

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

Claudia Olivetti and M. Daniele Paserman
Boston University and NBER

April 1, 2010

Abstract

This paper constructs a continuous and consistent measure of intergenerational mobility in the United States between 1850 and 1930 by linking individuals with the same first name across pairs of decennial Censuses. One of the advantages of this methodology is that it allows to calculate intergenerational correlations not only between fathers and sons, but also between fathers-in-law and sons-in-law, something that is typically not possible with historical data. Thus, the paper sheds light on the role of marriage in the intergenerational transmission of economic status from a historical perspective.

We find that the father-son correlation in economic status grows throughout the period, but is consistently lower than the correlation between fathers-in-law and sons-in-law. The gap declines over time, and seems to have closed by the end of the period. We present a simple model of investment in human capital, marital sorting and intergenerational mobility that can rationalize the findings.

*VERY VERY PRELIMINARY DRAFT. We thank Dara Lee and Laura Salisbury for excellent research assistance, and Robert Margo for many valuable comments and conversations.

1 Introduction

This paper makes three contributions: first, it constructs a consistent and continuous measure of intergenerational mobility in the US between 1850 and 1930. Second, it sheds light on the role of marriage in the intergenerational transmission of earning status and in the propagation of inequality across generations. Third, it develops a new methodology for estimating intergenerational correlations that can be used with widely available public use data.

Our estimator is based on a simple idea: in any given year, the set of individuals in the US Census Public Use Microdata Samples (PUMS) with a given first name is a random sample of the population with that first name. Hence, intergenerational correlations can be calculated as the correlation between the average earnings/socioeconomic status of individuals with a given first name in year t and the average earnings/socioeconomic status of the fathers of individuals with that name in year $t - k$. We formulate a simple econometric model of intergenerational earnings transmission, and show that our estimator is informative about the underlying structural parameters. Among the advantages of this methodology is that it can be readily applied to the correlation between the earnings/socioeconomic status of fathers and daughters, and, more interestingly, fathers-in-law and sons-in-law. This would not be at all possible without the availability of linked longitudinal data. Therefore, we can shed light on the role of marriage in the transmission of earning status across generations and in the propagation of inequality.

Empirically, we proceed as follows. First, we use data from the 1850-1930 IPUMS Linked Representative Samples to compare estimates obtained with our methodology to benchmark measures of intergenerational mobility obtained using conventional methods. We find that our methodology yields remarkably similar measures of intergenerational mobility, both in terms of levels and in terms of time trends (see Table 1). Second, we apply our methodology to the full IPUMS samples between 1850 and 1930. This yields a series of five intergenerational correlation coefficients at 20-year intervals, and four intergenerational correlation coefficients at 30-year intervals. These coefficients are calculated both for the correlation between fathers and sons and for the correlation between fathers and sons-in-law.

Our results indicate that: a) the intergenerational elasticity between fathers and sons at 20-year and 30-year intervals increases between 1850 and

1910, with most of the increase occurring after the turn of the century. These results are in accord with the findings of Ferrie (2005; 2007), who documented a marked decrease in intergenerational mobility in the United States between the late 19th century and the middle of the 20th century; b) the relationship in economic status between fathers in law and sons in law does not change substantially over the sample period; c) the father in law-son in law correlation is always higher than the corresponding father-son correlation. The gap declines over time, and almost disappears for the generation of children aged 0-15 in 1910. We discuss the implication of the latter finding for the propagation of income inequality across generations in the context of a simple overlapping-generations model of human capital investment and marriage.

2 Related Literature

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

Intergenerational Mobility: Modern Data. There is an extensive literature that studies intergenerational mobility using modern panel data sets (see Solon, 1999, and references therein, for a comprehensive survey). The estimates of father-son intergenerational mobility obtained for the U.S. range from 0.13 to 0.54 with the median estimate of the elasticity between father's log labor income and son's log labor income hovering around 0.4. Fewer studies provides estimates of the father-daughter correlation. These estimates range from 0.11 to 0.54, with a median estimate of 0.31.¹

Only a very limited number of papers in this literature have studied the correlations between father-in-law and son-in-law. Chadwick and Solon (2002) use PSID data to study intergenerational mobility in the daughter's family income. They find that for modern US data the father-son elasticity - estimated to be equal to 0.523 - tends to be somewhat larger than the father in law - son in law elasticity- estimated at 0.360. Lam and Schoeni (1993, 1994) compute correlations between son's income and the background characteristics (such as education) of father and father-in-law in Brazil. They find that the effect of father-in-law's schooling on wages is larger than the

¹Most recently, Hellerstein and Morrill (2009) study trends in fathers-daughters occupational mobility across cohorts of US women.

effect of father's schooling in Brazil, while the opposite is observed in the United States.

Intergenerational Mobility: Historical Data. The main contributor to the historical literature is Ferrie (1995, 2004) who uses the 1880 U.S. federal Census as well as extracts from the 1850-1910 public use Census samples to construct a large, nationally representative longitudinal data set that allows to study occupational mobility between fathers and sons. Based on this data, Ferrie (2005) shows that in the United States the degree of intergenerational mobility declined markedly between the end of the 19th and the middle of the 20th century ("the end of American exceptionalism"). Using a comparable data set for the United Kingdom, Long and Ferrie (2007) show that although in the late 19th-century intergenerational occupational mobility was higher in the US than in the UK, the two countries converged over time.

Sorting, Intergenerational Mobility and Inequality. Very few papers have studied the link between sorting and intergenerational mobility. One exception is the paper by Ermisch, Francesconi and Siedler (2006) that shows that positive assortative mating can explain 40 to 50 percent of the covariance between parents' and own family income both in Germany and in the UK.

The paper is also related to a large literature in economics and other fields on assortative mating.² Since the pioneering work by Becker (1991) and Lam (1988) a large literature has developed that studies how and why individuals tend to sort by education, income, ethnicity, religion, or nationality. Papers have looked at how sorting affects intergenerational transmission of cultural and religious traits (Bisin and Verdier, 2000); inequality and economic growth (Kremer, 1997; Fernández and Rogerson, 2001; and Fernández, Guner and Knowles, 2005); and the transmission of genetic traits and its effect on inequality and growth (Galor and Moav, 2002).

Names. Finally, the paper is also related to recent papers that have investigated the economic content and consequences of names. Fryer and Levitt (2004) and Bertrand and Mullainathan (2004) study the labor market effects of distinctively black names. Goldin and Shim (2004) analyze the patterns of maiden name retention among married women. Head and Mayer

²See Epstein and Guttman (1984) and the work by Mare (1991), Kalmijn (1994) and Schwartz and Mare (2005) for a comprehensive review of the (non-economic) literature on assortative mating.

(2008) investigate the social transmission of parental preferences through naming patterns.

Closely related to our project is the work by Guell et al. (2007), who use the informative content of family names to study intergenerational mobility in Spain. They develop a model whose endogenous variable is the joint distribution of surnames and income, and explore the relationship between mobility and the informative content of surnames, allowing for assortative mating to be a determinant of both. They find that the degree of mobility in Spain has substantially decreased over time.

There are no papers, at least to our knowledge, that study the role of marriage in the process of intergenerational mobility from a long-run, historical perspective in the US. This is mainly because of two reasons. First, since very few married women worked in the late 19th century and early 20th century (less than 5% of married women worked in 1900; Goldin, 1990), the analysis of long-run trends in father-daughter mobility is essentially infeasible. Second, it is not possible to construct a longitudinal data set of fathers-daughters on the basis of Census data because daughters change their last name upon marriage. This implies that young daughters in the 1880 Census cannot be identified in, say, the 1900 public use sample unless they remained single. The methodology developed in this project helps to fill this gap.

3 Econometric Methodology

3.1 Individual level data, grouped data, and pseudo-panels

To illustrate the econometric methodology, we consider a simple model of intergenerational earnings transmission, as in Becker and Tomes (1986) and Solon (1999). The son's log earnings in family f (y_{ft}) are given by:

$$\begin{aligned} y_{ft} &= \gamma_1 y_{ft-1} + e_{ft} + u_{ft}, \\ e_{ft} &= \lambda e_{ft-1} + v_{ft}; \end{aligned}$$

where $y_{f,t-1}$ are the father's log earnings, and u_{ft} and v_{ft} are idiosyncratic shocks. We can think of the autoregressive component of the error term (e_{ft}) as the family's "endowment," which is transmitted across generations. The idiosyncratic shocks to the income process, u_{ft} and v_{ft} , represent unexpected

innovations to a family’s earning ability and to its endowment, respectively, and can be thought of as “luck.” The parameters γ_1 and λ represent, respectively, the elasticity of son’s earnings with respect to father’s earnings, and the degree of persistence in the endowment process. For the process to be stationary, we require γ_1 and λ to be smaller than 1 in absolute value.

Benchmark: Linked individual-level data. We are interested in the probability limit of the least squares coefficient in a regression of son’s log earnings on father’s log earnings, $\hat{\gamma}_1$. We consider first as a benchmark the traditional case where we have individual level data that is linked across generations. Assuming stationarity, one can show that:

$$p \lim \hat{\gamma}_{1,INDIV} = \gamma_1 + \frac{\lambda(1 - \gamma_1^2)}{(1 + \gamma_1\lambda) + (1 - \gamma_1\lambda)(\sigma_u^2/\sigma_e^2)}. \quad (1)$$

A number of remarks are in order: a) the OLS estimate of the intergenerational elasticity is upward biased if $\lambda > 0$; b) the extent of upward bias is larger the larger is λ and the larger is the variance in the endowment (σ_e^2) relative to the variance of labor market “luck” (σ_u^2). This last point is intuitively reasonable: the larger the variance of the endowment, the more likely it is that any differences in earnings between sons are due to differences in their initial endowment than to differences in investment.

Grouped Data. Now assume that we can observe a fixed individual characteristic, such as a person’s first name, and that we group the data by that characteristic. Specifically, for any variable z , define the group-level average \tilde{z}_j as

$$\tilde{z}_j = \frac{1}{M_j} \sum_{f:N_f=j} z_{fj},$$

where $M_j = \sum_{f: N_f=j} 1$.

Now, still using the linked data, define $\tilde{y}_{j,t-1}$ as the average log earnings of fathers of children named j (taken from the census year in which the children are young), and let \tilde{y}_{jt} be the average log earnings (as adults) of children named j (taken from a census year 20 or 30 years later). We now consider the group-level (or “between”) regression of \tilde{y}_{jt} on $\tilde{y}_{j,t-1}$, and estimate it by weighted least squares, where the weights are given by the name frequencies. The probability limit of the “grouped” regression coefficient is:

$$p \lim \hat{\gamma}_{1,GROUPED} = \gamma_1 + \frac{\lambda(1 - \gamma_1^2)}{(1 + \gamma_1\lambda) + (1 - \gamma_1\lambda)(\sigma_u^2/\sigma_e^2)}. \quad (2)$$

Note the similarity between the probability limit of $\hat{\gamma}_{1,GROUPED}$ and $\hat{\gamma}_{1,INDIV}$. The difference between the grouped estimator and the individual level estimator depends on the size of σ_u^2/σ_e^2 relatively to σ_u^2/σ_e^2 . This in turn depends on how names are distributed in the population. If names are distributed completely randomly in the population, then σ_u^2/σ_e^2 converges in probability to $(\sigma_u^2/\bar{M}) / (\sigma_e^2/\bar{M})$, (where \bar{M} is the average frequency of names in the population), implying that the group-level estimator will have the same probability limit as the individual-level estimator.

However, it is likely that names contain some information about economic status. For example, Bertrand and Mullainathan (2004) document that in a sample of baby names in Massachusetts there is substantial between-name heterogeneity in the social background of mothers; similarly, Levitt and Fryer (2003) show that names provide a strong signal of socioeconomic status for blacks, but also that there are systematic and large differences in name choices by whites with different levels of education. Therefore, even though a person's first name is unlikely to be correlated with the idiosyncratic shock in the outcome variable u , it is reasonable to assume that names are correlated with the endowment e . If this is the case, the variance of \tilde{e} will typically be larger (maybe even substantially so) than the variance of e , with the result that the second term in the probability limit of the group-level estimator will typically be larger relative to its counterpart in the individual-level estimator.

The purpose of this analysis is to show that even the group-level estimator contains relevant information about the parameters of interest.

Pseudo-panels. We now consider the case where we do not have intergenerationally linked data. Instead we have a series of repeated cross sections with information on names, on family relationship, and on the outcome variable. We group the data in the earlier cross-section by children's first names, and the later cross-section by adults' first names, and then merge these two grouped data sets to create a pseudo-panel. Then, let \tilde{y}_{jt} be the average log earnings of individuals named j belonging to the children's generation (obtained from the cross-section in which the children's generation is of working age), and let \tilde{y}_{jt-1} be the average log earnings of the fathers of children named j (obtained from the cross-section in which the children's generation is young and still lives with their parents). If both cross-sections are random samples from the population, then \tilde{y}_{jt} and \tilde{y}_{jt-1} are consistent estimates of the population average log earnings for each respective cohort. The probability limit of the weighted least squares coefficient in a regression

of \tilde{y}_{jt} on \tilde{y}_{jt-1} is:

$$p \lim \hat{\gamma}_{1,PSEUDO} = \frac{[Cov(\tilde{e}_{jt}, \tilde{e}'_{jt})] / \sigma_e^2}{(1 + \gamma_1 \lambda) + (1 - \gamma_1 \lambda) \sigma_u^2 / \sigma_e^2} (\gamma_1 + \lambda). \quad (3)$$

This probability limit depends critically on the covariance between \tilde{e}_{jt} and \tilde{e}'_{jt} , which represent the averages of the endowment by name in two different cross-sections. If names are distributed randomly, these are independent draws and $\hat{\gamma}_{1,PSEUDO}$ converges to zero. However, if the average family effect of individuals with a given first name in one cross-section is correlated with the average family effect of individuals with that first name in a different cross-section, as would be the case if names reflect economic status, then the estimator will be informative about the underlying parameters of interest.

It is convenient to compare directly between the probability limits of the three estimators. Rearranging equations (1), (2) and (3) we obtain:

$$\begin{aligned} p \lim \hat{\gamma}_{1,INDIV} &= \delta \gamma_1 + (1 - \delta) \frac{\gamma_1 + \lambda}{1 + \gamma_1 \lambda}; \\ p \lim \hat{\gamma}_{1,GROUPED} &= \tilde{\delta} \gamma_1 + (1 - \tilde{\delta}) \frac{\gamma_1 + \lambda}{1 + \gamma_1 \lambda}; \end{aligned}$$

and

$$p \lim \hat{\gamma}_{1,PSEUDO} = \frac{Cov(\tilde{e}_{jt}, \tilde{e}'_{jt})}{\sigma_e^2} (1 - \tilde{\delta}) \frac{\gamma_1 + \lambda}{1 + \gamma_1 \lambda},$$

where

$$\begin{aligned} \delta &= \frac{\sigma_u^2 (1 - \gamma_1 \lambda)}{\sigma_u^2 (1 - \gamma_1 \lambda) + \sigma_e^2 (1 + \gamma_1 \lambda)}; \\ \tilde{\delta} &= \frac{\sigma_u^2 (1 - \gamma_1 \lambda)}{\sigma_u^2 (1 - \gamma_1 \lambda) + \sigma_e^2 (1 + \gamma_1 \lambda)}. \end{aligned}$$

The comparison shows that the pseudo-panel estimator will tend to be smaller than the grouped estimator.

3.2 Verifying the methodology: linked samples

In order to verify our methodology we use the IPUMS Linked Representative samples, 1850-1930. This data set links cases from the 1880 census to

1% samples of all other censuses of 1850-1930. The data have been linked so as to maximize representativeness of the linked cases (not sample size). Linkages are more likely for *uncommon* names (first and last). The IPUMS linked datasets contain information on nearly 500,000 people at two points in time, with one observation from the 1880 census and a second observation from one other census between 1850 and 1930. These data allow us to compare intergenerational correlations calculated using our methodology to those using traditional linked datasets.

In our analysis we restrict the sample to households with a white male 0-15 in the older census. In order to verify our methodology we first compute intergenerational correlations on the linked sample. We then compare this correlation to the one calculated based on the pseudo-panel obtained from the linked sample.

For example, if we are interested in comparing the two correlations on the 1850-1880 sample we proceed as follows. First, we use the linked 1850-1880 sample to compute the correlation between outcomes of fathers of children 0-15 in 1850 and childrens adult outcomes at ages 30-45 in 1880. Second, we use grouped data obtained on the basis of this linked sample to compute the pseudo-correlation between fathers and sons over the same period as follows. First, we group all children 0-15 in 1850 by first name, and calculate the average father's outcome by first name of the child. Second, we group all adults 30-45 in 1880 by first name and calculate the average own outcome. Finally, we merge the two data sets and calculate the pseudo-correlation (weighted by name frequency count).

We repeat this procedure for all the year pairs available in the IPUMS Linked Representative samples. The results of this exercise are reported in Table 1. Our methodology produces estimates of the intergenerational correlation that are remarkably similar to those obtained on the linked samples.

4 Data

We use the US Census 1% samples from IPUMS, 1850-1930. The data contain harmonized information on occupational status and predicted earnings. The availability of first names allows the creation of pseudo-panels linked by first name for both men and women. Note that, in contrast to the linked samples, our methodology for creating pseudo-panels is based on most *common* first names and on larger sample sizes. Using these data we can compute five

intergenerational correlation coefficients at 20-year intervals, and four intergenerational correlation coefficients at 30-year intervals for both sons-fathers pairs and sons in law - fathers in law pairs.

Table 2 reports the summary statistics for children's name in the initial year of the pseudo panel.

4.1 Measurement Issues

Two important measurement issues arise in implementing our methodology: how to calculate predicted earnings by occupation in the late 19th century and early 20th century data, and how to correctly match first names across censuses.

Measuring Earnings. One of the main challenges for the computation of historical intergenerational correlations is to obtain appropriate quantitative measures of socioeconomic status. Because income and earnings at the individual level are not available in the US Census until 1940, we are constrained to use measures of socioeconomic status that are based on individuals' occupational status. While this contrasts with the current practice among economists, who prefer to use direct measures of income or earnings if available, there is a long tradition in sociology to focus on occupational categories (Erikson and Goldthorpe, 1992). Note that, in the absence of intergenerationally linked data, we cannot use the approach of Ferrie (2005) and Long and Ferrie (2007), who calculated measures of mobility based on the transition matrices between five very broad occupational categories of fathers and sons.

Instead, our preferred measures of economic status will be the integrated occupational standing variables available in the IPUMS data sets. These variables are based on the IPUMS harmonized occupational coding contained in the OCC1950 variable. This variable contains occupation codes for every year converted into the 1950 occupational scheme.³ Based on this consistent occupational coding, IPUMS has developed a number of alternative numeric measures of occupational standing. For our benchmark analysis, we will use the OCCSCORE variable. This variable indicates the median total income (in hundreds of dollars) of the persons in each occupation in 1950. It is calculated based on total personal income from a published report that contained

³For 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>.

more cases than the public use samples. This variable has a number of useful advantages: first, its calculation is straightforward and is not based on any subjective assessments of an occupation's prestige; second, the measure is largely invariant across census years for a given occupation, apart from some minor variations in the post-1950 years; third, the measure is available continuously for all census years between 1850 and 2000. There are also some shortcomings that one needs to be aware of: mean earnings by occupation are not the same as actual personal income; the measure does not control for changes in relative occupational incomes over time; and the applicability of the 1950 scores to censuses that are far removed in time is open to question. It should also be noted that 1950 was a year of particularly high wage compression, and that wage differentials by geographic area were substantially smaller than at the turn of the century. All this may impact our estimated intergenerational correlations (it would probably bias the result toward overestimating *mobility*).

For this reason we use the 1901 occupational income distribution, based on the Cost of Living Survey, tabulated by Preston-Haines (1991). This measure of occupational income is also problematic. Since the purpose of the Cost of Living Survey was to investigate the cost of living of families in industrial locales of the U.S., the group of families selected represents to a fair degree the urban population of the nation at the time of the interviews. The nature of the survey also implies that farm-related occupations are not included. Moreover, the sample includes families with income not too far from the median. We plan to address some of these limitations in future drafts.

We also compute intergenerational correlations for three additional labor market outcomes. We consider the share of farmers and two alternative IPUMS measures of occupational standing based on the recoded 1950 occupational categories. ERSCOR50 assigns the percentile rank of each occupation's median income based on contemporaneous earnings data. EDSCOR50 indicates the percentage of persons in each occupation with one or more years of college education. Both these variables vary across censuses in the post-1950 period, but are invariant in the pre-1950 period.

Coding of names. A number of issues must be addressed when dealing with first names. First, obvious abbreviations (for example, "Wm"- "William", "Chas"- "Charles", "Geo"- "George") and misspellings must be corrected. Second, one must decide whether to differentiate between individuals who go by first name only ("William") and those who go by first names

with their middle initial (“William H.”). A preliminary examination of the data has shown that individuals who go by first name plus initial tend to have higher occupational standing variables, which suggests that it may be appropriate to treat them as separate categories. Finally, we must also address the issue of nicknames, and decide whether “Bill”, “Will”, “Willy”, “Willie”, should be considered as the same name as “William”. On one hand, one may think that an individual going by “Willie” instead of “William” could be indicative of a different (lower?) socioeconomic standing. On the other hand, it is not unreasonable to think that a person would go by a nickname as a child and use the formal version as an adult.

In our benchmark classification of names middle initials are ignored (that is, William is equal to William J.) and nicknames are treated as distinct names (that is, William is different from Bill).

For robustness purposes, we experiment with two alternative approaches. In the first robustness check we treat those that go by a simple first name as a separate category from those that go by that same first name and any middle initial (i.e., we will treat all the “William” as one category, and group all the “William J.,” “William H.,” etc. in a separate category). Nicknames are still treated as distinct names. In the second robustness check we group all names and nicknames from common root together (William = Bill = William J.).

5 Results

[To be completed]

- Figures 1 and 2, 20-year correlations in mobility. We explore the sensitivity of the results to different measures of intergenerational mobility (intergenerational elasticities versus correlations) and to different weighting schemes (weighting by the father name frequency, the son name frequency, or the geometric mean of the two). Elasticities are somewhat more sensitive to the weighting scheme. Both measures show a clear decrease in father-son mobility. The jump seems to occur around the turn of the 20th century. Father in law-son in law mobility is relatively stable over time, but the pattern differs slightly depending on the measure of mobility (elasticity versus correlation).
- Figure 3 and Table 3, comparison between father-son and father in law-son in law mobility. Father in law-son in law mobility is always

lower. The gap declines over time, and, for the generation of children aged 0-15 in 1910, seems to have almost closed completely (based on the correlation measure) and may have even reversed (based on the elasticity measure). The pattern for 30-year correlations and elasticities is similar.

- Table 4, sensitivity to alternative measures of occupational income. Using the 1900 income distribution (less compressed than 1950) the estimates of intergenerational mobility are roughly similar, and have similar patterns over time. Excluding farmers lowers the estimates substantially, indicating that a large fraction of the correlation is due to the persistence in income of farmers.
- Table 5, sensitivity to alternative measures of labor market outcomes. Some differences in the pattern of elasticities over time, especially in educational score. But, overall, convergence over time between father/son and father in law/son in law elasticities.
- Table 6, sensitivity to alternative name-coding schemes. Results are not sensitive to different coding of names.

6 Interpretation

We derive intergenerational links between son's income and father's income and between son-in-law's income and father-in-law's income using utility-maximizing behavior by parents in the spirit of the model by Becker and Tomes (1979, 1986), Mulligan (1999), and Ermisch, Francesconi and Siedler (2006). This is a simple overlapping generation model where agents live for two periods, the first as children, the second as adults. Agents only make economic decisions in the adult period. All individuals get married when adults and fertility is exogenous. Each family contains two parents and two children: one male, one female. We assume that parents are altruistic towards their children (Barro and Becker, 1989).

Parental utility is defined over own consumption and over the expected family income of their son and daughter when they are adults. 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 affect both earnings on

the labor market and the human capital of spouses on 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.

We assume the following functional form for parental preferences:

$$\beta_1 \log c_{t-1} + \beta_2 E [\log (Y_t^M)] + \beta_3 E [\log (Y_t^F)] \quad (4)$$

where c_{t-1} is parents' consumption and Y_t^M and Y_t^F represent son's and daughter's family income. The parameters β_2 and β_3 measure parents' altruism towards their son and daughter, respectively.

Parents choose c_{t-1} as well as the investment in human capital of their son, H_t^M , and daughter H_t^F , to maximize 4 subject to the budget constraint:

$$c_{t-1} + p_H (H_t^M + H_t^F) \leq y_{t-1}$$

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:

Labor income is a function of an individual's human capital according to the following functional form:

$$y_t = \gamma_0 (H_t)^{\gamma_1} \exp (E_t)$$

where E_t represents the combined effect of all determinants - other than human capital - of a man's lifetime earnings. 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 $\gamma_1 \in (0, 1)$. As in Becker and Tomes (1979, 1986) we decompose E_t in two components:

$$E_t = e_t + u_t$$

where e_t is the child's "endowment" of earning capacity and the i.i.d. stochastic term u_t , with variance σ_u^2 , is the child's "luck" on the labor market assumed to be independent on 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 .

Marriage market

Men and women sorts on the basis of their human capital. We model the matching function using the following functional form:

$$H_t^M = \alpha_0 (H_t^F)^{\alpha_1} \exp(\mu_t)$$

where α_1 measures the rate of return to (female) human capital on the marriage market and the stochastic term μ_t represents other (unobserved) factors in the marriage market such as “love”. We assume that μ_t is i.i.d. with variance σ_μ^2 . 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 on the marriage market according to the matching function, that is we assume that $\alpha_1 > 0$. This matching function can be interpreted as the rational expectation equilibrium outcome of a full-fledged competitive model of the marriage market (see Peters and Siow (2002) for the derivation of the matching function). In this context α_1 measures the degree of competition for males in the marriage market. That is, when $\alpha_1 > 1$ the market is favorable to women and when $\alpha_1 < 1$ the market is favorable to men.

Given this assumptions it follows that $Y_t^M = y_t = \gamma_0 (H_t)^{\gamma_1} \exp(E_t)$ is the son’s family income and $Y_t^F = y_t^{SIL} = \gamma_0 (\alpha_0 (H_t^F)^{\alpha_1} \exp(\mu_t))^{\gamma_1} \exp(E_t^{SIL})$ is the daughter’s family income. Substituting into 4 and taking first order condition we find that the parents optimal investment in human capital are given by:

$$(H_t^M)^* = \frac{\beta_2 \gamma_1}{\gamma_1 (\beta_2 + \beta_3 \alpha_1) + \beta_1} \frac{y_{t-1}}{p_H} \quad (5)$$

$$(H_t^F)^* = \frac{\beta_3 \alpha_1 \gamma_1}{\gamma_1 (\beta_2 + \beta_3 \alpha_1) + \beta_1} \frac{y_{t-1}}{p_H} \quad (6)$$

This implies that in equilibrium:

$$\begin{aligned} \log y_t &= \gamma_0 + \gamma_1 \log y_{t-1} + e_t + u_t \\ \log y_t^{SIL} &= \tilde{\gamma}_0 + \alpha_1 \gamma_1 \log y_{t-1} + e_t^{SIL} + u_t \end{aligned}$$

It follows that the correlation between y_t and y_{t-1} is given by:

$$\rho = \gamma_1 + \frac{\lambda (1 - \gamma_1^2)}{(1 + \gamma_1 \lambda) + (1 - \gamma_1 \lambda) (\sigma_u^2 / \sigma_e^2)},$$

and the correlation between y_t^{SIL} and y_{t-1} is given by:

$$\rho^{SIL} = \alpha_1 \frac{\sigma_y^2}{\sigma_y^2 + \sigma_\mu^2} \rho$$

where $\sigma_y^2 = \frac{\sigma_\varepsilon^2(1+\gamma_1\lambda) + \sigma_u^2(1-\gamma_1\lambda)}{(1-\gamma_1\lambda)(1-\gamma_1^2)}$.

Consequently in this model we can observe $\rho^{SIL} > \rho$ if and only if $\alpha_1 > 1 + \frac{\sigma_\mu^2}{\sigma_y^2}$. This relationship cannot occur if $\alpha_1 \leq 1$, that is if competition in the marriage market is 'favorable' to men, but it can obtain if $\alpha_1 > 1$, that is if competition in the marriage market is more favorable to women. In this case we will be more likely to observe $\rho^{SIL} > \rho$ the lower σ_μ^2 that is, the smaller the randomness component in matching (i.e. when there is no "love") and the higher is σ_y^2 , the inequality in the distribution of labor income. The latter will be higher if labor market returns to human capital, γ_1 , are high and if the persistence of children's "endowment", λ , is high.

7 Conclusion

To be completed.

[REFERENCES BELOW TO BE UPDATED]

References

- [1] Becker, Gary S. and Tomes, N. (1979). "An equilibrium theory of the distribution of income and intergenerational mobility," *Journal of Political Economy*, vol. 87(6), pp. 1153-89.
- [2] Becker, Gary S and Nigel, Tomes (1986). "Human Capital and the Rise and Fall of Families," *Journal of Labor Economics*, vol. 4(3), pp. S1-39.
- [3] Becker, Gary S. (1991). *A Treatise of the Family*, Cambridge, Mass.: Cambridge University Press.
- [4] Bertrand, Marianne and Sendhil Mullainathan. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review* 94 (4), pp. 991-1013.

- [5] Bisin, Alberto and Thierry Verdier (2000). “Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits.” *Quarterly Journal of Economics*, CXV, pp. 955–987.
- [6] Chadwick, Laura and Gary Solon (2002). “Intergenerational income mobility among daughters,” *American Economic Review*, vol. 92(1), pp. 335-44.
- [7] Chamberlain, Gary. “Panel Data,” in *Handbook of Econometrics Volume II* (Z. Griliches and Michael D. Intriligator, eds.), pp. 1248-1318, Amsterdam: North-Holland.
- [8] Epstein, E. and Guttman, R. (1984). “Mate selection in man: Evidence, theory, and outcome,” *Social Biology*, vol. 31(3-4), pp. 243-78.
- [9] Erikson, Robert and John H. Goldthorpe (1992). *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon Press.
- [10] Ermisch, John, Francesconi, Marco and Thomas Siedler, 2006. “Intergenerational Mobility and Marital Sorting,” *Economic Journal*, vol. 116(513), pages 659-679.
- [11] Ferrie, Joseph P. (2005). “History Lessons: The End of American Exceptionalism? Mobility in the United States since 1850.” *The Journal of Economic Perspectives*, 19 (3), pp. 199-215.
- [12] Fernández, Raquel, N. Guner, and J. Knowles (2005). “Love and Money: A Theoretical and Empirical Analysis of Household Sorting and Inequality.” *The Quarterly Journal of Economics*, 2005, 120 (1), pp. 273-344
- [13] Fernández, Raquel and R. Rogerson (2001). “Sorting and Long-Run Inequality” *Quarterly Journal of Economics*, 116, pp. 1305–1341.
- [14] Fryer, Roland G. and Steven D. Levitt (2004). “The causes and consequences of distinctively black names.” *The Quarterly Journal of Economics* 119 (3), pp.767-805.
- [15] Galor, Oded and Omer Moav (2002). “Natural Selection and the Origin of Economic Growth” *Quarterly Journal of Economics*, 117 (4), pp. 1133-1191.

- [16] Goldin, Claudia (1990). *Understanding the Gender Wage Gap: An Economic History of American Women*. Oxford University Press.
- [17] Goldin, Claudia, and Maria Shim (2004). "Making a Name: Women's Surnames at Marriage and Beyond," *Journal of Economic Perspectives*, 18(2): 143–160.
- [18] Güell, Maia; Rodríguez-Mora, José V. and Chris Telmer (2007). "Intergenerational Mobility and the Informative Content of Surnames." Manuscript., University of Edinburgh.
- [19] Hellerstein, Judith K. and Melinda S. Morrill (2009). "Dads and Daughters: The Changing Impact of Fathers on Women's Occupational Choices". Manuscript, University of Maryland.
- [20] 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.
- [21] Kalmijn, M. (1994). "Assortative mating by cultural and economic occupational status," *American Journal of Sociology*, vol. 100(2), pp. 422-52.
- [22] Kremer, Michael (1997). "How Much Does Sorting Increase Inequality?" *The Quarterly Journal of Economics*, vol. 112(1), pp 115-39.
- [23] Lam, David (1988). "Marriage markets and assortative mating with household public goods: theoretical results and empirical implications," *Journal of Human Resources*, 23 (4), pp. 1235-58.
- [24] Lam, David A. and Robert F. Schoeni (1993). "Effects of family background on earnings and returns to schooling: evidence from Brazil," *Journal of Political Economy*, vol. 101(4), pp. 710-40.
- [25] Lam, David A. and Robert F. Schoeni (1994). "Family ties and labor market in the United States and Brazil," *Journal of Human Resources*, vol. 29(4), pp. 1235-58.
- [26] Lieberman, Stanley (2000). *A Matter of Taste: How Names, Fashions, and Culture Change*. New Haven: CT Yale University Press.

- [27] 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.
- [28] Mare, Robert D. (1991). "Five decades of educational assortative mating," *American Sociological Review*, vol. 56(1), pp. 15-32.
- [29] Mulligan, Casey B. (1999). "Galton vs. the human capital approach to inheritance," *Journal of Political Economy*, vol. 107(6), Part 2, pp. S184-S224.
- [30] Peters, Michael and Aloysius Siow (2002), "Competing Premarital Investments," *Journal of Political Economy*, 110(3), 592-608.
- [31] Preston, Samuel H. and Michael R. Haines (1991). *Fatal Years: Child Mortality in Late Nineteenth Century America*, Princeton, NJ: Princeton University Press.
- [32] Raaum, Oddbjørn, Bratsberg, Bernt, Røed, Knut; Österbacka, Eva; Eriksson, Tor; Jäntti, Markus and Naylor, Robin A. (2007) "Marital Sorting, Household Labor Supply, and Intergenerational Earnings Mobility across Countries," *The B.E. Journal of Economic Analysis & Policy*: Vol. 7: Iss. 2 (Advances), Article 7.
- [33] Schwartz, Christine R., and Robert D. Mare. 2005. "Trends in Educational Assortative Marriage from 1940 to 2003." *Demography* 42: 621-46.
- [34] 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.
- [35] Van Poppel, Frans and Daniel Scott Smith (1999). "Introduction," *History of the Family*, 4(3), pp. 229-237.

Table 1. Intergenerational Correlations, Linked Representative Samples

	Linked sample, direct correlation between fathers and sons	Linked sample, pseudo-correlation between fathers and sons
1860-1880	0.4189 (0.0168)	0.4214 (0.0318)
1880-1900	0.4926 (0.0128)	0.5187 (0.0274)
1850-1880	0.3840 (0.0372)	0.3769 (0.0581)
1880-1910	0.4367 (0.0143)	0.4827 (0.0298)
1880-1920	0.3821 (0.015)	0.3991 (0.0300)
1880-1930	0.3571 (0.0153)	0.3722 (0.0314)

Table 2. Summary Statistics for Children's Names, 1850-1910

Year	Sample	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	Percent of children with names linked 30 years later	Share of total variation in log earnings explained by between-name variation	
									Raw	Adjusted
<i>Males</i>										
1850	1%	35,912	3,744	9.6	72.1	7.5	89.4	89.6	0.1215	0.0350
1860	1%	48,424	4,355	11.1	71.9	6.5	93.2	-	0.0962	0.0225
1870	1%	58,412	4,917	11.9	71.1	6.0	-	93.4	0.0945	0.0250
1880	1%	75,665	6,795	11.1	70.6	6.3	92.9	92.3	0.0987	0.0257
1900	1%	104,997	10,569	9.9	72.6	7.3	92.1	91.8	0.1123	0.0302
1910	1%	120,411	11,786	10.2	73.1	7.2	92.3	-	0.1127	0.0384
<i>Females</i>										
1850	1%	34,607	3,632	9.5	72.2	7.6	91.6	91.8	0.1396	0.0390
1860	1%	47,264	4,789	9.9	72.2	7.3	92.2	-	0.1377	0.0406
1870	1%	56,200	5,588	10.1	72.6	7.2	-	92.2	0.1412	0.0467
1880	1%	72,885	7,498	9.7	70.4	7.2	91.7	91.0	0.1358	0.0365
1900	1%	102,951	11,054	9.3	72.5	7.8	91.6	91.2	0.1599	0.0584
1910	1%	117,137	12,262	9.6	72.9	7.6	91.4	-	0.1695	0.0722

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

	20-year intervals		30-year intervals	
	Fathers- Sons	Fathers-Sons in law	Fathers- Sons	Fathers-Sons in law
First year in pair				
1850	0.3975 (0.0312)	0.5114 (0.0335)	0.3050 (0.0278)	0.4442 (0.0298)
1860	0.3587 (0.0263)	0.5271 (0.0258)	-	-
1870	-	-	0.3506 (0.0221)	0.3929 (0.0213)
1880	0.3764 (0.022)	0.4633 (0.0203)	0.3579 (0.0225)	0.4132 (0.022)
1890	-	-	-	-
1900	0.5194 (0.0194)	0.5388 (0.0143)	0.4191 (0.0163)	0.4562 (0.0145)
1910	0.5208 (0.015)	0.4603 (0.0117)	-	-

**Table 4. Intergenerational Elasticities 1850-1930.
Alternative Measures of Log Occupational Income.**

	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930
Log occupational income in:	A: Fathers-Sons				
1950	0.4312 (0.0325)	0.3803 (0.0271)	0.3931 (0.0225)	0.5043 (0.0192)	0.5213 (0.015)
1900	0.4516 (0.0305)	0.4105 (0.0252)	0.4284 (0.0203)	0.4516 (0.0147)	0.4777 (0.0125)
1950 ex. Farmers	0.2343 (0.0506)	0.2073 (0.0328)	0.1952 (0.0259)	0.2382 (0.0244)	0.3132 (0.0191)
1900 ex. Farmers	0.2960 (0.0401)	0.2400 (0.0293)	0.2604 (0.0223)	0.2667 (0.0169)	0.3111 (0.0154)
N (1950)	1193	1497	2290	3363	3826
N (1950 ex. Farmers)	745	946	1553	2393	2907
	B: Fathers in Law - Sons in Law				
1950	0.6006 (0.0363)	0.6170 (0.0279)	0.4600 (0.0203)	0.5576 (0.0145)	0.4853 (0.012)
1900	0.5844 (0.0357)	0.6478 (0.0285)	0.4729 (0.0203)	0.5418 (0.0138)	0.5015 (0.0116)
1950 ex. Farmers	0.3603 (0.066)	0.3045 (0.0345)	0.1806 (0.0218)	0.2580 (0.0183)	0.2591 (0.0145)
1900 ex. Farmers	0.3725 (0.0546)	0.3645 (0.0334)	0.2441 (0.021)	0.2849 (0.0163)	0.3191 (0.0143)
N (1950)	1010	1443	2392	3389	3742
N (1950 ex. Farmers)	583	864	1493	2347	2802

**Table 5. Intergenerational Elasticities 1850-1930.
Alternative Measures of Labor Market Outcomes.**

	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930
Variable	A: Fathers-Sons				
Log occupational income	0.4312 (0.0325)	0.3803 (0.0271)	0.3931 (0.0225)	0.5043 (0.0192)	0.5213 (0.015)
Occupational earnings score	0.3518 (0.027)	0.4309 (0.0281)	0.3846 (0.0191)	0.4232 (0.0148)	0.4265 (0.0124)
Occupational education score	0.1374 (0.0261)	0.2556 (0.0272)	0.2031 (0.0229)	0.2830 (0.0202)	0.4189 (0.0192)
Farmer	0.4833 (0.0264)	0.4574 (0.0213)	0.4085 (0.0164)	0.4103 (0.012)	0.3867 (0.0096)
N (log occ. income)	1193	1497	2290	3363	3826
	B: Fathers in Law - Sons in Law				
Log occupational income	0.6006 (0.0363)	0.6170 (0.0279)	0.4600 (0.0203)	0.5576 (0.0145)	0.4853 (0.012)
Occupational earnings score	0.5486 (0.034)	0.7401 (0.0334)	0.4767 (0.0198)	0.5541 (0.0141)	0.4840 (0.0121)
Occupational education score	0.2515 (0.033)	0.3569 (0.0328)	0.2852 (0.025)	0.4607 (0.0193)	0.5087 (0.018)
Farmer	0.3978 (0.0235)	0.5041 (0.0192)	0.4284 (0.0148)	0.4423 (0.0112)	0.3755 (0.0086)
N (log occ. income)	1010	1443	2392	3389	3742

**Table 6. Intergenerational Elasticities 1850-1930.
Alternative Names Concepts.**

	1850-1870	1860-1880	1880-1900	1900-1920	1910-1930
Name concept:	A: Fathers - Sons				
All	0.4312 (0.0325)	0.3803 (0.0271)	0.3931 (0.0225)	0.5043 (0.0192)	0.5213 (0.015)
Middle initials	0.3913 (0.0304)	0.3524 (0.0251)	0.3643 (0.021)	0.4518 (0.0179)	0.4862 (0.0145)
Nicknames	0.4530 (0.0334)	0.3937 (0.0275)	0.3739 (0.0225)	0.4489 (0.0191)	0.4606 (0.0154)
N (All)	1193	1497	2290	3363	3826
N (M.I.)	1430	1806	2730	4020	4714
N (Nicknames)	1144	1428	2157	3216	3684
	B: Fathers in Law - Sons in Law				
All	0.6006 (0.0363)	0.6170 (0.0279)	0.4600 (0.0203)	0.5576 (0.0145)	0.4853 (0.012)
Middle initials	0.5722 (0.0352)	0.5767 (0.0263)	0.4343 (0.0187)	0.4970 (0.0137)	0.4451 (0.0112)
Nicknames	0.5869 (0.0368)	0.6105 (0.0288)	0.4791 (0.0203)	0.5077 (0.0149)	0.4543 (0.0126)
N (All)	1010	1443	2392	3389	3742
N (M.I.)	1179	1734	2850	4028	4577
N (Nicknames)	978	1376	2279	3283	3641

Figure 1: 20-year Father-Son Intergenerational Elasticities and Correlations in Occupational Income: Alternative Weighting Schemes

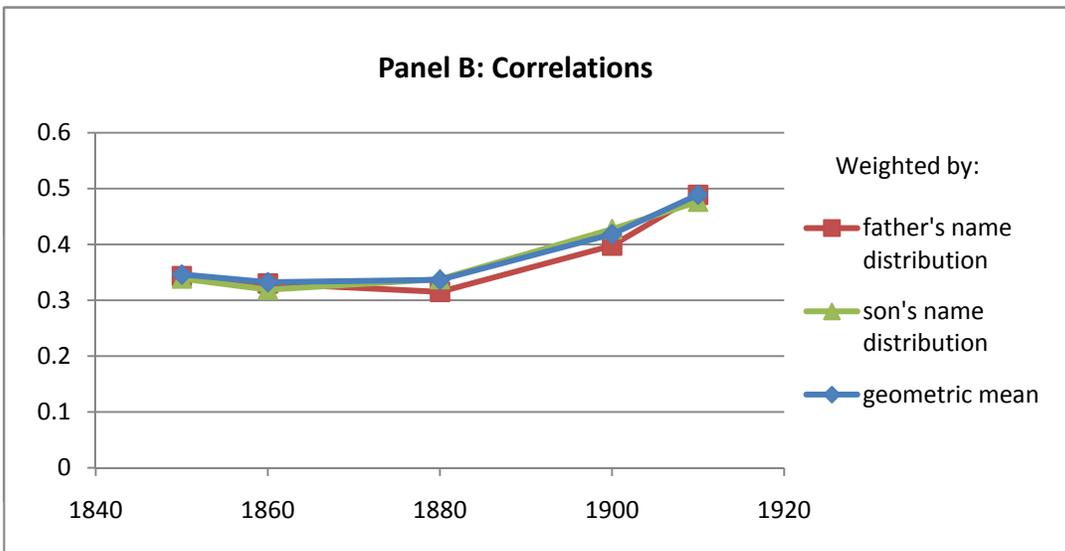
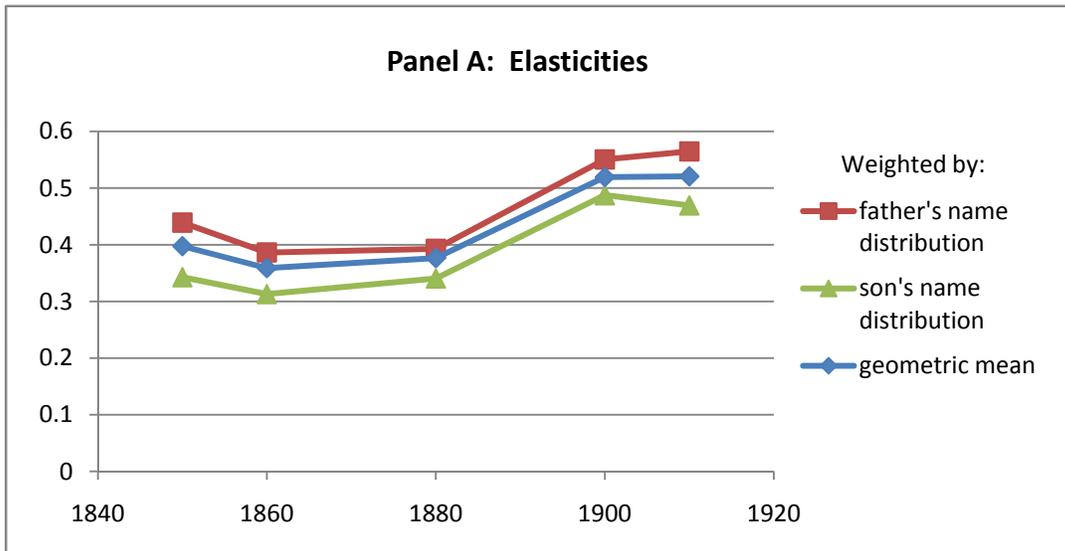


Figure 2: 20-year Father in Law -Son in Law Intergenerational Elasticities and Correlations in Occupational Income: Alternative Weighting Schemes

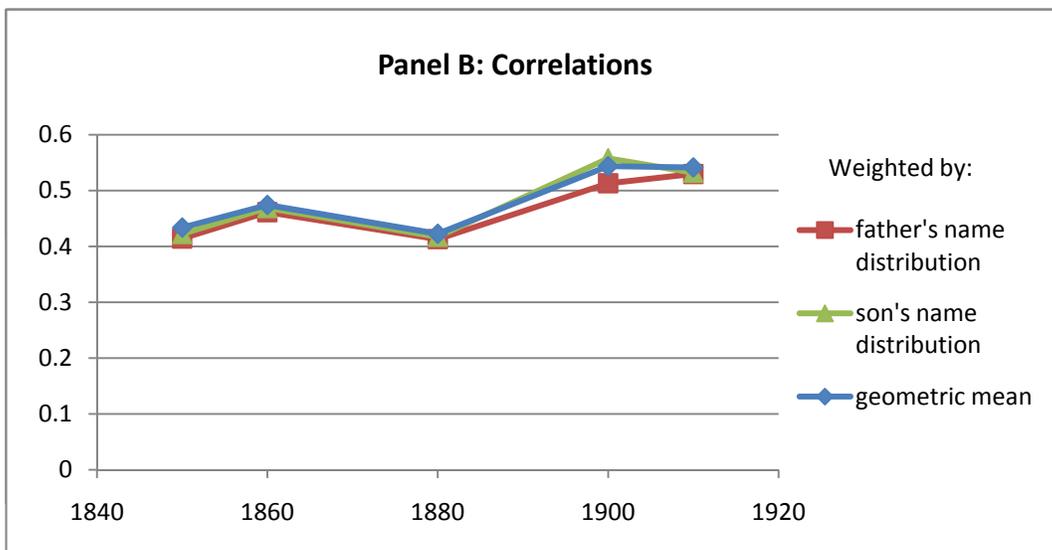
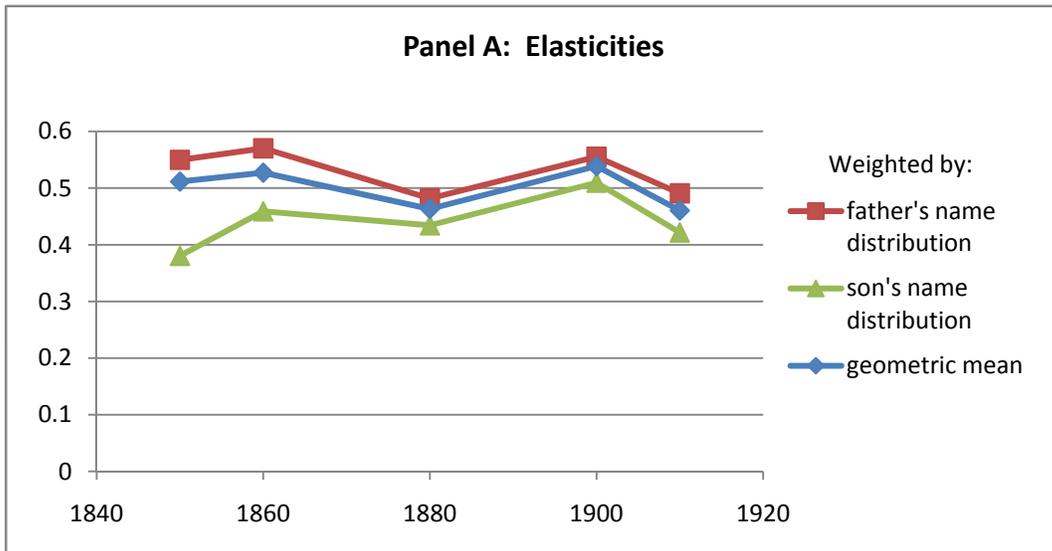
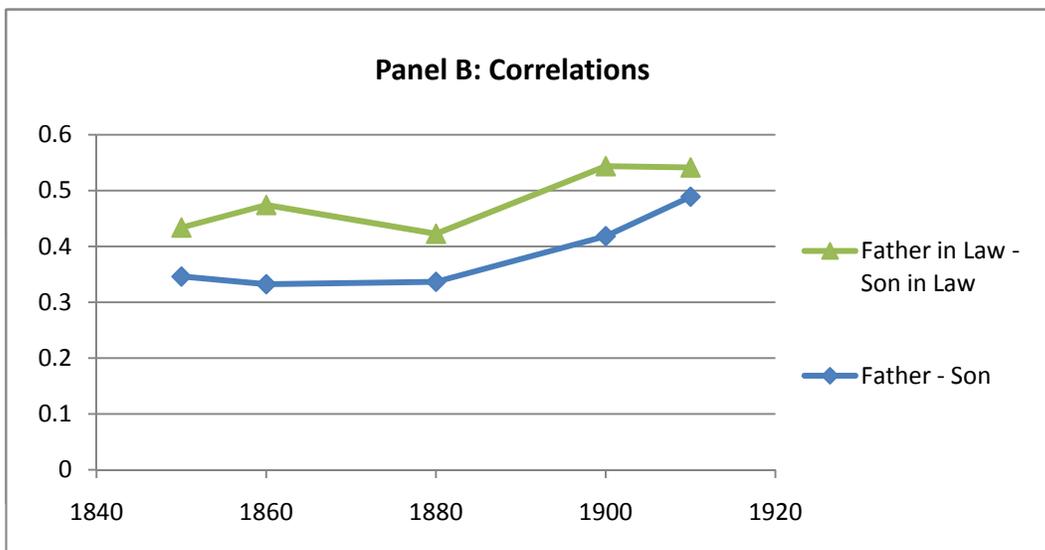
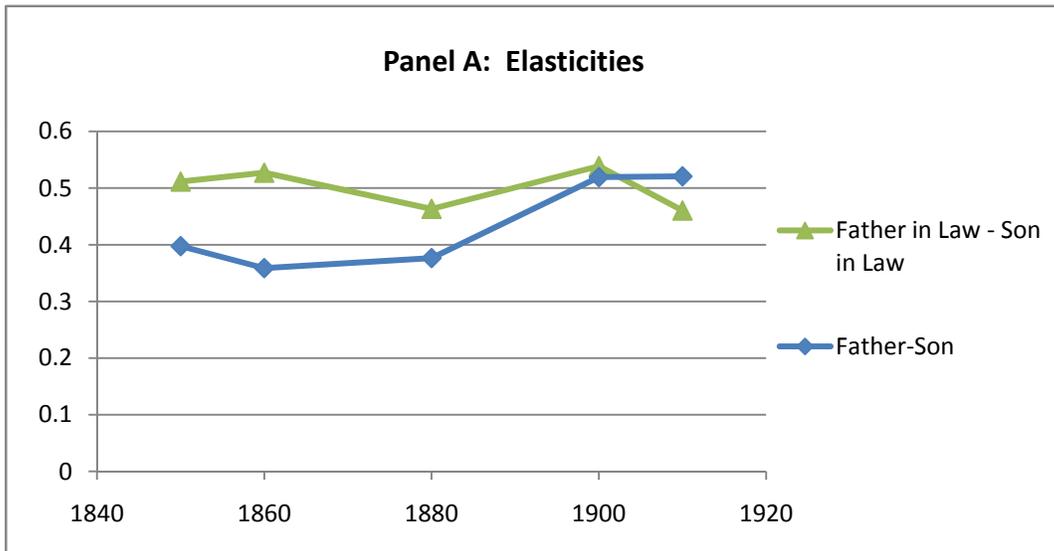


Figure 3: 20-year Father-Son and Father in Law - Son in Law Elasticities and Correlations in Occupational Income



Note: Weights are the geometric mean of the father's and son's (father in law and son in law) name frequencies.