

NBER WORKING PAPER SERIES

IN THE NAME OF THE SON (AND THE DAUGHTER): INTERGENERATIONAL
MOBILITY IN THE UNITED STATES, 1850-1930

Claudia Olivetti
M. Daniele Paserman

Working Paper 18822
<http://www.nber.org/papers/w18822>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2013

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, ECARES, LSE and Toulouse for many helpful suggestions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2013 by Claudia Olivetti and M. Daniele Paserman. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1930

Claudia Olivetti and M. Daniele Paserman

NBER Working Paper No. 18822

February 2013

JEL No. J11,J62,N31

ABSTRACT

This paper provides a new perspective on intergenerational mobility in the United States in the late 19th and early 20th centuries. We devise an empirical strategy that allows to calculate intergenerational elasticities between fathers and children of both sexes. The key insight of our approach is that the information about socio-economic status conveyed by first names can be used to create a pseudo-link not only between fathers and sons, but also between fathers and daughters. The latter is typically not possible with historical data.

We find that the father-son elasticity in economic status grows throughout the sample period. Intergenerational elasticities for daughters follow a broadly similar trend, but with some differences in timing. 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, with increasing returns to human capital contributing to explain the residual. Other mechanisms such as changes in fertility, migration, and investment in public schooling, appear to have had only a minor role in explaining the trends.

Claudia Olivetti
Boston University
Department of Economics
270 Bay State Road
Boston, MA 02215
and NBER
olivetti@bu.edu

M. Daniele Paserman
Department of Economics
Boston University
270 Bay State Road
Boston, MA 02215
and NBER
paserman@bu.edu

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. A low degree of intergenerational mobility can undermine the notion of equality of opportunity and may lead to persistent inequality. Recent research reveals that today intergenerational mobility in the U.S. is lower than in most other developed countries (Corak, 2013). This finding stands in contrast with the national *ethos* and the worldwide perception of the United States as the land of unlimited opportunity. Was this view ever justified?

Long and Ferrie (2007, forthcoming) date the “end of American exceptionalism” to the early decades of the 20th Century. Their analysis is based on the comparison of fathers’ and sons’ occupational status. This comparison, however, may be missing part of the picture. Daughters should be included if we want to know how the *average* well-being of a generation correlates with that of their parents. If there is a strong stratification in marriage by social class, assortative mating might magnify individual-level intergenerational persistence. Moreover, to the extent that mothers play a key role in the human capital accumulation of their children, investment in daughters could have important consequences on the transmission of status across multiple generations. Therefore, to reach a fuller understanding of the transmission of resources across generations, it is important to focus on daughters as well as sons.

This paper contributes to the existing literature by measuring the degree of intergenerational mobility between fathers and their children of both sexes. Typically, the estimation of intergenerational income elasticities is based on a regression of an individual’s economic status at time t on that of his or her own father at time $t - k$. This requires the use of longitudinal data sets that link fathers to their offspring. Starting with the seminal work by Ferrie (1995), 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.

One important contribution of this paper is that it develops an empirical strategy that enables us to estimate the intergenerational elasticity between fathers and daughters, as well as between fathers and sons, even when it is impossible to link individuals directly across generations. Since most married women in the second half of the 19th Century were not gainfully employed, we measure daughters’ economic well-being using their husbands’ socioeconomic status. The key insight of our approach is that the information about socioeconomic status conveyed by first names can be used to create a *pseudo-link* between fathers and sons, as well as fathers and daughters.

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 empirical strategy amounts to creating a “generated regressor” by using one sample to create a proxy for an unobserved regressor in a second sample. The key assumption is that the generated regressor can be imputed by first name. This is essentially a “two-sample two-stage least squares” estimator (TS2SLS, Inoue and Solon, 2010). In the first step, we use the sample of fathers and regress father's log earnings on a full set of children's first name dummies. In the second step, we use the sample of sons, and regress son's log earnings on the cross-sample predicted values from the first step. We sometimes refer to this estimator as a *pseudo-panel* estimator, as it is based on the creation of pseudo-links across generations.

We estimate father/son and father/son-in-law intergenerational income elasticities using 1% IPUMS samples between 1870 and 1930. Thus, we obtain a consistent and continuous measure of mobility for a relatively long period. This can be used to shed light on the timing of the decline in intergenerational mobility in the U.S. as well as its demographic and economic determinants.

Our results indicate that the intergenerational elasticity between fathers and sons increased 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 document a marked decrease in intergenerational mobility in the United States between the late 19th century and the middle of the 20th century. The intergenerational elasticity between fathers and sons-in-law displays a similar trend although the timing is slightly different. The father/son-in-law elasticity is higher than the father/son elasticity in the early part of the period but the two elasticities converge by the turn of the century. In the last period of the sample the father/son-in-law elasticity dips below the

father/son elasticity, a ranking that is similar to that found in modern studies. The main findings are robust to different methods of imputing income and coding names, to the use of different outcome variables and to differential mortality across socioeconomic groups and selection into marriage.

We also conduct a numerical exercise to study how the pseudo-panel estimator responds to changes in the name distribution. 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.

Finally, we investigate which historical developments may explain the trends and the gender differentials in intergenerational elasticity. We find that most of the increase in the intergenerational elasticity can be accounted for by regional differences in economic development, with increasing returns to human capital contributing to explain the residual. Other mechanisms such as changes in fertility, migration, and investment in public schooling, seem less likely to matter. Gender differentials in elasticities can be plausibly explained by imbalances in the sex ratio due to maternal and infant mortality, and changes in migratory flows.

Our paper is related to an extensive literature that studies intergenerational mobility using modern panel data (see the comprehensive surveys by Solon, 1999, and Black and Devereux, 2011). The bulk of the literature focuses on father-son intergenerational mobility and finds an intergenerational labor income elasticity hovering around 0.4. Only a 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/son-in-law elasticity - estimated at 0.360. Raaum et al. (2007) confirm this result for the US, the UK, and three Nordic countries. Associated to the increasing labor force participation of women, recent studies have focused on the father-daughter occupational mobility. Jäntti et al. (2006) document that in five of six developed countries, the father/son intergenerational elasticity is higher than the father/daughter one. Hellerstein and Morrill (2011) find that

the probability that a woman works in the same occupation as her father has increased over the course of the 20th Century.

Closely related to our project is the work by Güell 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. Others have instead exploited the distribution of surnames in data sets that are centuries apart to estimate long-run social mobility: Clark and Cummins (2012) use the distribution of surnames in England, and conclude that there is considerable persistence of status in the UK between 1800 and 2012. Collado et al. (2012), using data from two Spanish regions, find that socioeconomic status at the end of the 20th Century still depends heavily on the socioeconomic status of one's great-great grandparents.¹

The rest of the paper proceeds as follows. The next section presents a simple model of marriage and intergenerational mobility that defines the parameters of interest and facilitates the interpretation of our empirical results. Section 3 describes the econometric methodology. Section 4 presents the data and discusses measurement issues. The main results are presented in Section 5. Section 6 provides robustness checks and section 7 explores alternative factors underlying the trends. Section 8 concludes.

2 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 and 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

¹Clark (2013) summarizes similar studies from several other countries and concludes that surnames lose information about social status at rates much slower than implied by modern studies.

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: $\beta_1 \log c_{t-1} + \beta_2 E[\log(Y_{M,t})] + \beta_3 E[\log(Y_{F,t})]$, 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: $c_{t-1} + p_H (H_{M,t} + H_{F,t}) \leq y_{t-1}$ 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:

$$y_t = H_{M,t}^{\gamma_1} \exp(E_{M,t}), \quad (1)$$

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 expression:

$$Y_{F,t} = y_{SIL,t} = H_{F,t}^{\alpha_1} \exp(E_{F,t}), \quad (2)$$

where $E_{F,t}$ represents the combined effect of all determinants, other than human capital, of the earnings of the woman's husband.²

²Bhaskar and Hopkins (2012) show that a matching model of the marriage market with non-transferable utility and pre-marital investment with stochastic returns yield an equilibrium relationship between parental investments in sons and daughters that mirrors equation (2).

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 of 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 measures 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 opportunities, the optimal parental human capital investment is obtained by substituting (1) and (2) 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:

$$\log y_t = \gamma_1 \log y_{t-1} + e_t + u_t. \tag{3}$$

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

$$\log y_{SIL,t} = \alpha_1 \log y_{t-1} + \theta e_t + \mu_t. \tag{4}$$

Equations (3) and (4) 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:

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

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 (4) the expression for the father/son-in-law intergenerational elasticity is given by:

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

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.

3 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 \tilde{y}_{jt-1} , the average log earnings of fathers of children named j at time $t - 1$. We have thus created a “generated regressor” by using 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.³ 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. Because we are effectively creating a pseudo-panel of individuals linked by first names, we refer to this estimator as the “pseudo-panel” estimator, and label it $\hat{\eta}_{PSEUDO}$.⁴

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 \tilde{y}_{jt-1} as a predictor of 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.

There is abundant empirical evidence supporting the assumption that parents choose first names partly to signal their own standing in society, or their cultural and religious beliefs. 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, Fryer and Levitt (2004) show that names provide a strong signal of socioeconomic status for blacks, but also that there are systematic and large differences in name choices by whites with different levels of education. This practice is also widespread in other societies, both today and in the past. Head and Mayer (2008) investigate the social transmission of parental preferences through naming patterns in France. Hacker (1999) and

³The TS2SLS estimator is in itself a special case of a Two-Sample IV estimator (Angrist and Krueger, 1992).

⁴The second stage has a particularly simple structure because the right hand side variable is constant for every individual with the same first name. Therefore, in the special case of no additional regressors, the TS2SLS estimator is equivalent to a weighted least squares regression of \tilde{y}_{jt} on \tilde{y}_{jt-1} , where \tilde{y}_{jt} is the average log earnings of adults named j at time t , and the weights are equal to the frequency counts of first names in the son’s sample. This equivalence highlights the similarity between our approach and the synthetic cohort method pioneered by Browning 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. Aaronson and Mazumder (2008) use an estimation strategy that is also based on synthetic cohorts. They estimate intergenerational mobility in the US between 1940 and 2000 by imputing father’s income using state of birth.

Haan (2005) document a relationship between first names, religiosity and fertility in Canada and the US during the 19th Century. Cook et al. (2012) find that distinctively black names were already common in the post-Civil War period.

Some of these papers also document that first names can have consequences on economic outcomes. This may raise the concern that our empirical strategy may be confounded by parents strategically choosing “aspirational” names for their children. Parents may believe that by choosing names that are associated with a higher social class they may facilitate their children’s social mobility. Alternatively, parents might refrain from using ethnic names to prevent discrimination. These practices would make names a more noisy indicator of parental socio-economic status. Thus, the pseudo-panel estimator of the intergenerational elasticity would become more susceptible to attenuation bias, making us less likely to find any evidence of a significant effect. This can still be a concern for the interpretation of the *trends* in intergenerational elasticity, if the degree of strategic naming changes over time. As we document below, however, the informational content of names has remained remarkably stable over time. Moreover, we show in Section 7.1 that our estimates is fairly insensitive to changes in the distribution of names. In sum, we view it as highly unlikely that strategic naming affects our estimates in a meaningful way.

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. Our estimator boils down to a regression of son-in-law’s income on father-in-law’s income, where father-in-law’s income for men married to women named j is proxied by the average income of fathers of daughters named j at time $t - 1$.

4 Data

We now apply our methodology to data from the 1850 to 1930 US Census 1% samples from IPUMS (Ruggles et al., 2010), when first names are available.

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 minimize the impact of measurement error and temporary fluctuations in income (Solon, 1992) and are quite close to estimates based on predicted income by occupation (Björklund and Jäntti, 1997).

One of the advantages of the IPUMS data set is that it contains a harmonized classification of occupations, and several measures of occupational status that are comparable across years. For our benchmark analysis, we choose the OCCSCORE measure of occupational standing.⁵ 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 6.1.⁶

Coding of names. The second challenge, specific to our methodology, is how to correctly match first names across censuses. In our benchmark classification of names we ignore middle initials (that is, we treat “William” as equivalent to “William J.”) and we treat nicknames as distinct names (that is, “William” and “Bill” are considered two different names).⁷ 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 Appendix A. Specifically we examine the following variants to our basic specification: a) Separating names with and without middle initials; b) grouping the main root of a name with its most common nicknames and c) using the Soundex phonetic algorithm to deal with potential misspelling of names. All these different name coding schemes yield the same pattern of results as in the baseline.

The Distribution of Names. We first document some features of the distribution of first names in the sample. Table 1 reports the summary statistics for children’s names in the initial year of the pseudo-panel by gender. Both population (column 1) and the number of distinct names (column 2) grow between 1850 and 1910, but the average number of observations per name (column 3) is roughly constant. This pattern is common across genders. In every decade, a large proportion of names appears only once in the sample

⁵A number of other papers have used this same variable to measure occupational standing, among them Abramitsky and Platt-Boustan (2012), Cverk (2012), Jones and Tertilt (2008) and Katz and Margo (forthcoming).

⁶For 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>.

⁷The 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.).

(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’s occupational income on a set of name indicators. Note that if names were assigned at random, and we had a sufficiently large number of occurrences for every name, the between-name variation would not explain any of the total variation in father’s income, and the R^2 coefficient would be equal to zero. The entries in the column show that the between name variation varies by gender: it accounts for 11% to 13% of the total variation in fathers’ log earnings for boys and 13% to 15% for girls.⁸

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.

5 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 increases 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

⁸Because of the large number of singleton names, we could observe a positive R^2 even if names were assigned completely at random. To adjust for this, we calculate the R^2 that would result if names were assigned completely at random (R_{sim}^2), given the actual distribution of names. We then calculate an “adjusted R^2 ” as $(R^2 - R_{sim}^2) / (1 - R_{sim}^2)$. The adjusted coefficient remains sizable.

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 the timing of the increase is somewhat different. The father/son-in-law elasticity exhibits a first increase between 1870 and 1880 and then a further jump between 1900 and 1920, which coincides with the increase in the father-son elasticity. The two elasticities are almost identical in 1920 but they diverge at the end the period with the father/son-in-law elasticity dipping below the father/son elasticity. Interestingly, the ranking of son-in-law and son elasticities is consistent with modern estimates for the US and other developed economies (Chadwick and Solon, 2002, Raaum et al., 2007). We defer to section 7.2 for a discussion of the historical developments that can rationalize these findings.

The remaining rows in Table 3 show how our benchmark estimates are affected by sample selection issues due to either differences in child mortality by socioeconomic status, or to differences in the age distribution and marital status of sons and sons-in-law.

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

In all societies men marry later in life than women and the gender differential in age at first marriage tends to be largest in more traditional societies. The 19th century US is no exception. As documented in Ferrie and Rolf, (2008) and Fitch and Ruggles (2000), the male-female differential in median age at first marriage was quite large in the 19th Century, peaked in 1900 at more than 4 years, and then declined to about 2 years at the beginning of the 20th Century. In our samples this implies that sons-in-law are, on average, older than sons (especially at the beginning of the period) and that a fraction of the sons are unmarried. Failing to control for differences in the age distribution has the potential to affect the comparison of father/son-in-law and father/son elasticities. In particular, if the wage-age profile is concave, and sons are systematically younger than their brothers-in-law, we would systematically overestimate the father/son-in-law elasticity relative to the father/son

elasticity. In the third and sixth rows of Table 3 we attempt to make the son and son-in-law samples more comparable in terms of their demographic characteristics. In the third row, we restrict the sons sample to married individuals. In the sixth row, we only include individuals aged 20 to 35 in the sample of sons-in-law. There is some variation in the point estimates, but on the whole the results are very similar to the benchmark.

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.⁹ Not surprisingly, the individually linked data yields intergenerational elasticities that are 28-33% higher than those obtained with the pseudo-panel estimator. This is likely because of attenuation bias induced by measurement error in father’s earnings when we take averages by first names. Thus, our estimate can be interpreted as a lower bound to the actual intergenerational elasticities.¹⁰

6 Robustness

6.1 Measuring Income

As is well known, the 1950 income distribution was relatively compressed (Goldin and Margo, 1992). Moreover, the 1950 occupational classification may not reflect accurately the relative standing of occupations that were common during the late 19th Century and early 20th Century. This issue is important from our standpoint as ‘farmers’ represent a large part of our sample and farming occupations and farm ownership were associated with higher socioeconomic status during our sample period than in 1950.¹¹ As pointed out by Xie and Killewald (2011) measures of occupational mobility during this period of structural transformation can be sensitive to the treatment of farmers. We address these concerns by studying whether our estimates are sensitive to alternative imputations of occupational income, paying special attention to the imputation of farmers’ income. The results are reported in Table 4. The first row of each panel reproduces the benchmark estimates from Table 3.

⁹Source: IPUMS Linked Representative Sample (Ruggles et al., 2010). Since the linking is done using information on first and last names, no linked data on married women is available. Therefore, we can only compute father-son elasticities.

¹⁰The intergenerational *correlation* may differ from the elasticity if the dispersion of earnings varies substantially across generations. We find that this is not the case: the magnitude and trends of intergenerational correlations are almost identical to the elasticities reported in Table 3.

¹¹The 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.

We start by imputing income using the 1900 occupational-earnings distribution obtained from the tabulations in Preston and Haines (1991). These tabulations are based on the 1901 Cost of Living Survey, which was designed to investigate the cost of living of families in industrial locales in the United States. The main advantage of using the 1900 occupational income distribution is that the list of occupational categories matches more closely the list, types and ranking of occupations that were common during much of the sample period. This categorization, however, suffers from two limitations. First, the 1901 survey collected data for the “typical” urban family, meaning that by construction the resulting income distribution is more compressed than what one would obtain in a representative sample. Second, while Preston and Haines do impute income for some agricultural occupations, they explicitly refrain from imputing an average income for generic farm owners and farm tenants. We experiment with two different methods for imputing farmers’ income.

Our first imputation method assigns to all individuals coded as “farmer” in the 1950 coding scheme a weighted average of income for all farming occupations present in the Preston-Haines classification. 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 (\$335.04) to all farmers (see Appendix Table 1 for the details of the calculation). As an alternative, we impute farmers’ income based on the methodology originally described by Mitchell et al. (1922) and recently used by Abramitzky et al. (2012). For owner-occupier farmers, we calculate 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 (this results in an imputed annual income of \$576). 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. Under this second imputation rule we assign an income of \$475.93 to all “farmers” in the 1950 coding scheme.

The intergenerational elasticities estimates based on the different imputation methods are reported in the second and third rows of each panel in Table 4. The father/son intergenerational elasticity is not very sensitive to using the 1900 occupational income distribution nor to the exact imputation of farmers’ income. The estimate of the father/son-in-law intergenerational elasticity is very similar to the benchmark under the first imputation method but the results differ somewhat when we impute farmers’ income using the information on the value of farm product. In this case, the estimated elasticity is 8 to 9 points lower relative

to the benchmark in the first part of the sample period (1870 to 1900), but the difference becomes smaller over time. On the whole, the results are quite similar 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.

The next two rows of Table 4 show the estimated elasticities if we completely remove farmers from the analysis, using either the 1950 or the 1900 occupational income distribution. Both the son and son-in-law intergenerational elasticities are substantially lower than those in the benchmark analysis. This reflects the unsurprising fact that farming status is highly correlated across generations so that excluding farmers altogether raises intergenerational mobility. The trends are similar to the benchmark. In addition, we have also calculated the father/son elasticities based on the linked sample excluding farmers (see the last row of the first panel). For the two pair of years in which we can calculate elasticities we find lower estimates relative to the benchmark of Table 3.

We conclude from Table 4 that our basic findings of an increasing trend in elasticities is not affected by the treatment of farmers' income. Therefore, the decline in intergenerational mobility does not seem to be driven by the structural transformation of the U.S. economy, from agricultural to industrial, over this period.

In Table 5 we assess the robustness of our results to additional alternative measures of occupational income. In the second row of each panel we replace occupational income with an individual's percentile rank in the distribution. The rationale for using rank is that it does not depend on the, potentially noisy, imputed level of occupational income. We find an attenuated trend for the father-son elasticity while the father/son-in-law elasticity mirrors the baseline fairly closely. In the next row we re-estimate the model using average occupational incomes in 1990. The 1990 distribution has the advantage of being substantially more dispersed than the 1950 distribution, and therefore allows us to assess whether our measures of intergenerational mobility are affected by the variance of measured earnings. The estimated elasticities are lower than the benchmark estimates, 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 measures of income, 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.

6.2 Life Cycle Effects

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.¹² 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.

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 controls 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 2).

7 What factors can explain the trends?

7.1 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 3 that our estimator 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

¹²In 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.

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 2 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. We match the following moments: 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 ($Cov(y_t, \tilde{y}_{jt-1})/V(\tilde{y}_{jt-1})$ 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^2 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:

$$y_t = \gamma_1 y_{t-1} + e_t + u_t \quad (5)$$

$$e_t = \lambda e_{t-1} + v_t, \quad (6)$$

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 $\{1, 2, \dots, J\}$ is given by:

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

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. Furthermore, they are independent of each other as well as of all other variables in the model. The parameter σ_{CON}^2 determines the concentration of names in the population: the higher σ_{CON}^2 the more likely it is that some names will appear frequently while others are very rare. The parameter σ_{SES}^2 instead determines the sensitivity of names to socio-economic status. In the extreme case of $\sigma_{SES}^2 = 0$ names are assigned randomly. The larger σ_{SES}^2 , 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 $\theta = (\gamma_1, \lambda, \sigma_u^2, \sigma_v^2, \sigma_{CON}^2, \sigma_{SES}^2)$ 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_{CON}^2 and σ_{SES}^2 indicate that the distribution of first names is fairly concentrated, and names do carry economic content.

Sensitivity to name distribution. We now fix $(\gamma_1, \lambda, \sigma_u^2, \sigma_v^2)$ 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_{CON}^2 and σ_{SES}^2 . 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_{CON}^2 . 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_{CON}^2 , 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_{CON}^2 rises from 2.5 to 15.

On the other hand, the estimated elasticity is strongly affected by the parameter σ_{SES}^2 . In particular, the estimated value of η is very close to zero when $\sigma_{SES}^2 = 0$, i.e. names carry no information about a family’s socio-economic status, as discussed in Section 3. The table shows that σ_{SES}^2 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_{SES}^2

would be associated with an increase in the R^2 from 0.105 to 0.172. This is much larger than the observed change in the R^2 over the period (from 0.1108 to 0.1256 – see Table 1, column 8).

It is also unlikely that the gender differential in elasticities is driven by gender differences in naming patterns. The last column of Table 1 showed that in 1850 the economic content of boys' and girls' names was essentially identical, but in the following decades girls' names carry stronger information about their fathers' socioeconomic status. While this pattern may be consistent with the trend in father/son and father/son-in-law elasticities between 1870 and 1880 (the opening of a gender gap in the economic content of names coincides with a divergence in elasticities), it stands in contrast with our findings for the remainder of the period – the gap in the economic content of names widens, but the gap in elasticities closes down and even reverses direction in the last period.

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 from 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^2 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.

7.2 Fertility and Migration

Changes in fertility. The total fertility rate dropped from 5.42 in 1850 to 2.45 in 1930 (Haines, 2008). The drop in fertility is likely to have affected the ability of parents to invest in their children's human capital: a larger family size is associated with 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. The observed elasticity would increase if the fertility decline occurs earlier for the high income group than for the low income group. In this case, the resources of high income parents would be split among fewer children, giving each one an even stronger initial advantage relatively to children from lower income families. 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 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 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 coefficient occurs for cohorts born at the beginning of the 20th century. Thus changes in fertility do not seem to be able to explain the observed trends.

We can further assess this point by directly controlling for fertility in our baseline regression. Ideally, to account for changes in fertility and for potential asymmetries in the allocation of family resources across children, we could control for the number of siblings and birth order. However, information on these variables is not available in the adult sample. Therefore, we control for the average number of siblings and the distribution of birth orders by first name in the children's sample. The results are reported in Table 10. The differences relative to the baseline results are minimal, with the possible exception of the first two cohorts. If anything, accounting for fertility makes the trend in intergenerational mobility even more pronounced.

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

While this hypothesis is appealing at first glance, it appears to be inconsistent with the evolution over time in the intergenerational elasticity estimate. Immigration to the US had an early peak in the 1880s and then a second, larger peak between 1900 and 1915.¹³ 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 1900 and 1910 cohorts.

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 control directly for the effect of immigration in our basic estimates. These could be downward biased if immigrant fathers tend to be employed in low-paying occupations, but their children

¹³U.S. Department of Homeland Security, Office of Immigration Statistics, *Yearbook of Immigration Statistics* (various years).

quickly rise through the social ranks. It is easy to address these issues by simply controlling for the immigrant status of sons and, from 1880 onwards, fathers.¹⁴ For the son-in-law specification we control for the immigrant status of both spouses and their fathers.

The results are presented in the second and third row of Panels A (sons) and B (sons-in-law) in Table 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. 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. 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.¹⁵

Internal mobility Long and Ferrie (forthcoming) argue that residential mobility, either across state or county lines, is a prime candidate to explain the high level of intergenerational mobility in the US in the 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 1930.¹⁶ Therefore, the trends in mobility across states 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

¹⁴We have also estimated the model restricting the whole analysis to exclude all immigrants, or even all children with immigrant fathers. The results are almost identical.

¹⁵One 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.

¹⁶Source: our own calculations from the IPUMS samples

“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, in the remaining rows of Table 11, we add to our basic specification controls for internal migrant status of both generations. We define internal migrants as individuals living in a different state than their state of birth. Contrary to our conjecture, the inclusion of these controls has essentially no effect on the intergenerational elasticity estimates. If anything, the estimates in the first part of the sample period seem to be slightly *upward* biased.

Based on this evidence it does not appear that inter-state mobility explains much of the trend in the intergenerational elasticity estimates. However, as documented in Long and Ferrie (forthcoming), of greater importance for geographic mobility were the movements across counties within a state. They estimate that between 1870 and 1880 the fraction of 20-29 years old white, native born males who changed county was twice as large as the fraction who changed states (50% vs. 26%). Because of lack of data, however, we cannot control for internal mobility at the county level. Therefore, it could still be the case that the rate of return to geographic mobility dropped between the second half of the 19th Century and the first few decades of the 20th Century.

7.3 Regional Differences

The trends in intergenerational elasticity could also be explained by differences in the degree of economic development. If fathers and sons tend to live in the same region, large economic disparities across regions will translate into a high correlation between father’s and son’s income. The period under examination was indeed characterized by large regional disparities. The industrial revolution did not spread uniformly across the United States. Regional income diverged significantly in the second half of the 19th Century. Income per capita in the South fell sharply during the Civil War, absolutely and relatively to other regions, and recovered at a slow pace. By 1900 income per capita in the South was barely half of the national average (see Kim and Margo (2004), p. 2991). Moreover, as discussed previously, geographic mobility was on the decline in the first part of the 20th Century. Therefore, it is possible that the sharp increase in intergenerational elasticity observed in the latter part of our sample period reflects these large geographic disparities in economic development.

To assess this possibility, we include in our basic regressions controls for state of residence. The results are presented in Table 12. In all years, controlling for state of residence (second row) substantially lowers the estimated intergenerational elasticity coefficient, but the decrease is most pronounced in the last two cohorts. Consequently, intergenerational mobility is no lower in the 1910-1930 period than in the 1850-1870 period.

The third row of the table controls directly for two measures of economic development at the state level: urbanization rates, and the fraction of workers employed in agriculture. Both measures capture regional divergence in the second half of the 19th Century. The fraction of the population living in urban areas in the Northeast climbed from 44% to 68% between 1880 and 1910, but then increased by only 4 additional percentage points between then and 1930; by contrast, the urbanization rate in the South was only 11% in 1880, 21% in 1910, and 32% in 1930. Similarly, the fraction of workers employed in agriculture in 1910 was 11% in the Northeast, but 57% in the South.¹⁷ The results are almost identical to those reported in the second row. In other words, much of the differences across states can be parsimoniously summarized by these two variables.

To further understand the role of regional differences, we also conduct our analysis separately for each region of birth. Specifically, for every individual born in a specific region in Census year t , we proxy his father's income by the average income of fathers of children with that first name in Census year $t - 20$, and who lived in the same region. The results are presented in Table 13. The region-specific intergenerational elasticity is almost always lower than the national elasticity, providing further support for the notion that a substantial part of the national estimate is accounted for by regional differences in development. There is also a fairly stable ranking of regions in terms of elasticity, especially towards the end of the sample period, with the Northeast being the most mobile, the South the least mobile, and the Midwest somewhere in between. The trends also differ somewhat across regions: the father/son elasticity declines in the Northeast over time, stays relatively flat in the Midwest, and increases markedly in the South after 1880 (the end of Reconstruction). Similar patterns are observed for the father/son-in-law elasticities, even though the trends are slightly more mixed.

The ranking across regions of the elasticity estimates also sheds light on the role of compulsory schooling and investments in public education in the intergenerational transmission of economic status. The intergenerational elasticity is lowest in the Northeast, where all states had compulsory schooling in 1900 (Lingwall, 2010); is highest in the South, where only 3 of 16 states (plus the District of Columbia) had compulsory schooling; and is somewhere in the middle in the Midwest, where 10 of 12 states had introduced compulsory schooling by 1900. These results are consistent with Solon (2004), who shows that intergenerational income elasticity decreases with the progressivity of public investment in human capital. Intuitively, government investment in public education crowds out some private investment (from mostly high income families), and at the same time allows some low-income children

¹⁷Source: our own calculations from IPUMS data.

to acquire human capital and raise their labor market earnings.

7.4 Returns to Human Capital

What additional factors might contribute to explain the trends in father/son and father/son-in-law elasticity after accounting for regional differences in economic development and geographic mobility?

One possible explanation for the overall increase in the father/son elasticity is the improvement in men's labor market outcomes during this period. In the context of our model, this can be thought of as an increase in the labor market return to human capital. Cverk (2012) shows that men's career prospects, measured by occupational upgrading, improved substantially between 1880 and 1930. Katz and Margo (forthcoming) document a substantial increase in the share of white collar jobs in the overall economy between 1850 and 1910 (from 6.9% to 19.7%), and a contemporaneous upward trend in relative wages of white collar workers relative to common laborers and artisans. They argue that these trends combined reveal an increasing relative demand for more educated workers over the course of the 19th Century. Margo (2000) provides further evidence of a long-term rise in the returns to educated labor beginning before the Civil War and continuing until the turn of the 20th Century.¹⁸

Changes in labor market returns to human capital can also explain the trend in father/son-in-law elasticity. Equation (4) in our model can be interpreted as the reduced form of a matching tournament model of marriage with pre-marital investment that has stochastic returns (Bhaskar and Hopkins, 2012). In this framework the rate of return to female human capital is determined endogenously as a function of male returns to human capital and marriage market conditions. In a society where women do not work, the incentive to invest for girls increases with the labor market returns of boys. Thus the improvement in men's labor market outcomes would be consistent with our finding that the father-son and father-daughter elasticity share a common trend over the period of interest.

However, there are periods where the two elasticities diverge. For example, the father/son-in-law elasticity is greater than the father/son elasticity between 1880 and 1920 and then dips below it in 1930. These differences may be driven by changes in the sex ratio (defined as the ratio of men to women), which affects the relative position of women in the marriage market. As women become scarce, even lowest quality women become desirable and can fetch a high quality mate. This would push the return to female human capital down. On

¹⁸This 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 Villafior (1987) for an in-depth analysis of wage growth between 1820 and 1865.

the other hand, the increased competition on the male side of the market leads to male over-investment in human capital and, as a result, an increase in the variance of the quality of potential husbands. This, in turn, increase women's incentives to invest in human capital, pushing up the returns.¹⁹ A similar but opposite argument holds if there is a decline in the sex ratio.

It follows that trends in fecundity and immigration over the sample period that affected the sex ratio may help explain differences in the evolution of the father-son and father-daughter elasticity. Differential fecundity by gender implies that marriageable women are scarce and this affects their relative power in the marriage market.²⁰ The scarcity of fecund women is especially important when infant and maternal mortality are high and people have more children. Both infant mortality and fertility were very high in 1850 but plummeted by the early decades of the 20th Century (Haines, 2008).²¹ Maternal mortality declined from 850 deaths per 100,000 births in 1900 to 660 by 1917 (Loudon, 1992). These developments would lead to an increase in the number of eligible women and thus to a decline in the sex ratio, and, consequently, a higher return to female investment.

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

8 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 devised an empirical strategy 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 broadly follows the same trend, with some differences in timing,

¹⁹Bhaskar and Hopkins (2012) show that the net effect on female returns to human capital is negative. The sex ratio imbalance induces a greater investment by the abundant sex.

²⁰See Siow (1998) and Bergstrom and Bagnoli (1993) for alternative mechanisms underlying this relation.

²¹The infant mortality rate was 216.8 per 100,000 births in 1850, 110.8 in 1900 and 60 in 1930.

and drops below the father/son elasticity at the end of the sample period. An exploration of historical demographic and economic trends reveals that divergence in the degree of economic development across US regions, coupled with a decrease in geographic mobility, can account for almost all of the changes in the degree of intergenerational mobility over time. This is not to say that there were no changes in intergenerational mobility over the sample period: on the contrary, the US became a less mobile society in part because there was less geographic mobility, and in part because the different regions were experiencing a different growth trajectory.

References

- [1] Aaronson, Daniel and Bhashkar Mazumder (2008). “Intergenerational Economic Mobility in the United States, 1940 to 2000,” *Journal of Human Resources*, 43(1), pp. 139-172.
- [2] Abramitzky, Ran; Leah Platt Boustan and Katherine Eriksson (2012). “Europe’s Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration,” *American Economic Review*, 102(5), pp. 1832-1856.
- [3] Angrist, Joshua D. and Alan B. Krueger (1992). “The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples,” *Journal of the American Statistical Association*, 87(418), pp. 328-36.
- [4] 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.
- [5] Bandiera, Oriana; Imran Rasul and Martina Viarengo (2012). “The Making of Modern America: Migratory Flows in the Age of Mass Migration,” *Journal of Development Economics*, <http://dx.doi.org/10.1016/j.jdeveco.2012.11.005>.
- [6] Becker, Gary S. and Nigel Tomes (1979). “An equilibrium theory of the distribution of income and intergenerational mobility,” *Journal of Political Economy*, 87(6), pp. 1153-89.
- [7] Becker, Gary S. and Nigel Tomes (1986). “Human Capital and the Rise and Fall of Families,” *Journal of Labor Economics*, 4(3), pp. S1-39.

- [8] Bergstrom, Ted and Mark Bagnoli (1993). "Courtship as a Waiting Game," *Journal of Political Economy*, 101(1), pp. 185-202.
- [9] 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.
- [10] Bhaskar, V. and Ed Hopkins (2012) "Marriage as Rat Race: Noisy Pre-Marital Investments with Assortative Matching," manuscript, University College London.
- [11] Björklund, Anders and Markus Jäntti (1997). "Intergenerational Income Mobility in Sweden Compared to the United States." *American Economic Review*, 87(5), pp. 1009-1018.
- [12] Black, Sandra E. and Paul J. Devereux (2011). "Recent Developments in Intergenerational Mobility." in (O. Ashenfelter and D. Card, eds.), *Handbook of Labor Economics*, Volume 4B, pp. 1487-1541, Amsterdam: North-Holland.
- [13] Browning, Martin; Angus Deaton and Margaret Irish (1985). "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle," *Econometrica*, 53(3), pp. 503-544.
- [14] Chadwick, Laura and Gary Solon (2002). "Intergenerational Income Mobility among Daughters," *American Economic Review*, vol. 92(1), pp. 335-44.
- [15] Clark, Gregory (2013) "What is the True Rate of Social Mobility? Evidence from the Information Content of Surnames," manuscript, UC Davis.
- [16] Clark, Gregory and Neil Cummins (2012) "What is the True Rate of Social Mobility? Surnames and Social Mobility in England, 1800-2012," manuscript, UC Davis.
- [17] Collado, M. Dolores; Ignacio Ortuño Ortín and Andrés Romeu (2012). "Long-run Intergenerational Social Mobility and the Distribution of Surnames," manuscript, Universidad de Alicante.
- [18] Cook, Lisa D.; Trevon D. Logan and John M. Parman (2012) "The Long-Term Consequences of Distinctively Black Names: Evidence from the American Past," manuscript, Michigan State University.

- [19] Corak, Miles (2013). “Inequality from Generation to Generation,” in (R. Rycroft, ed.) *The Economics of Inequality, Poverty, and Discrimination in the 21st Century*, ABC-CLIO.
- [20] Cvrcek, Tomas (2012) “America’s Settling Down: How Better Jobs and Falling Immigration led to a Rise in Marriage, 1880-1930” *Explorations in Economic History*, 49(3), pp. 335-351.
- [21] Duncan, Otis D. (1966). “Methodological Issues in the Analysis of Social Mobility,” in (N. J. Smelser, S. M. Lipset, eds.) *Social Structure and Mobility in Economic Development*, pp. 51-97, Chicago: Aldine.
- [22] Erikson, Robert and John H. Goldthorpe (1992). *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon Press.
- [23] Ermisch, John; Marco Francesconi and Thomas Siedler (2006). “Intergenerational Mobility and Marital Sorting,” *Economic Journal*, 116(513), pp. 659-679.
- [24] Ferrrie Joseph P. (1995). “A New Sample of Americans Linked from the 1850 Public Use Micro Sample of the Federal Census of Population to the 1860 Federal Census Manuscript Schedules,” NBER Historical Working Paper No. 71.
- [25] 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.
- [26] Ferrie, Joseph P. and Karen Rolf (2008). “The May-December Relationship since 1850: Age Homogamy in the United States”, manuscript, Northwestern University.
- [27] Fitch, Catherine A. and Steven Ruggles (2000). “Historical trends in marriage formation: The United States 1850-1990,” in (L.J. Waite, C. Bachrach, M. Hindin, E. Thomson and A.Thornton, eds.) *The Ties that Bind: Perspectives on Marriage and Cohabitation*, pp. 59-88, Hawthorne: Aldine de Gruyter.
- [28] 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.
- [29] Goldin, Claudia (1999). “Egalitarianism and the Returns to Education during the Great Transformation of American Education,” *Journal of Political Economy* 107 (December): S65-S94.

- [30] Goldin, Claudia and Lawrence Katz (2008). *The Race between Education and Technology*. Cambridge, MA: The Belknap Press of Harvard University Press.
- [31] Goldin, Claudia and Robert A. Margo (1992). “The Great Compression: The Wage Structure in the United States at Mid-century,” *Quarterly Journal of Economics*, 107 (1), pp. 1-34.
- [32] Güell, Maia; José V. Rodríguez-Mora and Chris Telmer (2007). “Intergenerational Mobility and the Informative Content of Surnames,” manuscript, University of Edinburgh.
- [33] Haan, Michael (2005). “Studying the Impact of Religion on Fertility in Nineteenth-Century Canada: The Use of Direct Measures and Proxy Variables,” *Social Science History*, 29 (3), pp. 373-411.
- [34] Hacker, J. David (1999). “Child Naming, Religion, and the Decline of Marital Fertility in Nineteenth-Century America,” *The History of the Family: An International Quarterly*, 4 (3): pp. 339-65.
- [35] Haines, Michael R. (1996). “Long-Term Marriage Paterns in the United States from Colonial Times to the Present,” *The History of the Family*, 1 (1): pp. 15–39.
- [36] Haines, Michael (2008). “Fertility and Mortality in the United States,” in (R. Whaples ed.) EH.Net Encyclopedia, <http://eh.net/encyclopedia/article/haines.demography>.
- [37] Head, Keith and Thierry Mayer (2008). “Detection Of Local Interactions From The Spatial Pattern Of Names In France,” *Journal of Regional Science*, 48(1): pp. 67-95.
- [38] Hellerstein, Judith K. and Melinda S. Morrill (2011). “Dads and Daughters: The Changing Impact of Fathers on Women’s Occupational Choices,” *Journal of Human Resources*, 46(2): pp. 333-372.
- [39] Inoue, Atsushi and Gary Solon (2010). “Two-sample Instrumental Variables Estimators,” *Review of Economics and Statistics*, 92(3), pp. 557-561.
- [40] Jäntti, Markus; Brent Bratsberg, Røed Knut; Oddbjørn Raaum; Robin A. Naylor; Eva Österbacka; Anders Björklund and Tor Erkiison (2006). “American Exceptionalism in a New Light: A Comparison of Intergenerational Earnings Mobility in the Nordic Countries, the United Kingdom and the United States,” Discussion Paper no. 1938, Institute for the Study of Labor (IZA), Bonn.

- [41] Jones, Larry E. and Michle Tertilt (2008). “An Economic History of Fertility in the United States: 1826-1960,” in (Peter Rupert ed.) *Frontiers of Family Economics*, Volume 1, pp. 165-230, Emerald Press.
- [42] Katz, Lawrence F. and Robert A. Margo. *Forthcoming*. “Technical Change and the Relative Demand for Skilled Labor: The United States in Historical Perspective,” in (L. Boustan, C. Frydman, and R. A. Margo eds.) *Human Capital in History: The American Record*, Cambridge, MA: National Bureau of Economic Research Conference Volume.
- [43] Kim, Sukko (1998). “Economic Integration and Convergence: U.S. Regions, 1840-1990,” *Journal of Economic History*, 58 (3): pp. 659-683.
- [44] Kim, Sukko and Robert A. Margo (2004). “Historical Perspectives on U.S. Economic Geography,” in (V. Henderson and J. Thisse, eds.) *Handbook of Regional and Urban Economics*, Volume 4: Cities and Geography, pp. 2982-3019, Amsterdam: North-Holland.
- [45] 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*, 101(4), pp. 710-40.
- [46] Lam, David A. and Robert F. Schoeni (1994). “Family ties and labor market in the United States and Brazil,” *Journal of Human Resources*, 29(4), pp. 1235-58.
- [47] Lingwall, Jeff (2010). “Compulsory Schooling, the Family, and the ‘Foreign Element’ in the United States, 1880-1900,” manuscript, Carnegie Mellon University, <http://www.heinz.cmu.edu/research/372full.pdf>.
- [48] 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.
- [49] Long, Jason and Joseph P. Ferrie (forthcoming). “Intergenerational Occupational Mobility in Britain and the U.S. Since 1850,” *American Economic Review*.
- [50] Loudon, Irvine (1992). *Death in Childbirth: An International Study of Maternal Care and Maternal Mortality 1800-1950*, Oxford: Clarendon Press.
- [51] 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.

- [52] Robert A. Margo (2000). “The History of Wage Inequality in America, 1920 to 1970,” *Macroeconomics* 0004035, EconWPA.
- [53] 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.
- [54] Murphy, Kevin M. and Robert H. Topel (1985). “Estimation and Inference in Two-Step Econometric Models,” *Journal of Business and Economic Statistics*, 3(4), pp. 370-379.
- [55] Preston, Samuel H. and Michael R. Haines (1991). *Fatal Years: Child Mortality in Late Nineteenth Century America*, Princeton, NJ: Princeton University Press.
- [56] Raaum, Oddbjørn; Bernt Bratsberg, Knut Røed, Eva Österbacka, Tor Eriksson, Markus Jäntti and Robin A. Naylor (2007). “Marital Sorting, Household Labor Supply, and Intergenerational Earnings Mobility across Countries,” *The B.E. Journal of Economic Analysis & Policy*, Vol. 7: Iss. 2 (Advances), Article 7.
- [57] Ruggles, Steven; J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek (2010). *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database], Minneapolis: University of Minnesota.
- [58] Siow, Aloysius (1998). “Differential Fecundity, Markets, and Gender Roles,” *Journal of Political Economy*, vol. 106(2), pp. 334-354.
- [59] Solon, Gary R. (1992). “Intergenerational Income Mobility in the United States,” *American Economic Review*, 82 (3), pp. 393-408.
- [60] Solon, Gary R. (1999). “Intergenerational Mobility in the Labour Market,” in (O. Ashenfelter and D. Card, eds.) *Handbook of Labor Economics*, Volume 3A, pp. 1761-800, Amsterdam: North-Holland.
- [61] Solon, Gary R. (2004). “A Model of Intergenerational Mobility Variation over Time and Place,” in (M. Corak, ed.) *Generational Income Mobility in North America and Europe*, Cambridge: Cambridge University Press.
- [62] Sutch, Richard (2011). “Hard Work, Nonemployment, and the Wealth-Age Profile: Evidence of a Life-Cycle Strategy in the United States during the Nineteenth Century,” manuscript presented at NBER Development of the American Economy, Program Meeting, <http://users.nber.org/confer/2011/DAEs11/Sutch.pdf>.

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

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 of 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 Appendix Table 3 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 the next row, we deal with the potential misspelling of names by using the Soundex algorithm.²² 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 tends to reduce the attenuation bias and raise the estimate of the elasticity. On 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.

The trends in father-son-elasticities are not much affected by the different name coding schemes. The point estimates in row 2 and 3 are quite similar to the benchmark, while based on the Soundex algorithm the father-son elasticity are higher by 6 to 13 percentage points. As for the father-son-in-law elasticities in all specifications there is less evidence of an upward trend. This is driven by fluctuations in the point estimates in the first part of the sample while all specifications exhibit a sharp decline in the elasticity in 1930.

²²The 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

We generate a population of N families. We generate incomes for each family for T periods based on the income transmission process described in equations (5) and (6). 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.²³

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 $\{1, 2, \dots, J\}$ is given by:

$$P(j|e_{t-1}) = \frac{\exp(\delta_{CON,j} + \delta_{SES,j}e_{t-1})}{\sum_{j'=1}^J \exp(\delta_{CON,j'} + \delta_{SES,j'}e_{t-1})} \quad (7)$$

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.

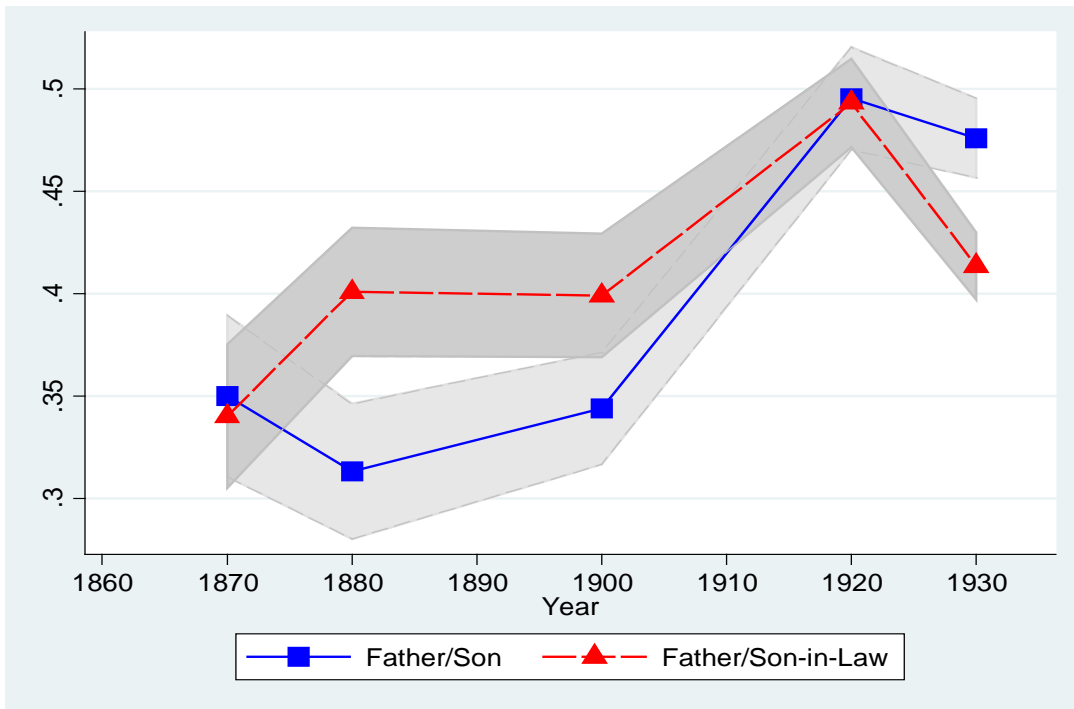
We use these two samples to calculate the six simulated moments that will be matched to the data. Starting from an initial guess for the parameter vector $\theta = (\gamma_1, \lambda, \sigma_u^2, \sigma_v^2, \sigma_{CON}^2, \sigma_{SES}^2)$ we replicate this process $R = 15$ times, and compute the simulated 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.²⁴ 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 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 $p = 0.09$.

²³In practice, preliminary simulations showed that $T = 14$ was sufficient to guarantee that the initial conditions had no effect on the distributions.

²⁴Ideally, 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.

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



Note: The figure presents point estimates and 90% confidence intervals for the father/son and father-son-in-law intergenerational elasticities. The values on the horizontal axes represent the year from which the son's (son-in-law's) sample are drawn. The elasticities are obtained from a regression of son (son-in-law) log occupational income on imputed father's (father-in-law's) log occupational income. See text for details of the imputation procedure. Occupational income is based on average earnings in the occupation in 1950.

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	<i>Males</i>							
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
	<i>Females</i>							
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

Note: Column (7) shows the share of children that have one of the 50 most popular names, by gender. Column (8) shows the R^2 from a regression of father's log occupational income on a full set of name dummies. Unless noted otherwise, the source for this and all following Tables are the 1850 to 1930 Integrated Public Use Micro Samples of the US decennial population censuses (Ruggles et al., 2010).

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

	1850	1860	1870	1880	1900	1910	1920	1930
Males								
Rank:	<i>Most Prestigious</i>							
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
	<i>Least Prestigious</i>							
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:	<i>Most Prestigious</i>							
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
	<i>Least Prestigious</i>							
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
	<div style="display: flex; justify-content: space-between; margin-bottom: 5px;"> Exact name, nickname or alternative spelling appears more than once (most prestigious). Exact name, nickname or alternative spelling appears more than once (least prestigious). </div>							

Notes: Entries in the table represent the five children names with the highest and lowest average father occupational score, by gender and Census year. Only names that appear at least 20 times are considered for the ranking.

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

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

Table 4. Intergenerational Elasticities 1850-1930.
1900 Income Distribution and Farmers' Income.

	(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 (0.0239)	0.3133 (0.0200)	0.3440 (0.0166)	0.4953 (0.0152)	0.4760 (0.0118)
1900	0.3502 (0.0222)	0.3542 (0.0189)	0.3823 (0.0155)	0.4471 (0.0121)	0.4436 (0.0101)
1900, imputed farmer wage	0.3467 (0.0284)	0.2879 (0.0229)	0.3634 (0.0196)	0.4660 (0.0150)	0.4701 (0.0127)
1950 excluding farmers	0.1899 (0.0476)	0.1561 (0.0359)	0.1463 (0.0280)	0.2540 (0.0322)	0.2922 (0.0277)
1900 excluding farmers	0.2487 (0.0460)	0.2075 (0.0374)	0.2320 (0.0329)	0.2992 (0.0312)	0.2954 (0.0259)
1950 ex. farmers (linked sample)		0.2860 (0.0495)	0.3266 (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 (0.0213)	0.4009 (0.0191)	0.3992 (0.0183)	0.4932 (0.0131)	0.4136 (0.0100)
1900	0.3115 (0.0203)	0.4229 (0.0192)	0.4120 (0.0182)	0.4900 (0.0126)	0.4387 (0.0100)
1900, imputed farmer wage	0.2509 (0.0242)	0.3161 (0.0205)	0.3166 (0.0208)	0.4415 (0.0146)	0.4221 (0.0120)
1950 excluding Farmers	0.2150 (0.0465)	0.2003 (0.0303)	0.1802 (0.0284)	0.3270 (0.0288)	0.3220 (0.0227)
1900 excluding Farmers	0.1986 (0.0403)	0.2290 (0.0316)	0.2224 (0.0297)	0.3490 (0.0289)	0.3744 (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]				

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

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

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

Notes: Entries in the table represent OLS coefficients from a regression of son's (son-in-law's) log occupational income on imputed father's (father-in-law's) log occupational income. Different rows use different measures of occupational income. ERSCOR50 reports the percentage of persons in occupations having lower standardized median earnings than the respondent's occupation. SEI is a constructed measure that assigns a Duncan Socioeconomic Index (SEI) score to each occupation using the 1950 occupational classification. Standard errors in parentheses. The number of observations used in each regression, and the number of distinct first names used to impute father's (father-in-law's) income are reported in brackets at the bottom of each panel.

Table 6. 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 (0.0239)	0.3523 (0.0240)	0.3133 (0.0200)	0.3307 (0.0199)	0.3440 (0.0166)	0.3466 (0.0164)	0.4953 (0.0152)	0.4855 (0.0151)	0.4760 (0.0118)	0.4605 (0.0117)
Father's age		0.0096 (0.0093)		0.0009 (0.0080)		0.0289 (0.0060)		0.0196 (0.0055)		0.0183 (0.0043)
Father's age squared		-0.0001 (0.0001)		-0.0001 (0.0001)		-0.0004 (0.0001)		-0.0002 (0.0001)		-0.0002 (0.0001)
Son's age		0.1075 (0.0069)		0.0879 (0.0058)		0.1014 (0.0048)		0.0907 (0.0044)		0.1174 (0.0039)
Son's age squared		-0.0017 (0.0001)		-0.0013 (0.0001)		-0.0015 (0.0001)		-0.0014 (0.0001)		-0.0018 (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 (0.0213)	0.3330 (0.0219)	0.4009 (0.0191)	0.3873 (0.0192)	0.3992 (0.0183)	0.3987 (0.0183)	0.4932 (0.0131)	0.4869 (0.0134)	0.4136 (0.0100)	0.4077 (0.0102)
Father's age		0.0062 (0.0100)		0.0106 (0.0085)		0.0016 (0.0073)		0.0093 (0.0059)		0.0046 (0.0040)
Father's age squared		-0.0001 (0.0001)		-0.0002 (0.0001)		-0.0000 (0.0001)		-0.0001 (0.0001)		-0.0001 (0.0000)
Son's age		0.0447 (0.0029)		0.0328 (0.0020)		0.0282 (0.0018)		0.0179 (0.0013)		0.0249 (0.0013)
Son's age squared		-0.0006 (0.0000)		-0.0004 (0.0000)		-0.0004 (0.0000)		-0.0002 (0.0000)		-0.0003 (0.0000)
N, no. of names	[23280, 976]		[30081, 1376]		[45804, 2063]		[68439, 2888]		[79314, 3326]	

Notes: Entries in the table represent OLS coefficients from a regression of son's (son-in-law's) log occupational income on imputed father's (father-in-law's) log occupational income. Son's age is taken from the individual level data. Father's age is imputed using the average father's age by son's first name in the earlier sample. Standard errors in parentheses. The number of observations used in each regression, and the number of distinct first names used to impute father's (father-in-law's) income are reported in brackets at the bottom of each panel.

Table 7. 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			
$Cov(y_t, \tilde{y}_{j,t-1}) / V(\tilde{y}_{j,t-1})$	0.314	0.313	1860 and 1880 1% samples			
$V(\tilde{y}_{j,t-1})$	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

Notes: Entries in the top part of the table represent the data moments and the simulated moments at the distance-minimizing parameters reported in the bottom part of the table. The parameter estimates are obtained using equal-weighted minimum distance. See text for details of the simulation.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Concentration 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	$\eta=0.0345$ [share50= 0.3444] ($R^2=0.1078$)	0.1131 [0.344] (0.1139)	0.2301 [0.3437] (0.1269)	0.3107 [0.3452] (0.1421)	0.3735 [0.3468] (0.1592)	0.4343 [0.3542] (0.1897)	0.4662 [0.3651] (0.209)
5	0.0275 [0.5526] (0.0894)	0.1073 [0.5524] (0.0967)	0.2203 [0.5521] (0.1084)	0.3087 [0.5517] (0.1232)	0.3757 [0.552] (0.1406)	0.4385 [0.5542] (0.1718)	0.4616 [0.5584] (0.1901)
7.833	0.0139 [0.6976] (0.0713)	0.1160 [0.6965] (0.0774)	0.2246 [0.6958] (0.0898)	0.3144 [0.6952] (0.1053)	0.3794 [0.6949] (0.1215)	0.4494 [0.6947] (0.1519)	0.4746 [0.6972] (0.1716)
10	0.0146 [0.7638] (0.0605)	0.1169 [0.7638] (0.0666)	0.2324 [0.7636] (0.0774)	0.3148 [0.7623] (0.0922)	0.3890 [0.7615] (0.1098)	0.457 [0.7609] (0.138)	0.48 [0.761] (0.1596)
15	0.0122 [0.8444] (0.0441)	0.1209 [0.8447] (0.0498)	0.2419 [0.8438] (0.0599)	0.3385 [0.8428] (0.0736)	0.4009 [0.842] (0.09)	0.4703 [0.8408] (0.1191)	0.4892 [0.8396] (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 intergenerational 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 parameters γ_1 , λ , σ^2_u and σ^2_v are set to the minimizing values listed in the bottom panel of Table 7. 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 Estimated 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	$\eta=0.0502$ [share50=0.6953] ($R^2=0.105$)	0.1070 [0.6952] (0.1057)	0.1543 [0.6952] (0.1081)	0.2239 [0.695] (0.1109)	0.2931 [0.6956] (0.1168)	0.3763 [0.6948] (0.1268)
0.2	0.1024 [0.6953] (0.1039)	0.1591 [0.6952] (0.1053)	0.2080 [0.6952] (0.1081)	0.2796 [0.695] (0.1115)	0.3496 [0.6956] (0.1182)	0.4340 [0.6948] (0.1292)
0.3	0.1518 [0.6953] (0.1022)	0.2084 [0.6952] (0.1041)	0.2591 [0.6952] (0.1074)	0.3330 [0.695] (0.1113)	0.4039 [0.6956] (0.1186)	0.4897 [0.6948] (0.1305)
0.421	0.2049 [0.6953] (0.0992)	0.2613 [0.6952] (0.1016)	0.3144 [0.6952] (0.1053)	0.3915 [0.695] (0.1097)	0.4641 [0.6956] (0.1175)	0.5522 [0.6948] (0.1303)
0.5	0.2331 [0.6953] (0.0967)	0.2892 [0.6952] (0.0993)	0.3438 [0.6952] (0.1031)	0.4233 [0.695] (0.1077)	0.4975 [0.6956] (0.1156)	0.5878 [0.6948] (0.1289)
0.6	0.2582 [0.6953] (0.0928)	0.3137 [0.6952] (0.0956)	0.3701 [0.6952] (0.0994)	0.4526 [0.695] (0.104)	0.5291 [0.6956] (0.1118)	0.6236 [0.6948] (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 parameters σ^2_{SES} , σ^2_{CON} , σ^2_u and σ^2_v are set to the minimizing values listed in the bottom panel of Table 7. 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 (0.0239)	0.3133 (0.0200)	0.3440 (0.0166)	0.4953 (0.0152)	0.4760 (0.0118)
Control for number of siblings	0.2836 (0.0255)	0.2735 (0.0214)	0.3444 (0.0168)	0.5024 (0.0157)	0.4740 (0.0121)
Control for birth order	0.3277 (0.0247)	0.2860 (0.0207)	0.3433 (0.0166)	0.4974 (0.0154)	0.4642 (0.0119)
N, no. names (baseline)	[37077, 1182]	[50847, 1478]	[80255, 2234]	[109079, 3253]	[122468, 3720]
B: Fathers-Sons in Law					
Baseline	0.3402 (0.0213)	0.4009 (0.0191)	0.3992 (0.0183)	0.4932 (0.0131)	0.4136 (0.0100)
Control for number of siblings	0.2920 (0.0239)	0.3044 (0.0210)	0.3949 (0.0190)	0.4651 (0.0140)	0.3815 (0.0109)
Control for birth order	0.3289 (0.0215)	0.3659 (0.0197)	0.3962 (0.0184)	0.4734 (0.0133)	0.3951 (0.0104)
N, no. names (baseline)	[23280, 976]	[30081, 1376]	[45804, 2063]	[68439, 2888]	[79314, 3326]

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

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

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

Table 12. Differences in Economic Development across U.S. States

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

Notes: Entries in the table represent OLS coefficients from a regression of son's (son-in-law's) log occupational income on imputed father's (father-in-law's) log occupational income. Percent urban and percent agriculture are the percentage of the whole state population who are urban and employed in agriculture, respectively. Standard errors in parentheses. The number of observations used in each regression, and the number of distinct first names used to impute father's (father-in-law's) income are reported in brackets at the bottom of each panel.

Table 13. Intergenerational Elasticities 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 (0.0239)	0.3133 (0.0200)	0.3440 (0.0166)	0.4953 (0.0152)	0.4760 (0.0118)
Northeast	0.2948 (0.0383)	0.2539 (0.0337)	0.1677 (0.0310)	0.2187 (0.0279)	0.1918 (0.0224)
Midwest	0.1499 (0.0468)	0.2521 (0.0368)	0.2677 (0.0315)	0.2771 (0.0279)	0.2701 (0.0230)
South	0.4593 (0.0564)	0.1591 (0.0337)	0.2878 (0.0311)	0.3081 (0.0293)	0.3641 (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 (0.0213)	0.4009 (0.0191)	0.3992 (0.0183)	0.4932 (0.0131)	0.4136 (0.0100)
Northeast	0.2014 (0.0380)	0.2221 (0.0382)	0.3111 (0.0409)	0.2743 (0.0333)	0.2100 (0.0261)
Midwest	0.3471 (0.0520)	0.3811 (0.0353)	0.3289 (0.0337)	0.3371 (0.0238)	0.3015 (0.0183)
South	0.3975 (0.0478)	0.3303 (0.0286)	0.3192 (0.0306)	0.4649 (0.0252)	0.3791 (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]

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

Appendix Table 1. Calculation of 1900 Wage Assigned to Farmers

Name of 1910 occupation assigned 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

Notes: The table records all the occupations in the 1910 Census that were coded as farmers in the 1950 occupational classification scheme, and the fraction of each occupation out of the total. Column (2) shows the weights used to calculate the default average wage for farmers. This is the average wage assigned to all farmers in Table 4, second row of each panel. Column (3) records the income data for specialized farming occupations from Preston and Haines (1991). See text for details on the calculation of income for owner-occupier farmers and farm tenants.

Appendix Table 2: 30-year elasticities

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

Notes: Entries in the table represent OLS coefficients from a regression of son's (son-in-law's) log occupational income on imputed father's (father-in-law's) log occupational income. Standard errors in parentheses. The number of observations used in each regression, and the number of distinct first names used to impute father's (father-in-law's) income are reported in brackets at the bottom of each panel.

Table A3. Sensitivity to Different Name Coding Schemes.

	(1)	(2)	(3)	(4)	(5)
	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930
A: Fathers-Sons					
Name concept:					
All	0.3500 (0.0239)	0.3133 (0.0200)	0.3440 (0.0166)	0.4953 (0.0152)	0.4760 (0.0118)
Middle initials	0.3400 (0.0230)	0.3112 (0.0191)	0.3291 (0.0156)	0.4189 (0.0136)	0.4389 (0.0111)
Nicknames	0.3673 (0.0246)	0.3310 (0.0207)	0.3412 (0.0176)	0.4489 (0.0159)	0.4268 (0.0123)
Soundex codes	0.4212 (0.0304)	0.4041 (0.0250)	0.4771 (0.0223)	0.5571 (0.0184)	0.5530 (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 (0.0213)	0.4009 (0.0191)	0.3992 (0.0183)	0.4932 (0.0131)	0.4136 (0.0100)
Middle initials	0.3441 (0.0208)	0.3619 (0.0179)	0.3771 (0.0170)	0.4249 (0.0122)	0.3834 (0.0096)
Nicknames	0.4360 (0.0258)	0.4152 (0.0204)	0.4135 (0.0189)	0.4551 (0.0140)	0.3882 (0.0107)
Soundex codes	0.5907 (0.0305)	0.5543 (0.0257)	0.5570 (0.0256)	0.6122 (0.0176)	0.4944 (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]

Notes: Entries in the table represent OLS coefficients from a regression of son's (son-in-law's) log occupational income on imputed father's (father-in-law's) log occupational income. Standard errors in parentheses. The "middle initials" row groups separately names with and without middle initials: "John" is a separate category from "John M.," but "John M." and "John H." belong to the same category. The "nicknames" row groups together common nicknames associated to a given first name: "Johnny" is the same as "John". The "Soundex codes" row groups together all names that have the same Soundex code. At the bottom of each panel, the number of observations used in each regression, and the number of distinct first names used to impute father's (father-in-law's) income.