]	[0.8428]	[0.842]	[0.8408]	[0.8396]
	(0.0441)	(0.0498)	(0.0599)	(0.0736)	(0.09)	(0.1191)	(0.1394)

Notes: The entries in the table represent the estimated moments based on 15 simulated pseudo-panels. The first number in each cell is the pseudo-elasticity; the second number (in parentheses) is the population share of the 50 most popular names; and the third number (in square brackets) is the R^2 in a regression of father's socioeconomic status on a full set of name fixed effects. The values of γ_1, λ, σ^2_u and σ^2_v are set to minimize the distance between simulated and empirical moments listed in Table 4. The entry in bold represents the minimizing combination of σ^2_{SES} and σ^2_{CON}.

Table 9. The Effects of Changes in the Income Generating Process on Intergenerational Pseudo-Elasticities
Simulation Results.

	(1)	(2)	(4)	(5)	(6)	(7)
Persistence of income (γ_1):	Persistence of income shock (λ):					
	0	0.1	0.191	0.3	0.4	0.5
0.1	η=0.0502	0.1070	0.1543	0.2239	0.2931	0.3763
	[share50=0.6953]	[0.6952]	[0.6952]	[0.695]	[0.6956]	[0.6948]
	(R^2=0.105)	(0.1057)	(0.1081)	(0.1109)	(0.1168)	(0.1268)
0.2	0.1024	0.1591	0.2080	0.2796	0.3496	0.4340
	[0.6953]	[0.6952]	[0.6952]	[0.695]	[0.6956]	[0.6948]
	(0.1039)	(0.1053)	(0.1081)	(0.1115)	(0.1182)	(0.1292)
0.3	0.1518	0.2084	0.2591	0.3330	0.4039	0.4897
	[0.6953]	[0.6952]	[0.6952]	[0.695]	[0.6956]	[0.6948]
	(0.1022)	(0.1041)	(0.1074)	(0.1113)	(0.1186)	(0.1305)
0.421	0.2049	0.2613	**0.3144**	0.3915	0.4641	0.5522
	[0.6953]	[0.6952]	**[0.6952]**	[0.695]	[0.6956]	[0.6948]
	(0.0992)	(0.1016)	**(0.1053)**	(0.1097)	(0.1175)	(0.1303)
0.5	0.2331	0.2892	0.3438	0.4233	0.4975	0.5878
	[0.6953]	[0.6952]	[0.6952]	[0.695]	[0.6956]	[0.6948]
	(0.0967)	(0.0993)	(0.1031)	(0.1077)	(0.1156)	(0.1289)
0.6	0.2582	0.3137	0.3701	0.4526	0.5291	0.6236
	[0.6953]	[0.6952]	[0.6952]	[0.695]	[0.6956]	[0.6948]
	(0.0928)	(0.0956)	(0.0994)	(0.104)	(0.1118)	(0.1251)

Notes: The entries in the table represent the estimated moments based on 15 simulated pseudo-panels. The first number in each cell is the pseudo-elasticity; the second number (in parentheses) is the population share of the 50 most popular names; and the third number (in square brackets) is the R^2 in a regression of father's socioeconomic status on a full set of name fixed effects. The values of σ^2_{SES}, σ^2_{CON}, σ^2_u and σ^2_v are set to minimize the distance between simulated and empirical moments listed in Table 4. The entry in bold represents the minimizing combination of γ_1 and λ.

Table 10. Fertility and Birth order

	(1)	(2)	(3)	(4)	(5)
	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930
	A: Fathers-Sons				
Baseline	0.3500	0.3133	0.3440	0.4953	0.4760
	(0.0239)	(0.0200)	(0.0166)	(0.0152)	(0.0118)
Control for number of siblings	0.2836	0.2735	0.3444	0.5024	0.4740
	(0.0255)	(0.0214)	(0.0168)	(0.0157)	(0.0121)
Control for birth order	0.3277	0.2860	0.3433	0.4974	0.4642
	(0.0247)	(0.0207)	(0.0166)	(0.0154)	(0.0119)
N, no. names (baseline)	[37077, 1182]	[50847, 1478]	[80255, 2234]	[109079, 3253]	[122468, 3720]
	B: Fathers-Sons in Law				
Baseline	0.3402	0.4009	0.3992	0.4932	0.4136
	(0.0213)	(0.0191)	(0.0183)	(0.0131)	(0.0100)
Control for number of siblings	0.2920	0.3044	0.3949	0.4651	0.3815
	(0.0239)	(0.0210)	(0.0190)	(0.0140)	(0.0109)
Control for birth order	0.3289	0.3659	0.3962	0.4734	0.3951
	(0.0215)	(0.0197)	(0.0184)	(0.0133)	(0.0104)
N, no. names (baseline)	[23280, 976]	[30081, 1376]	[45804, 2063]	[68439, 2888]	[79314, 3326]

NOTE: control for birth order = control for share of 'first name' that is first born, second born, 3+ born

Table 11. Immigration and Internal Migration

	(1)	(2)	(3)	(4)	(5)
	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930
	A: Fathers-Sons				
Baseline	0.3500	0.3133	0.3440	0.4953	0.4760
	(0.0239)	(0.0200)	(0.0166)	(0.0152)	(0.0118)
Control for immigrant status	0.2992	0.2769	0.3247	0.4705	0.4659
	(0.0235)	(0.0198)	(0.0165)	(0.0151)	(0.0118)
Control for internal migrant status	0.2984	0.2766	0.3249	0.4708	0.4667
	(0.0235)	(0.0198)	(0.0164)	(0.0151)	(0.0118)
Control for immigrant status and father's		0.2367	0.2883	0.4420	0.4368
		(0.0195)	(0.0163)	(0.0150)	(0.0117)
Control for internal migrant status and father's		0.2328	0.2862	0.4387	0.4342
		(0.0195)	(0.0163)	(0.0150)	(0.0117)
N, no. names (baseline)	[37077, 1182]	[50847, 1478]	[80255, 2234]	[109079, 3253]	[122468, 3720]
	B: Fathers-Sons in Law				
Baseline	0.3402	0.4009	0.3992	0.4932	0.4136
	(0.0213)	(0.0191)	(0.0183)	(0.0131)	(0.0100)
	0.2720	0.3625	0.3676	0.4773	0.4086
Control for immigrant status	(0.0211)	(0.0190)	(0.0182)	(0.0131)	(0.0101)
	0.2722	0.3619	0.3640	0.4733	0.4043
Control for internal migrant status	(0.0211)	(0.0190)	(0.0182)	(0.0131)	(0.0100)
		0.3254	0.3122	0.4433	0.3815
Control for immigrant status and father's		(0.0188)	(0.0180)	(0.0131)	(0.0101)
		0.3215	0.3051	0.4372	0.3743
Control for internal migrant status and father's		(0.0188)	(0.0180)	(0.0130)	(0.0100)
N, no. names (baseline)	[37077, 1182]	[50847, 1478]	[80255, 2234]	[109079, 3253]	[122468, 3720]

Table 12. Intergenerational Elasticities 1850-1930.
By Region of Birth.

	(1)	(2)	(3)	(4)	(5)
	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930
	A: Fathers-Sons				
All	0.3500	0.3133	0.3440	0.4953	0.4760
	(0.0239)	(0.0200)	(0.0166)	(0.0152)	(0.0118)
Control for state of residence	0.2765	0.1943	0.2108	0.2746	0.2799
	(0.0228)	(0.0189)	(0.0156)	(0.0142)	(0.0111)
Control for indicators of economic development by state of residence	0.2784	0.1975	0.2013	0.2633	0.2656
	(0.0228)	(0.0188)	(0.0156)	(0.0142)	(0.0110)
N, no. names (all)	[37077, 1182]	[50847, 1478]	[80255, 2234]	[109079, 3253]	[122468, 3720]
	B: Fathers-Sons in Law				
All	0.3402	0.4009	0.3992	0.4932	0.4136
	(0.0213)	(0.0191)	(0.0183)	(0.0131)	(0.0100)
Control of region of residence	0.2474	0.2947	0.2509	0.3199	0.2600
	(0.0205)	(0.0182)	(0.0175)	(0.0127)	(0.0099)
Control for indicators of economic development by state of residence	0.2513	0.2988	0.2517	0.3177	0.2550
	(0.0204)	(0.0181)	(0.0174)	(0.0127)	(0.0098)
N, no. names (all)	[23280, 976]	[30081, 1376]	[45804, 2063]	[68439, 2888]	[79314, 3326]

Table 13. Intergenerational Elasticities 1850-1930.
By Region of Birth.

	(1)	(2)	(3)	(4)	(5)
	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930
	A: Fathers-Sons				
All	0.3500	0.3133	0.3440	0.4953	0.4760
	(0.0239)	(0.0200)	(0.0166)	(0.0152)	(0.0118)
Northeast	0.2948	0.2539	0.1677	0.2187	0.1918
	(0.0383)	(0.0337)	(0.0310)	(0.0279)	(0.0224)
Midwest	0.1499	0.2521	0.2677	0.2771	0.2701
	(0.0468)	(0.0368)	(0.0315)	(0.0279)	(0.0230)
South	0.4593	0.1591	0.2878	0.3081	0.3641
	(0.0564)	(0.0337)	(0.0311)	(0.0293)	(0.0229)
N, no. names (all)	[37077, 1182]	[50847, 1478]	[80255, 2234]	[109079, 3253]	[122468, 3720]
N, no. names (northeast)	[11461, 580]	[14846, 672]	[19327, 727]	[23818, 891]	[29959, 1040]
N, no. names (midwest)	[7091, 442]	[12713, 629]	[25372, 1039]	[35418, 1406]	[38069, 1589]
N, no. names (south)	[7709, 474]	[11481, 607]	[16570, 973]	[23490, 1558]	[30305, 1965]
	B: Fathers-Sons in Law				
All	0.3402	0.4009	0.3992	0.4932	0.4136
	(0.0213)	(0.0191)	(0.0183)	(0.0131)	(0.0100)
Northeast	0.2014	0.2221	0.3111	0.2743	0.2100
	(0.0380)	(0.0382)	(0.0409)	(0.0333)	(0.0261)
Midwest	0.3471	0.3811	0.3289	0.3371	0.3015
	(0.0520)	(0.0353)	(0.0337)	(0.0238)	(0.0183)
South	0.3975	0.3303	0.3192	0.4649	0.3791
	(0.0478)	(0.0286)	(0.0306)	(0.0252)	(0.0178)
N, no. names (all)	[23280, 976]	[30081, 1376]	[45804, 2063]	[68439, 2888]	[79314, 3326]
N, no. names (northeast)	[6602, 448]	[8102, 559]	[9741, 602]	[12819, 769]	[16865, 923]
N, no. names (midwest)	[4877, 354]	[7883, 586]	[14957, 964]	[22529, 1340]	[24911, 1457]
N, no. names (south)	[5337, 408]	[7200, 587]	[10413, 926]	[16556, 1335]	[21104, 1625]

Appendix Table 1. Summary Statistics for Farm Status

Children, 0-15			Adults, 20-35		
Census year	Total	Percent with farmer father	Census year	Total	Percent farmer, or with farmer husband
Males					
1850	35,597	57.31	1870	42,287	24.36
1860	48,114	51.84	1880	58,974	25.98
1880	75,004	48.63	1900	89,874	16.07
1900	103,817	40.63	1920	122,916	13.44
1910	117,612	35.46	1930	138,323	8.87
Females (married adults only)					
1850	34,272	57.22	1870	26,055	39.50
1860	46,874	50.64	1880	33,434	40.68
1880	72,160	47.66	1900	49,824	31.03
1900	101,516	40.01	1920	74,175	22.92
1910	114,074	34.83	1930	85,809	15.84

Appendix Table 2. Calculation of 1900 Wage Assigned to Farmers

Name of 1910 occupation assigned 1950 occupation of "farmer" at least once	Percent of "farmers" with this occupation in 1910	Default weight assigned to this occupation's wage (out of 1)	1901 wage assigned to this 1910 occupation
Farmers, general farms (owners)	57.55%		**576**
Farmers, general farms (tenants)	35.97%		**334**
Farm laborers, home farm	2.07%	0.384	255
Gardeners	1.37%	0.253	413
Dairy farmers	1.05%		**576**
Stock raisers	0.75%	0.139	334
Fruit growers	0.73%	0.135	334
Poultry raisers	0.17%	0.032	334
Florists	0.15%	0.028	593
Farm laborers, working out	0.04%	0.008	255
Nurserymen	0.04%	0.007	593
Dairy foremen, general farms	0.03%	0.006	750
Sugar cane farmer	0.02%		**576**
Apiarists	0.02%	0.003	334
Livery stable keepers and managers	0.01%	0.002	502
Coffee farmers	0.01%		**576**
Stock herders, drovers, and feeders	0.01%	0.001	334
Other and not specified pursuits	0.01%	0.001	334
Garden laborers	0.00%	0.001	255
Orchard and nursery laborers	0.00%	0.001	255
Corn shellers, hay balers, grain threshers, etc.	0.00%	0.000	255
Policemen	0.00%	0.000	887
Default wage to farmers:			335.04
Wage to farmers with income imputation:			475.93

Appendix Table 3. Summary Statistics for Immigrant Status

Children, 0-15				Adults, 20-35			
Census year	Total	Percent immigrant	Percent with foreign-born father	Census year	Total	Percent immigrant	Percent with foreign-born father
Males							
1850	35,597	4.20	15.17	1870	42,287	26.54	-
1860	48,114	3.85	25.27	1880	58,974	19.95	36.24
1880	75,004	2.83	31.38	1900	89,874	20.10	42.38
1900	103,817	2.50	28.61	1920	122,916	19.16	39.33
1910	117,612	2.99	28.08	1930	138,323	13.16	34.34
Females (married adults only)							
1850	34,272	4.39	15.40	1870	26,055	25.12	-
1860	46,874	3.90	25.28	1880	33,434	19.78	33.51
1880	72,160	2.80	31.83	1900	49,824	20.02	39.70
1900	101,516	2.39	28.91	1920	74,175	19.41	37.99
1910	114,074	3.08	28.29	1930	85,809	13.90	33.09

Appendix Table 4: 30-year elasticities

	(1)	(2)	(3)	(4)
	1850-1880	1870-1900	1880-1910	1900-1930
Sample:				
Sons: baseline	0.2311	0.3108	0.3189	0.3871
	(0.0185)	(0.0165)	(0.0156)	(0.0123)
N, no. names	[37778, 1240]	[64972, 1645]	[83447, 2240]	[115713, 3313]
Sons in law: baseline	0.2913	0.3315	0.3726	0.4144
	(0.0189)	(0.0167)	(0.0174)	(0.0108)
N, no. names	[26311, 1093]	[43954, 1655]	[56494, 2105]	[87271, 3152]

Table A5. Intergenerational Elasticities 1850-1930.
Sensitivity to Different Name Coding Schemes.

	(1)	(2)	(3)	(4)	(5)
	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930
Name concept:	**A: Fathers-Sons**				
All	0.3500	0.3133	0.3440	0.4953	0.4760
	(0.0239)	(0.0200)	(0.0166)	(0.0152)	(0.0118)
Middle initials	0.3400	0.3112	0.3291	0.4189	0.4389
	(0.0230)	(0.0191)	(0.0156)	(0.0136)	(0.0111)
Nicknames	0.3673	0.3310	0.3412	0.4489	0.4268
	(0.0246)	(0.0207)	(0.0176)	(0.0159)	(0.0123)
Soundex codes	0.4212	0.4041	0.4771	0.5571	0.5530
	(0.0304)	(0.0250)	(0.0223)	(0.0184)	(0.0155)
N, no. names (All)	[37077, 1182]	[50847, 1478]	[80255, 2234]	[109079, 3253]	[122468, 3720]
N, no. names (M.I.)	[36685, 1419]	[50243, 1789]	[79227, 2676]	[107721, 3910]	[120706, 4605]
N, no. names (Nicknames)	[37172, 1138]	[50947, 1415]	[80315, 2107]	[109098, 3111]	[122501, 3581]
N, no. names (Soundex)	[39262, 887]	[54941, 995]	[84686, 1248]	[116154, 1595]	[130274, 1623]
	B: Fathers-Sons in Law				
All	0.3402	0.4009	0.3992	0.4932	0.4136
	(0.0213)	(0.0191)	(0.0183)	(0.0131)	(0.0100)
Middle initials	0.3441	0.3619	0.3771	0.4249	0.3834
	(0.0208)	(0.0179)	(0.0170)	(0.0122)	(0.0096)
Nicknames	0.4360	0.4152	0.4135	0.4551	0.3882
	(0.0258)	(0.0204)	(0.0189)	(0.0140)	(0.0107)
Soundex codes	0.5907	0.5543	0.5570	0.6122	0.4944
	(0.0305)	(0.0257)	(0.0256)	(0.0176)	(0.0134)
N, no. names (All)	[23280, 976]	[30081, 1376]	[45804, 2063]	[68439, 2888]	[79314, 3326]
N, no. names (M.I.)	[22954, 1142]	[29682, 1644]	[45239, 2459]	[67637, 3496]	[77963, 4083]
N, no. names (Nicknames)	[23627, 945]	[30152, 1309]	[45814, 1958]	[68445, 2787]	[79322, 3227]
N, no. names (Soundex)	[25482, 566]	[32626, 705]	[48695, 855]	[72906, 1113]	[84541, 1198]

a. Separate group for those with middle initials: "John" is different from "John M." who is the same as "John H."
b. Common nicknames grouped associated first name: "Johnny" is the same as "John".