Three-generation Mobility in the United States, 1850-1940: The Role of Maternal and Paternal Grandparents*

Claudia Olivetti
Boston College and NBER

M. Daniele Paserman
Boston University and NBER

Laura Salisbury
York University and NBER

March 2016

Abstract

This paper estimates intergenerational elasticities across three generations in the United States in the late 19th and early 20th centuries. We extend the methodology in Olivetti and Paserman (2015) to explore the role of maternal and paternal grandfathers for the transmission of economic status to grandsons and granddaughters.

We document three main findings. First, grandfathers matter for income transmission, above and beyond their effect on fathers' income. Second, the socio-economic status of grandsons is influenced more strongly by paternal grandfathers than by maternal grandfathers. Third, maternal grandfathers are more important for granddaughters than for grandsons, while the opposite is true for paternal grandfathers. We present a model of multi-trait matching and inheritance that can rationalize these findings.

Keywords: Intergenerational Mobility, Multiple Generations, Gender, Marriage, Assortative Mating

JEL codes: J62, J12

---

*We thank Hoyt Bleakley, Robert Margo, Suresh Naidu, and seminar participants at the "Inequality Across Multiple Generations" workshop, the Society of Labor Economists, the NBER Development of the American Economy Summer Institute, Boston University, Boston College, Federal Reserve Bank of Minneapolis, Mannheim University, MIT, Ohio State University, University of Texas and University of Toronto for helpful comments and suggestions.

1 Introduction

The dramatic increase in income inequality over the past four decades has led to a renewed interest in how economic status is transmitted across generations. A high degree of inequality that persists across generations undermines the very notion of equality of opportunity. The availability of large administrative datasets has pushed the envelope of research on intergenerational mobility, allowing scholars to explore in much more detail the nature of the transmission mechanism across generations (see for example Chetty et al., 2014a and 2014b). One of the most interesting recent developments is the study of the transmission of economic status across multiple generations (Solon 2013, Mare 2011). This extends a large literature that examines intergenerational mobility across two generations, typically focusing on fathers and sons (see Solon, 1999 and Black and Devereux, 2011 for extensive surveys).

While previous literature on intergenerational mobility has focused on transmission from parents to children, the transmission mechanism may be substantially more complex. For example, grandparents may make independent human capital investments in grandchildren, or they may influence parental incentives to invest. Grandchildren might also benefit from the financial resources and social connections of their grandparents. The biological process underlying the transmission of traits is similarly complex, spanning multiple generations. Moreover, both institutions and biology can potentially lead to a differential effect of paternal and maternal grandparents. For example, in a patrilineal society, wealth is largely transmitted through the paternal line. At the same time, the resources of maternal grandparents may facilitate mothers' direct investment in their children, or they may amplify the effect of investment by fathers or paternal grandparents.[1]

In this paper, we estimate intergenerational elasticities across three generations for the United States during the late 19$^{\text{th}}$ and early 20$^{\text{th}}$ centuries, extending a methodology originally developed by Olivetti and Paserman (2015), which exploits the socioeconomic content of first names. Our unique contribution is the analysis of the effects of *both* maternal and paternal grandparents on *both* granddaughters and grandsons.

A handful of studies have tackled the measurement of multigenerational effects on income transmission using historical data. Ferrie and Long (2015) measure occupational mobility

[1]The study of multigenerational processes and influences has recently received increasing attention in the sociological and demographic literature on social mobility. See for example, Zeng and Xie (2014).

across three generations by tracing men through national censuses from 1850 to 1910 in the UK and the US. Clark (2014) examines intergenerational mobility over the very long term by tracing the performance of men with particular surname characteristics over time. These studies find evidence of significant multigenerational effects. However, both studies use surnames in some capacity to trace families over time; as such, they are unable to assess the importance of maternal grandparents in societies in which women change their last names upon marriage, as they cannot be connected to their grandchildren by surname. Moreover, these studies are unable to characterize mobility for married granddaughters, for the same reason. An exception is the work by Lindahl et al. (2015) who estimate the persistence of human capital over four generations of individuals linking data from Malmö's parish registries in the 1930s to the modern census records. They find that the intergenerational correlation in educational attainment of grandchildren does not depend on their own gender or that of their grandparents. The paper is silent about any difference between maternal and paternal grandparents.

The key insight behind our methodological approach is that the information about socioeconomic status conveyed by *first names* can be used to create *pseudo-links* between grandfathers (G1), fathers (G2) and children (G3). Specifically, the empirical strategy amounts to imputing father's income, which is unobserved, using the average income of fathers of children with a given first name. Extending this idea, one can also impute grandfather's income as a weighted average of the name-specific average income of the fathers' fathers, with weights equal to the fraction of fathers with that name among all the fathers of G3 children with a given first name.[2] For example, if half of the fathers of children named "John" in 1880 are named "Adam," and the other half are named "Bob," the income of the grandfather of "John" would be imputed as the simple average of the average income of fathers of "Adam" and the average income of fathers of "Bob" in 1860.

The intuition for why this methodology works can be explained using a simple example. Assume that the only possible names for boys in generation G3 are Adam and Zachary, with high socioeconomic status G2 parents more likely to name their child Adam, and Zachary being more common among low socioeconomic status parents. In a society with a high degree

[2] The data only allows us to calculate the intergenerational elasticity in an index of occupational status based on the 1950 income distribution. Somewhat loosely, we will sometimes refer to our estimates as estimates of the intergenerational income elasticity, or simply intergenerational elasticity.

of intergenerational mobility, we would not expect the adult Adams to have much of an advantage on the adult Zacharys. Moreover, in the previous generation (G1) the fathers of men who name their sons Adam should be almost indistinguishable from the fathers of men who name their sons Zachary. On the other end in a more rigid society the adult Adams grow to be richer than the adult Zacharys, and the G1 fathers of men who name theirs sons Adam are expected to be richer. Therefore, one can obtain a good measure of intergenerational mobility by correlating the average incomes of people with a given name, that of fathers of people with that name, and that of fathers of fathers who assign that name.

A distinct advantage of our approach is that it allows us to measure the importance of maternal grandparents as well as paternal grandparents. Our methodology applies equally well to women: just replace Adam and Zachary in the previous example with Abigail and Zoë, and use husband's income as the measure of women's socioeconomic status.

Olivetti and Paserman (2015) use this methodology to provide the first estimate of intergenerational mobility between fathers and daughters in the late 19^{th} and early 20^{th} Centuries. In the case of three generations, the methodology allows us to estimate four different channels of intergenerational transmission of socioeconomic status: fathers-sons-grandsons, fathers-sons-granddaughters, fathers-daughters-grandsons and fathers-daughters-granddaughters. Moreover, we are able to model intergenerational income transmission by including the income of *both* paternal and maternal grandparents in the same regression. It is important to emphasize at this point that even though our methodology does not necessarily recover the intergenerational elasticity estimates that would be obtained with a true intergenerationally linked data set, it is still able to provide comparable estimates of the evolution of long-run mobility across all the possible gender lines. Thus, our analysis is able to explicitly test the relative importance of paternal and maternal grandparents, which affords it the potential to uncover different mechanisms through which gender differentials in intergenerational mobility may arise.

Using 1% extracts from the Decennial Censuses of the United States between 1850 and 1940, we find evidence that, even after controlling for the income in generation G2 ("father's income"), the income of generation G1 ("grandfather's income") has a large and positive effect on the income of generation G3 ("grandchild's income"). In addition, we find interesting gender differentials in the strength of the correlation between the three generations. Our results indicate that the transmission of economic status occurs mostly along gendered lines. That is, paternal grandfathers matter more than maternal grandfathers for the income of grandsons, while the opposite is true for granddaughters. Furthermore, holding the gender of the second generation constant, we find that maternal grandfathers are more important for granddaughters than for grandsons, while the opposite is true for paternal grandfathers.

We consider a number of possible explanations for the gender asymmetry in the multi-generational transmission of economic status. Explanations based on direct investment of grandparents in grandchildren (driven by residential location of married children), the timing of intergenerational transfers, or married women's rights to own property or sign contracts do not seem appropriate for the period of analysis (late 19th and early 20th Century), as the US had mostly completed its transition to an industrial society. We propose an alternative framework, which builds on Chen et al. (2013), in which individuals' desirability in the marriage market is a function of 'market' and 'non-market' traits that are transmitted from parents to children. This framework can rationalize our findings if there are gender asymmetries in the relative importance of market and non-market traits, as well as differences in the degree of inheritability across traits and genders.[3]

Our findings suggest that traditional estimates of intergenerational mobility that assume a first-order autoregressive process for income may substantially understate the true extent of intergenerational persistence in economic status, in accordance to other recent papers that link multiple generations (e.g., Ferrie and Long, 2015; Lindahl et al., 2015). Solon (2015) discusses several mechanisms that would lead the intergenerational transmission process to decay at a rate that is slower than exponential. The standard Becker-Tomes (1979) model generates the prediction that son's income is a function of father's income and an error term that can be decomposed into an AR(1) component (often interpreted as a family's "endowment" of earning ability) and an idiosyncratic term. This model predicts that the coefficient on grandfather's income in an AR(2) regression is negative. However, Solon shows that if one relaxes the assumption that the endowment process is AR(1), (for example, if grandparents

[3] There are relatively few studies that look at gender differentials in the link between grandparents and grandchildren. One notable exception is Duflo (2003), who studies the extension of pension benefits to black South-Africans in the 1990s. She shows that pensions received by grandmothers, particularly maternal grandmothers, had a large impact on the anthropometric status of granddaughters, but little effect on that of grandsons. In contrast, no similar effect is found for pensions received by grandfathers. Vidal-Fernandez and Posadas (2013) provide evidence that by providing childcare services, maternal grandmothers may have a positive (causal) impact on the labor supply of their daughters.

contribute directly to their children's human capital accumulation) it becomes possible to obtain a positive coefficient on grandparents' income. By making the endowment multidimensional, our model can be viewed as an alternative way of relaxing the AR(1) assumption for the endowment process.

The rest of the paper is organized as follows. The next section discusses the methodology as well as the data used for the analysis and some measurement issues. The main results and some robustness checks are presented in Section 3. Section 4 presents the theoretical frameworks that we use to provide a possible interpretation for our findings. Section 5 discusses how regional differences in the transmission of economic status across generations can be interpreted in light of our model, and Section 6 concludes.

2 Methodology and Data

Consider an individual i belonging to G3 who is a child at time $t-s$ and adult at time t (in practice, we will look at generations separated by 20 or 30 years). Let y_{it} be individual i's log earnings at time t, y_{it-s} be his father's (G2) log earnings at time $t-s$, and y_{it-2s} be his father's father's log-earnings (G1) at $t-2s$. With individually linked data, y_{it}, y_{it-s} and y_{it-2s} are all observed, and the intergenerational elasticity estimate is obtained by regressing y_{it} on y_{it-s} and y_{it-2s}.

Assume instead that we only observe three separate cross-sections, and it is impossible to link individuals across the three. This means that both y_{it-s} and y_{it-2s} are unobserved, and it becomes necessary to impute them. 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.[4]

Linking generation G2 to generation G3

To link individuals from generations G2 and G3, we follow the same approach used in Olivetti and Paserman (2015). For a G3 adult at time t named j, we replace y_{it-s} with $\tilde{y}'_{jt-s}$, the average log earnings of G2 fathers of children named j, obtained from the time $t-s$ cross section (the "prime" indicates that this average is calculated using a different sample). We

[4] There has been a flurry of recent research that uses the informative content of *surnames* to obtain estimates of social mobility across generation. See Güell, Rodríguez-Mora and Telmer (2015), Collado, Ortuño Ortín and Romeu (2012) and Clark (2014). Our study differs in that we exploit the informative content of *first* names.

restrict the sample of G2 fathers to be in the same age range as the sample of G3 adults; this is to facilitate our links to G1, which will be explained in the following subsection.

This methodology amounts to creating a "generated regressor" by using one sample to create a proxy for an unobserved regressor in a second sample.[5] A key requirement of this methodology is that first names carry information about socioeconomic status, or alternatively, that names are not distributed randomly in the population.[6] This ensures that the average income of fathers of children with a given first name is a reasonable proxy for the fathers' actual income. Otherwise the generated regressor would be just noise and our estimator would converge to zero.[7]

Linking generation G1 to generation G2

Adding a link to generation G1 is slightly more complicated. We would like to impute G1's income to a G3 adult named j as the average income of the grandfathers of children named j in year $t-2s$. However, two difficulties arise: first, G3 adults in year t would not have been born in year $t-2s$, so it is impossible to make a "direct" pseudo-link to year $t-2s$. Second, making "direct" pseudo-links from G1 to G3 would require households to be multigenerational, i.e. containing children and grandfathers residing together, which was not typically the case.

However, we can still apply the same principle used for the G2-G3 link, extended to an additional generation. For example, suppose that children named Adam in year $t-20$ have fathers named David, Edward and Fred, in equal proportions. The income assigned to G1 for the group of G3 adults named "Adam" is the weighted average, with weights $\frac{1}{3}$, $\frac{1}{3}$, $\frac{1}{3}$, of the average income at time $t-40$ of all fathers of children named David, Edward and Fred, respectively.

Formally, we proceed as follows. First, we calculate $q'_{j,k}$ as the fraction of fathers (G2) named k of children (G3) named j. This value is taken from Census year $t-s$, in which

[5]Note that there is a trade-off in imputing income. On the one hand, using finer cells to impute father's income (such as last names, or first names by state of birth) can achieve higher precision in the imputed values. On the other hand, it might reduce the cell size used to compute the average and exacerbate measurement error.

[6]The empirical evidence strongly supports the assumption that parents choose first names partly to signal their own standing in society, or their cultural and religious beliefs (see, for example, Bertrand and Mullainathan, 2004, and Fryer and Levitt, 2004).

[7]See Olivetti and Paserman (2015) for a more detailed discussion of the econometric properties of this estimator.

G2 individuals are adults, and G3 children still live at home with their parents. Second, we calculate $\tilde{y}''_{k,t-2s}$, the average log earnings of G1 fathers of children named k (this average is calculated from Census year $t-2s$ and we use the "double-prime" to indicate that this average is calculated using yet a different sample). Finally, we calculate the average log earnings of the grandfathers of G3 adults named j as:

$$\tilde{y}''_{j,2t-s} = \sum_{k} q'_{j,k} \tilde{y}''_{k,t-2s}$$

In other words, $\tilde{y}''_{j,t-2s}$ is a weighted average of the name-specific average log earnings of the fathers of G2 fathers, with the weights equal to the fraction of G2 individuals with that name among all the fathers of G3 children named j.

One can then obtain an estimate of the income elasticity across the three generations by running a regression of $y_{i,t}$ on $\tilde{y}'_{j,t-s}$ *and* $\tilde{y}''_{j,t-2s}$. This regression is run at the individual G3 level, with all G3 adults with the same first name having identical imputed incomes of G2 and G1.[8]

How does this estimator compare to the one based on individually linked data? As shown in Olivetti and Paserman (2015) its difference relative to the standard estimator can be decomposed into three parts.

The first component is the traditional attenuation bias deriving from the fact that we replace true father's and grandfather's income with an imputed value, thus introducing measurement error. The three-generation case includes an additional source of measurement error stemming from the fact that we also need to estimate the weights $q'_{j,k}$ in equation 1. This extra source of measurement error should bias us strongly against finding any effect of grandparents' income on grandchildren. In practice, we will see that this is not the case.

The second component captures the direct effect of first names that goes above and beyond any direct effect of father's income. First, there is a direct labor market premium (or penalty) potentially associated with a given first name. This may reflect factors such as ethnicity, religion, state of birth, or any other signal of social status associated with a given first name. In addition, parents might engage in "aspirational naming." This would be the case if ambitious

[8] It can be shown that this estimator can be written as a special case of a "two-sample two-stage least squares" estimator (Inoue and Solon, 2010), where there are two variables that are instrumented, and each instrument is constructed using a different sample. Details available upon request.

and motivated parents who assign children high socio-economic status names also transfer to them their work-ethic and push them to succeed in the labor market. Both forces are likely to make the estimator larger than that obtained using individually linked data.

The third component of the difference derives from the use of different samples to impute the income of the previous generations. This means that the estimator omits the intergenerational correlation in motivation, genetic ability, social capital and other unobservables that are not embodied in first names. If these unobservables are positively correlated across generations, as is reasonable to assume, then our estimator will be pulled down relative to the linked estimator.

Overall, the pseudo-panel estimator can be either lower or higher than the linked estimator, depending on which of the three effects dominates. In practice, Olivetti and Paserman (2015) show that in samples in which one can calculate both, the pseudo-panel estimator is approximately 30% lower than the linked estimator.

The description above was presented in terms of the father-son-grandson relationship. It is easy to see, however, that the methodology can also be applied to fathers-son-granddaughters, fathers-daughters-grandsons, and fathers-daughters-granddaughters, thus allowing the analysis of gender differentials in the transmission of economic status across multiple generations. Because married women rarely worked during this period, we will always proxy a woman's socioeconomic status by that of her husband. This implies that the income of the middle generation is always that of the man. In other words, in all our specifications we will regress the income of a member of generation G3 (this could be either the grandson or the granddaughter's husband) on the income of generation G2 (this is always the man) and on the income of generation G1 (which could be either the maternal or paternal grandfather).

Data and Measurement Issues

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

Individual level data are available from IPUMS for every decadal Census from 1850 to 1940, with the exception of 1890. This means that we can calculate our three-generation measures of intergenerational mobility for three triplets observed at a distance of 20 years from one another (1860-1880-1900, 1880-1900-1920, and 1900-1920-1940); and for three triplets of observations observed at a distance of 30 years from one another (1850-1880-1910, 1870-1900-1930, and 1880-1910-1940). This gives us a unique long-run perspective on the transmission of economic status across generations.

A challenge that applies to all computations 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' occupational status.[9]

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). 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. This variable indicates the median total income (in hundreds of dollars) of persons in each occupation in 1950.

2.1 Assessing the Methodology

In order to assess our methodology, we compare our intergenerational income elasticities across three generations to those obtained using the IPUMS Linked Representative Samples. This comparison requires us to restrict our analysis to males, and to focus on two samples (1860-1880-1900 and 1850-1880-1910) for which linked data across two cohorts are available.

Using these data, we define G3 income as the income of the adult in the latest year of the triplet (1900 or 1910). G2 income is the income of this person's father in 1880; this is a true link. To obtain G1 income, we create a pseudo link: because we observe the name of the individual's father in 1880, we can calculate the average income of fathers of boys with this

[9] Income from salaries and wages is available in 1940. However, there is no information on income from self-employment, which was potentially quite important at the time. Because of this limitation, and to maintain consistency with the rest of the analysis, we choose not to use this variable.

name in the first year of the triplet (1860 or 1850).

Table 1 compares estimates of intergenerational mobility obtained using either the linked data (Panel A) or the pseudo-linked data based on the 1% IPUMS samples and the methodology described above (Panel B).

Column 1 and 3 in the table show the standard two-generation intergenerational elasticity. In panel A the estimates are based on actual father-son pairs. In panel B the fathers' occupational income is imputed based on sons' first names. As in Olivetti and Paserman (2015), the latter estimate based on the pseudo link is smaller by 12 to 24 log points.[10]

Column 2 and 5 show the estimates of a regression of G3 income on G1 income. The estimates in both panels are based on imputation of G1's income. However, in panel A the imputed value is simply the average income of G1 fathers of G2 children with a given first name. In panel B it is a weighted mean of average G1 incomes, weighted by the distribution of G2 first names.

This double averaging poses a challenge to our methodology. It implies that the distribution of income for G1 is substantially more compressed than that of G2 and G3. This is apparent from the standard deviation of the average log occupational income of each of the three generations (calculated at the G3 name level). In our sample of G2 and G3 males in 1860-1880-1900, this value is 0.314 for G3, 0.298 for G2, and only 0.091 for G1. Similarly, in our sample of G2 and G3 males in 1850-1880-1910, the standard deviations of mean log occupational income by first name for G3, G2, and G1 are 0.341, 0.304, and 0.122, respectively. As a result, the OLS estimates of the G1-G3 intergenerational elasticity in panel B are implausibly large (0.6 and 0.29) compared to those in panel A, which do not require double averaging for the imputation of G1's income (0.24 and 0.17). In addition, standard errors on the G1 coefficient in panel B are quite large, despite the sample in panel B being 20 to 30 times larger than that in panel A. The same patterns also arise in Column 3 and 6 where we include both G1 and G2 income on the right hand side.

To address this issue, we change our explanatory variables from log occupational scores

[10]Olivetti and Paserman (2015) estimate the two-generation elasticity to be 0.34 for 1880-1900 and 0.32 for 1880-1910. Our estimates here are somewhat smaller because we impose the additional restriction that G2 individuals be between the ages of 20 and 35 (30 to 45 for the 30-year interval); this facilitates our links to G1. However, a consequence of this restriction is that the number of observations in G2 declines, leading to increased attenuation bias in the estimated two-generation elasticity.

to percentile ranks of the mean of log occupational income by first name.[11] The results are reported in Table 2. While there is still a discrepancy in the G2-G3 elasticities between panel A and B, as one would expect because of the imputation, the coefficients on G1 are fairly similar. Moreover, the coefficient on G1 income is estimated as precisely as the one on G2 income. In the remainder of the paper, we will use percentile ranks as our explanatory variables.

The coefficients in columns 1 and 4 in Panel B indicate that going from the bottom to the top percentile of earnings in generation G2 is associated with an increase in G3's log occupational income of about 0.26. Going from the bottom to the top percentile in G1 income is associated with a 0.13 to 0.15 increase, implying a fairly strong correlation between the socioeconomic status of grandsons and paternal grandfathers. Moreover, we find that G1's rank has a significant positive effect on the log income of G3, even after controlling for the rank of G2 (columns 3 and 6). The statistically significant coefficient on G1 income implies that the intergenerational income transmission process might not be well characterized by an AR(1) process. Ignoring the second order autoregressive term will lead to overstating the extent of long-run mobility across generations. Our estimates in Table 2 imply that a given shock to a dynasty's income would take at least one more generation to fade out relative to what would be predicted by an AR(1) process. For example, take the estimates in column 4 and 6 in Panel B. Based on the AR(1) estimate, it would take 2 generations for 90% of the shock to dissipate. For 99% of the shock to go away, it would take 4 generations. The corresponding numbers for the AR(2) model are, instead, 3 and 5 generations.

We can compare these numbers to Ferrie and Long (2015, Table 3) who estimate three-generation mobility along the paternal line. They use data linked across three generations (1850-1880-1910), and they impute income based on the 1860 occupational wealth distribution. Their estimates are roughly similar to ours, with a slightly lower G2-G3 elasticity in both the AR(1) and the AR(2) models, but a slightly higher G1-G3 elasticity. Their AR(2) estimates imply the same level of persistence of a given income shock as that predicted by our calculations.

[11] The results are essentially identical if we use the percentile rank of mean occupational income in levels or if we also use percentile ranks for the dependent variable. Chetty et al. (2014b) also advocate using a rank-rank specification to account for non-linearities in the log-log relationship between father's and son's incomes.

3 Results

3.1 Basic Results

We now apply our methodology to study how both grandsons and granddaughters are affected by grandparents, along both the maternal and the paternal line. Table 3 presents regressions of G3 earnings on the percentile rank of earnings of G1 and G2, using all possible gender combinations and all decade triplets in which the distance between generations is 20 years (1860-1880-1900; 1880-1900-1920; 1900-1920-1940). Panel A reports the results for grandsons (G3 Male), while Panel B reports the estimates obtained for granddaughters (G3 Female). There are two columns for each decade triplet reporting the elasticities along the paternal and the maternal line, respectively. Table 4 follows the same structure for the decade triplets in which the distance between generations is 30 years (1850-1880-1910; 1870-1900-1930; 1880-1910-1940).

We first note that the coefficient on G1 is positive and significant in almost all cases at both 20 and 30-year intervals. This offers further support for the existence of multigenerational effects. As in Olivetti and Paserman (2015), we find that the G2 coefficient increases between 1900 and 1920, then tends to level off by 1940: one-generation mobility in the US declines between the late part of the 19^{th} and the early part of the 20^{th} Century. The 30-year G2 elasticities in Table 4 exhibits a broadly similar time trend.

We also find interesting differences in the G1-G3 intergenerational elasticity by gender, which we illustrate in Figures 1 and 2. We first look at the relative importance of maternal and paternal grandfathers. For grandsons (left panels of Figures 1 and 2), the elasticity with respect to paternal grandfather's income (the solid blue line) is consistently larger than the elasticity with respect to maternal grandfather's income (the dashed red line), although the differences are mostly not significant at conventional levels. For granddaughters (right panels of Figures 1 and 2), the pattern appears to be reversed (maternal grandfathers have a larger influence than paternal grandfathers), even though the evidence is more mixed.

We then look at whether paternal and maternal grandfathers affect grandchildren of different genders differently (this amounts to comparing the solid blue lines and red dashed lines across the left and right panels of the figures). Paternal grandfathers' income tend to be more strongly associated with grandsons' income than granddaughters' income. The opposite pattern appears when looking at the effect of maternal grandfathers – they tend to have a greater influence on their granddaughters than their grandsons.

Finally, there seems to be an increasing importance over time of grandparents' income on grandsons, with the trend being more pronounced for paternal grandfathers. On the other hand, the effect of grandparents on granddaughters appear to be mostly flat over time.

3.2 Robustness

One initial concern is that comparisons by gender may be sensitive to the way our samples are constructed. For example, we measure a woman's socioeconomic status by the earnings of her husband. This means that all women in our sample are married, whereas men need not be married to be included. Then, we may be measuring differences in intergenerational income transmission by marital status rather than gender. Furthermore, we do not place restrictions on the age of these husbands in our baseline specification; therefore, our results may reflect the fact that we are measuring income at different points in the life cycle for women and men.

To ensure that our results are not being driven by these details of our sample construction, we redo the analysis imposing different restrictions on G2 and G3. The additional restriction we impose on G2 is that individuals in the sample be married to a spouse in the same age range as the individual (20-35 or 30-45, depending on the sample years). We impose two additional restrictions on G3. First, we restrict individuals to be married; second, we restrict individuals to be married to a spouse in his or her age range. We calculate the G1-G3 intergenerational elasticity for each of 6 combinations of these sample restrictions (including the baseline sample restrictions). The results, using the 1860-1880-1900 sample, are reported in Table A1 in the Appendix.[12] Altogether, it appears that these different sample restrictions have only a minimal effect on the baseline results.

To summarize the results of the above robustness analysis, we compile all G1-G3 intergenerational elasticities estimated under different sample restrictions in each decade triplet. There are 144 such estimates.[13] We regress these on indicators for chronological order (earliest, middle, or latest sample), the interval that separates generations (20 or 30 years), the type

[12] The full set of results, using all samples, can be downloaded at `http://dept.econ.yorku.ca/~lsalisbu/3GenMobility_AppendixTables_2016.xlsx`.

[13] Four possible combinations of grandparents (paternal and maternal) and grandchildren (male and female); two time intervals (at 20 and 30-year); three time periods; two possible sample restrictions on G2; and three possible sample restrictions on G3: $4 \times 2 \times 3 \times 2 \times 3 = 144$.

of grandfather, the gender of G3, and categorical variables indicating which sample restrictions are imposed. Standard errors are clustered at the specification level. We report these results in Table 5. In column (1) we pool all specifications, and in the remaining columns we separate them by gender. Column (2) contains only G3 males, and column (3) contains only G3 females; similarly, column (4) contains only paternal grandfathers, and column (5) contains only maternal grandfathers.

This exercise supports the existence of an upward time trend in the G1-G3 intergenerational elasticity, and it suggests that this elasticity declines as the interval at which cohorts are constructed increases. While there is no overall tendency for the coefficient to be higher when G2 is male, the picture changes dramatically when we separate G3 by gender. For G3 males, the effect of the paternal grandfather is clearly stronger. For G3 females, the opposite is true. There is no significant difference in the G1-G3 intergenerational elasticity by gender of G3 overall, but this masks significant differences when we separate by gender of G2 (i.e. comparing the effect of maternal and paternal grandfathers). The impact of the paternal grandfather is much stronger for G3 males than G3 females. The impact of the maternal grandfather tends to be stronger for G3 females, although this is not quite statistically significant.

3.3 Gender Differentials: Extension of Methodology

In order to assess the gender differences documented above more directly, we extend our methodology to include both grandfathers in the same regression. To see how this can be accomplished, consider the following example. Suppose there is one G3 child named Adam in 1880, and his parents (both between the ages of 20 and 35) are named Bill and Barbara. The (G2) income of both Bill and Barbara is defined as Bill's income, as we are defining a woman's income as that of her husband. Paternal G1 income will be the average income of fathers of children named Bill in 1860; similarly, maternal G1 income will be the average income of fathers of children named Barbara in 1860. These values can both be included in a regression of Adam's (G3) income in 1900 on the income of G2 and G1. This approach has two advantages: first, it allows us to estimate the effect of one grandparent's income, holding the income of the other grandparent constant; second, we can directly test whether or not paternal grandparents have a greater effect on grandchildren's income than maternal grandparents, which is what our results so far suggest.

These results are presented in Table 6. Rather than estimating G1-G3 elasticities for each pseudo-panel separately, we pool all three panels constructed at 20-year (columns 1 and 2) or 30-year (columns 3 and 4) intervals, and include decade controls in our regressions. This allows us to neatly characterize gender differences in the transmission of socioeconomic status over the entire sample period. The regressions are run separately for G3 males and females, and the coefficients are reported side by side. The third to last row contains the p-value from a test of the equality of the coefficients on paternal and maternal G1, for a given G3 gender. For G3 males, the coefficient on paternal G1 is significantly higher than the coefficient on maternal G1 in both the panels constructed at 20 and 30 year intervals. For G3 females, the coefficient on maternal G1 is larger than the coefficient on paternal G1 in both cases, but it is only significant (at the 10% level) when the panels are constructed at 30 year intervals.

Looking across columns, we see that the coefficient on paternal G1 tends to be higher for G3 males than for G3 females, and this is statistically significant in both cases (the p-value for this test is reported in the second to last row of the table). The opposite is true of the coefficient on maternal G1: this is higher for G3 females, and this difference is always significant (p-value reported in the last row of the table).[14]

These results reinforce our previous findings on gender differences in the transmission of income across three generations. We discuss alternative economic mechanisms that can rationalize these gendered patterns of intergenerational transmission in the next Section.

We test the sensitivity of these results to our method of measuring occupational income. Our findings make use of a ranking of occupations based on the 1950 occupational wage distribution. If the 1950 wage distribution differs from the wage distribution during the period we focus on, this may affect our results. Most importantly, our results are likely to be sensitive to the placement of farmers in the occupational wage distribution, as farmers comprise a very large fraction of the occupations in our sample.

To test the sensitivity of our results to the occupational ranking, we use an occupational income distribution from 1900, and we impute a wage for farmers using data from the 1900 Census of Agriculture. The 1900 occupational wage distribution is obtained from the tabula-

[14] We have also experienced with including interactions between paternal and maternal grandparents, as well as interactions between grandparents' and parents' income. Significant interaction effects could point to the presence of either substitutability or complementarity between different grandparents (or between grandparents and parents) in the production of grandchildren's human capital. In the vast majority of specifications we did not find any evidence of significant interaction effects, and in any case the pattern of signs was not consistent.

tions 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.[15] Preston and Haines explicitly refrained from imputing an average income for generic farm owners. To fill this gap, we impute farmer's income using data from the 1900 Census of Agriculture and a method based on Abramitzky et al. (2012).[16]

To further test that our results are not sensitive to our occupational wage measure, we also rank occupations based on personal property reported in 1860 and 1870. This is advantageous because it corresponds to the earlier periods in our analysis. We calculate mean personal property of household heads by occupation, pooling data from 1860 and 1870 and adjusting for price differences between these two decades. One issue is that farmers' personal property consists largely of equipment or resources used in productive activities. Including this property will likely overstate the status of farmers considerably. Therefore, we adjust farmers' personal property downward by the average value of farm equipment and livestock in 1860 and 1870, using national average values from the census of agriculture (Haines and ICPSR 2010).[17]

We report results using the 1900 wage distribution and the 1860-1870 occupational wealth distribution in Table 7. We estimate the coefficient on maternal and paternal grandparents simultaneously, as we did in Table 6, pooling all three panels constructed at 20 or 30 years intervals. When we use the 1900 wage distribution, the magnitude of the G1-G3 elasticity is quite comparable to that obtained using the 1950 occupational wage distribution. The gender differences in the G1-G3 elasticity also remain, although these differences are not always significant at conventional levels. Similarly, when we use the 1860-1870 wealth distribution, though the G1 coefficients are typically larger in this case.

We have also tested the sensitivity of our findings to including controls for age, immigrant status, literacy, and sibling age rank (results not reported).[18] These controls have a minimal

[15] One limitation of this measure is that the 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.

[16] In an nutshell, 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 the income for specialized farm workers (stock raisers, fruit growers, etc.) in the Preston-Haines tabulations. For more details see Olivetti and Paserman (2015), Appendix C.

[17] This follows Olivetti and Paserman (2015). An alternative that yields similar results is to calculate the occupational ranking as mean personal property by occupation excluding the South (Ferrie and Long, 2015).

[18] Sibling age rank is calculated in the following way: the sibling age rank of an adult named i in year t is equal to the mean sibling age rank among children named i in year $t-20$ (or $t-30$). We test the sensitivity

effect on the coefficients on G1 earnings, and they do not alter the underlying patterns we find.[19] Altogether, we conclude that status is transmitted across three generations in a way that depends on gender, even though some of the results are somewhat sensitive to the exact measurement of income.

4 Interpretation

The empirical analysis in the previous sections has uncovered a number of interesting stylized facts about the intergenerational correlations between grandchildren and their paternal and maternal grandfathers. To summarize, we have documented that the income of grandsons is correlated more strongly with that of their paternal grandfathers. Conversely, granddaughters' income is most strongly related to that of their maternal grandfathers. In addition, paternal grandfathers are more important for their grandsons than their granddaughters, while the ranking is reversed for maternal grandfathers. What type of models can rationalize these findings?

4.1 Paternal and maternal grandparents

We first focus on the differential effect of paternal and maternal grandparents. There are a number of reasons to expect that economic status is passed along the paternal and maternal lines in different ways.

For example, if sons inherit a larger portion of the family wealth than daughters, this would mechanically lead to finding a stronger correlation between the economic status of

of our findings to including this control because we want to argue that different naming patterns for sons and daughters cannot explain our results. In particular, it is possible that first born sons tend to be named after their fathers and first born daughters tend to be named after their mothers. If this is the case, then our estimate of paternal G1 income for G3 males would be averaged over fewer G2 names (and thus more accurate) than our estimate of maternal G1 income for G3 males. Conversely, our estimate of maternal G1 income for G3 females would be averaged over fewer G2 names than our estimate of paternal G1 income for G3 females. This may generate gender-based differences in measurement error, which may bias us in favor of finding gender-based differences in G1-G3 transmission. We argue that such a difference would be largely driven by first-born children, so, if this is occurring, our results should be sensitive to controlling for sibling age rank; they are not. In addition, in tables A2 and A3, we show that there is no gender asymmetry in the concentration of parents' and children's names which might bias our results in this direction.

[19] We have also tested the sensitivity of our results to treating first names differently when creating our pseudo-panels. We use two alternative name groupings. First, we divide individuals with the same first name into two groups: those who report a first initial and those who do not. Second, we define "names" to be soundex codes; this will group together people with differently spelled but phonetically similar names. The results are broadly similar, and are available upon request.

grandchildren and their paternal grandfathers. During the Colonial period there was in fact unequal treatment of daughters in terms of inheritance laws: primogeniture or the "double portion" provision for the eldest son ensured that gender and birth order affected the amount of wealth inherited. However, by 1800 primogeniture had been formally abandoned everywhere (Alston and Schapiro, 1984) and in most states sons and daughters were entitled to equal shares of personal and real property (Shammas et al., 1987). In practice, around the turn of the 19th century daughters were still treated somewhat unequally, but by 1890 the treatment of sons and daughters had become virtually identical (Shammas et al., 1987, based on a sample of probate records from Bucks County, Pennsylvania). Therefore, this mechanical correlation seems unlikely to be relevant for our period of analysis.

It is also possible that paternal grandfathers directly invested more in their grandchildren than maternal grandfathers. This pattern could arise, for example, if the custom of passing surnames along the male line led grandfathers to develop a stronger preference for grandchildren that carry their family name. Alternatively, postmarital location norms could have facilitated the direct investment of paternal grandfathers in their grandchildren. In *virilocal* societies, married couples live with the husband's family. If the U.S. was largely virilocal, we would expect transmission across generations to be strongest along the paternal line. However, an examination of the 1880 and 1900 IPUMS samples reveals that only 10-12 percent of married couples under 35 resided in the same household as a parent, even though that parent was significantly more likely to belong to the husband, especially in agricultural families. The importance of this channel is likely to have declined over our sample period with the decline of the farming sector in the U.S. economy. In addition, given that women married and had children much earlier than men, maternal grandparents were more likely to know their grandchildren, which would allow them to invest directly in their human capital, thus reversing the relative importance of maternal and paternal grandparents.

Even when intergenerational transfers of wealth or human capital do not flow directly from grandparents to grandchildren, paternal grandfathers may have exerted greater control over these transfers than maternal grandfathers. This matters if the preferences of grandfathers with respect to grandchildren's consumption are systematically different from those of the middle generation. Specifically, consider a dynastic model of intergenerational transmission, in which each generation has quasi-hyperbolic, or $\beta - \delta$ preferences over its own consumption and that of future generations.[20] If each generation heavily discounts the utility of future generations relative to its own utility, but the discount factor between any two future generations is relatively low, this creates a tension between grandparents' and parents' desired allocation of consumption across the three generations. Namely, grandparents (G1) prefer to allocate more to their grandchildren (G3), relative to what would be chosen by the parents (G2). So, the G1-G3 elasticity should be greater when G1 is better able to enforce his preferred allocation across the three generations.

Which institutions may have enabled paternal grandfathers to enforce their preferences to a greater extent than maternal grandfathers? Botticini and Siow (2003) argue that in virilocal societies, altruistic parents will leave dowries to their daughters and bequests to their sons to mitigate a free-rider problem.[21] Furthermore, through most of the 19$^{\text{th}}$ Century, women completely relinquished control of their assets to their husbands. Under the doctrine of coverture, a husband owned any wages earned by his wife and any property she brought to the marriage (Geddes and Lueck, 2002).

These institutions imply that in the presence of a daughter, the G1 patriarch would have had little say over the allocation of resources between the second and third generations. The daughter was less likely to live in close proximity, and the rights on any transfers she received upon marriage would have been transferred to the husband. On the other hand, sons were more likely to receive a bequest only upon the patriarch's death. The ability of the patriarch to withhold the transfer of resources to his male offspring would have made it easier for him to monitor and influence the allocation of resources between son and grandchildren, guaranteeing a sufficiently high investment in G3's human capital.

This mechanism may have been relevant for most of the 19$^{\text{th}}$ Century but is less likely to have played a role during our sample period. As we documented above, virilocality was already on the decline in the second half of the 19$^{\text{th}}$ Century. Also, formal dowries were relatively

[20]Quasi-hyperbolic preferences have been made popular in recent years to model the intra-personal self-control problems in consumption and savings decisions and other contexts (Laibson, 1997; O'Donoghue and Rabin, 1999; DellaVigna and Paserman, 2005). However, one of the first applications of $\beta - \delta$ preferences (Phelps and Pollak, 1968) was to an intergenerational growth model that would be applicable here.

[21]Others emphasize consumption smoothing and the role of marital arrangements for solving agency problems (see for example Rosenzweig and Stark, 1989, based on data on rural India; and Fafchamps and Quisumbing, 2005a and 2005b, on rural Ethiopia). We investigate the insurance motive by running a regression that includes an interaction term between parent and grandparents income. We did not find any evidence that grandparents have a larger effect if parents are poorer, independent of G3 gender.

uncommon. Botticini and Siow (2003) show that in late 18th Century Connecticut, between 46 and 67 percent of married daughters were assigned inter vivos transfers from their family of origin, likely at the time of their marriage, but by the 1820s, only 40 percent received such transfers. Finally, most States abandoned the doctrine of coverture in the second half of the Century. By 1880, three quarters of the states allowed women the ownership and control over all property they brought to the marriage.

Therefore, it appears that most of the institutions that could have led to a greater importance of paternal grandparents were no longer prevalent by the time we begin our analysis. Furthermore, all of the above explanations do not differentiate by the gender of the grandchildren, and have difficulty in rationalizing our finding of a gender asymmetry in the relative importance of paternal and maternal grandparents. In the next section, we propose a formal model that can account for these differences.

4.2 Multi-trait Matching and Inheritance

We adapt the model of intergenerational mobility and multi-trait matching in Chen et al. (2013) to allow for multiple generations. Based on this model, the observed gender differences in social mobility can be rationalized based on asymmetries between market and non-market traits which we discuss below.

We assume that individuals are characterized by two distinct traits. The first (which we denote by y) directly affects an individual's earning potential, and includes elements such as cognitive skills or education. The second (denoted by x) can also affect earnings potential but to a lesser extent. It can include physical attractiveness, health, kindness and other attributes signaling reproductive capacity – all things that potentially have little impact on market productivity but are valued in the marriage market. For convenience, we refer to y as the "market" trait and x as the "non-market trait," even though we should keep in mind that both can matter for labor market outcomes (which is what we measure empirically).

The basic premise of the model is that individuals' attractiveness in the marriage market is also a function of both traits. Every individual is characterized by a unique index of attractiveness, which depends on the individual's x and y traits: $h_i^G(x_i, y_i) = x_i + \phi^G y_i$, $G = F, M$. Critically, we assume that there is an asymmetry in the relative importance of the two traits across genders. Specifically, the non-market trait x has higher weight in determining women's desirability, (i.e., $\phi^F < 1$), while the market trait y is more important for men ($\phi^M > 1$). This difference can be explained based on biological differences in reproductive roles and on the persistence of gender roles within households (see, for example, Buss, 1989, Eagly et al., 2000, 2004). Even today, evidence based on on-line dating and speed-dating shows that men and women value different attributes in prospective partners (see, for example, Fisman et al., 2006).[22]

The matching equilibrium in the marriage market features perfect assortative mating: the highest ranking man is matched with the highest ranking woman, the second highest man with the second highest woman, and so on. To further simplify matters, we assume that each trait can take only one of two levels: $x \in \{\underline{x}, \ \overline{x}\}$ and $y = \{\underline{y}, \ \overline{y}\}$.

Therefore, the equilibrium in the marriage market takes on a particularly simple form, summarized by the table below:

Ranking of Couples	Females	Males
1	$(\overline{x}, \overline{y})$	$(\overline{x}, \overline{y})$
2	$(\overline{x}, \underline{y})$	$(\underline{x}, \overline{y})$
3	$(\underline{x}, \overline{y})$	$(\overline{x}, \underline{y})$
4	$(\underline{x}, \underline{y})$	$(\underline{x}, \underline{y})$

There are four categories of individuals: men and women endowed with high levels of both traits (i.e., the highest ranked individuals) are paired with each other, as do men and women endowed with low levels of both traits (the lowest ranked individuals). However, in the middle two categories, there is some mixing: men with high levels of the market trait ($\overline{y}$) and low levels of the non-market trait ($\underline{x}$) are matched with women with low levels of the market trait and high levels of the non-market trait ($\underline{y}$ and $\overline{x}$), while men with ($\underline{y}$, $\overline{x}$) are matched with women with ($\overline{y}$, $\underline{x}$).

To understand the implications of this matching model for intergenerational mobility, we must consider how the two traits are transmitted across generations. We assume that for both traits, a child can either be endowed with the same level of the trait as his/her parent, or he/she can "switch" – i.e., if the parent is endowed with a high level of the trait, the child

[22] A handful of studies in economics has emphasized the importance of biological gender differentials on gender roles and market outcomes. See for example, Siow (1998) and Cox (2003).

will be endowed with a low level, and vice versa. Let π_x and π_y be the probabilities that, respectively, traits x and y "switch."[23] We capture the fact that traits are relatively persistent across generations by constraining the "switching" probabilities to be weakly smaller than 1/2. Clearly, lower values of the switching probabilities imply that a trait is highly persistent across generations.

The next key assumption is that the transmission of traits x, y is gender-segregated: specifically, we assume that the father passes on his traits to the son, and the mother passes on her traits to the daughter. While this assumption is clearly extreme (in reality it is likely that children inherit traits from both their parents), we view it as a convenient simplification, which captures the fact that children will be more inclined to view the parent of their same sex as a role model to imitate.[24]

Finally, we also assume that the transmission of the x and y traits are independent of each other and across genders. Putting everything together, we can derive a two-generation transition probability matrix where the (j,k) element is the probability of generation $t+1$ being in rank k conditional on generation t being in rank j. The two-generation transition matrices for men and women, Π_M and Π_F are defined as follows.

$$\Pi_M = \begin{bmatrix} (1-\pi_x)(1-\pi_y) & \pi_x(1-\pi_y) & (1-\pi_x)\pi_y & \pi_x\pi_y \\ \pi_x(1-\pi_y) & (1-\pi_x)(1-\pi_y) & \pi_x\pi_y & (1-\pi_x)\pi_y \\ (1-\pi_x)\pi_y & \pi_x\pi_y & (1-\pi_x)(1-\pi_y) & \pi_x(1-\pi_y) \\ \pi_x\pi_y & (1-\pi_x)\pi_y & \pi_x(1-\pi_y) & (1-\pi_x)(1-\pi_y) \end{bmatrix} \tag{1}$$

[23] It is straightforward to allow the switching probabilities differ both by trait and by gender, reflecting both institutional and biological factors. This richer model complicates the analysis somewhat, but does not add meaningfully to the economic insight.

[24] Classic studies in social psychology and psychoanalysis offer theoretical underpinning for the assertion that the transmission of certain traits is gendered. Maccoby and Jacklin (1974) argue that children emulate their parent of the same gender because such behavior is socially reinforced, or that children emulate the parent with whom they spend the most time, which is typically the parent of the same gender. Acock and Yang (1984) provide empirical support for some of the predictions of this theory. Chodorow (1978) suggests that mothers pass traits expressly related to "mothering" onto their daughters, which occurs because daughters are more likely to personally identify with their mother than their father. Boyd (1989) reviews alternative models of the mother-daughter relationship as well as empirical research on this topic. Lamb (1976) argue that fathers play a bigger role in the development of their sons than their daughters. More recently a number of studies in economics (Thomas, 1994), developmental psychology (Keller, 2002, Weinberg et al., 1999) and cultural and social anthropology (Godoy et al., 2006) provide empirical evidence of differential investment of fathers and mothers along gender lines.

$$
\Pi_F = \begin{bmatrix} (1-\pi_x)(1-\pi_y) & (1-\pi_x)\pi_y & \pi_x(1-\pi_y) & \pi_x\pi_y \\ (1-\pi_x)\pi_y & (1-\pi_x)(1-\pi_y) & \pi_x\pi_y & \pi_x(1-\pi_y) \\ \pi_x(1-\pi_y) & \pi_x\pi_y & (1-\pi_x)(1-\pi_y) & (1-\pi_x)\pi_y \\ \pi_x\pi_y & \pi_x(1-\pi_y) & (1-\pi_x)\pi_y & (1-\pi_x)(1-\pi_y) \end{bmatrix} \quad (2)
$$

Note that because of the nature of the matching equilibrium, the transition matrices are not identical for men and women. For example, a man born to the highest rank will move to the second highest rank only if the x trait switches and the y trait does not switch, an event that occurs with probability $\pi_y(1-\pi_x)$. On the other hand, a woman in the highest rank will move to the second highest rank only if trait x stays the same but trait y switches, an event that occurs with probability $(1-\pi_x)\pi_y$.

Based on these transition matrices we can obtain four three-generation transition probability matrices, whose (j,k) element is equal to the probability that a grandchild belongs to rank k, conditional on the grandfather belonging to rank j. There are four such matrices depending on the gender of the grandchild and the gender of the middle generation. These three generation matrices, which we denote with $\Omega_{g,G}$, for $g = \{M, F\}$ and $G = \{MAT, PAT\}$, are obtained from the product of the two-generation matrices:

$$
\begin{aligned}
\Omega_{M,PAT} &= \Pi_M\Pi_M \\
\Omega_{M,MAT} &= \Pi_F\Pi_M \\
\Omega_{F,PAT} &= \Pi_M\Pi_F \\
\Omega_{F,MAT} &= \Pi_F\Pi_F.
\end{aligned}
$$

Based on these matrices, we can calculate the three-generation rank correlations, $\rho_{g,G}$:

$$
\rho_{g,G} = \frac{\frac{1}{4}r'\Omega_{g,G}r - E(R)^2}{V(R)},
$$

where $r = \begin{pmatrix} 1 & 2 & 3 & 4 \end{pmatrix}'$ and R is the random variable denoting an individual's rank, and has a discrete uniform distribution between 1 and 4. Explicit formulas for these correlations are presented in Appendix A.

We are interested in finding how the relative size of the three-generation correlations depend on π_x and π_y. Specifically, we look at three differences. The first, $(\rho_{M,PAT} - \rho_{F,MAT})$, compares the three-generation correlation going through the male line (grandfather-father-son) to the one going through the female line (grandmother-mother-daughter). The second, $(\rho_{M,PAT} - \rho_{M,MAT})$, compares the effect of paternal and maternal grandfathers on grandsons. The third, $(\rho_{F,MAT} - \rho_{F,PAT})$, compares the effect of paternal and maternal grandmothers on granddaughters.

Figure 3 displays how the signs of these three differences vary over the parameter space.[25] The difference $(\rho_{M,PAT} - \rho_{F,MAT})$ depends only on the difference between π_x and π_y. It is positive in the area below the 45 degree line (i.e. $\pi_y < \pi_x$), meaning that the y-trait is relatively more persistent than the x-trait. Intuitively, males, whose social status depends more heavily on the market trait y, are more likely to preserve the ranking of their fathers and paternal grandfathers if this trait is relatively persistent.

The difference $(\rho_{M,PAT} - \rho_{M,MAT})$ is positive in three out of the four zones marked in the figure. First, consider zone I and II, where the y-trait is more persistent than the x-trait. As an example, we ask what these values imply for the descendants of a grandfather who has high levels of both the x and y traits, and therefore belongs to the highest rank. The low value of π_y implies that G2 sons are likely to maintain the high value of the market trait, and therefore are likely to remain in one of the top two ranks. Since the traits are passed along the male line, the grandson is also likely to stay in one of the top two ranks. Hence, the correlation between grandson and paternal grandfather is likely to be high. Compare this to the outcome of the maternal grandson (the son of a G2 female). The G2 daughter inherits her traits from her mother, who, because of perfect assortative mating, is also endowed with high levels of both x and y. The relatively high value of π_x implies that the G2 daughter has a relatively high probability of ending up in the third rank, characterized by low levels of non-market skills ($\underline{x}$) and high levels of market skills ($\overline{y}$), and will therefore marry an $\left(\overline{x}, \underline{y}\right)$husband. But then, the male grandson will inherit the low levels of the y trait from his father, and remain in one of the two lowest ranks. Within two generations, the maternal grandson will have experienced considerable downward mobility in economic status.

The difference $(\rho_{M,PAT} - \rho_{M,MAT})$ is also positive in zone IV, characterized by π_x being

[25]See Appendix A for calculations. We show that the four zones are defined by the following inequalities: in Zone I, $\pi_y < -\frac{3}{2} + 4\pi_x$; in Zone II, $-\frac{3}{2} + 4\pi_x < \pi_y < \pi_x$; in Zone III, $\pi_x < \pi_y < \frac{3}{8} + \frac{1}{4}\pi_x$; in Zone IV, $\pi_y > \frac{3}{8} + \frac{1}{4}\pi_x$.

strongly more persistent than π_y. This might seem counterintuitive but an example along the lines of the previous one illustrates how this pattern may arise. Let's start again from a grandfather who has high levels of both traits. Because of the high value of π_x and the low value of π_y, the G2 son will likely be in either the first or the third rank: he maintains the high level of the non-market traits, but may lose his market advantage. In turn, the paternal grandson will also be in either the first or the third rank. Take instead a maternal grandson. The G2 daughter is likely to maintain a higher level of x and remain in either the first or second rank. Her husband, will have a high level of y but could have either high or low level of x. Because x is persistent and y is not, the maternal grandson will have rank 1 or 3 if his father has a high x trait, but rank 2 or 4 if his father has a low x trait. The end result is that the maternal grandson is more likely to be more removed from his $(\overline{x}, \overline{y})$ grandfather than the paternal grandson, even if the non-market trait, dominant along the female line, is more persistent. Similar arguments apply to descendants of grandfathers who start out in one of the other categories.

Finally, the difference in the effect of maternal and paternal grandfathers for granddaughters, $(\rho_{F,MAT} - \rho_{F,PAT})$, is positive in zones I, III and IV. The intuition of this result mirrors exactly the one we just provided for grandsons, but with the roles of π_x and π_y reversed: either π_x is more persistent than π_y or π_y is extremely persistent while π_x is not.

The empirical pattern of inequalities that we find in the data is consistent with the parameters π_x and π_y being in Zone I. In this area the switching probability for the x trait (non-market skills) must be sufficiently high, while the switching probability for the y trait (market skills) must be relatively low. This asymmetry in the degree of inheritability of market and non-market traits can be justified on the basis of potential differences in the importance of parental investment. Parents with a high level of the market trait have higher disposable income and invest more in their children. If the market trait (e.g., education) is more amenable to parental investment than the non-market trait (e.g., physical appearance or reproductive ability), it follows that the market trait is more persistent.[26] Teaching a child how to read and write may be easier than manipulating his or her reproductive ability.[27] The institutional

[26] See Mailath and Postlewaite, 2006, for a theoretical justification of the link between parental investment and the persistence of market and non-market traits.

[27] Modern studies have found that the intergenerational correlation in health, while positive, tends to be a fair bit smaller than the intergenerational correlation in income (Currie and Moretti, 2007).

set up might also reinforce this mechanism. For example, if private and public investment are complementary in the production of human capital, an increase in public spending in education could result in an even greater parental investment by wealthier parents and a stronger persistence in the market trait.[28]

In short, with a relatively parsimonious set of assumptions, our simple model is able to deliver a rich set of predictions that matches the pattern of intergenerational correlations that is observed in the data.

5 Case Study: Regional Differences

It is possible that the relative heritability of market and non-market traits differs by region. If this is the case, then our baseline results may mask regional heterogeneity in the G1-G3 transmission process. Table 8 presents the results obtained when we estimate the model separately by region of residence at 20 year intervals (Panel A) and at 30 year intervals (Panel B). The results reveal marked regional differences in the role of gender in the transmission of economic status across generation. We discuss primarily the differences between the Northeast (columns 1 and 2) and the South (columns 5 and 6). In the Midwest (columns 3 and 4) most of the gender differences are insignificant, and therefore it is difficult to make strong claims about this region.

In the Northeast, the strongest relationship is the one between grandsons and paternal grandfathers. Furthermore, paternal grandfathers seem to have an effect on their grandchildren of both genders, while the effect of maternal grandfathers is never significant. Finally, grandsons are more strongly affected by their grandfathers on both sides. In contrast, in the South the strongest relationship is the one between granddaughters and maternal grandfathers. Both grandfathers have a significant impact on the outcomes of grandchildren of both genders. However, maternal grandfathers clearly matter more for granddaughters than paternal grandfathers, while the evidence for grandsons is more mixed. Finally, maternal grandfathers have a significantly larger impact on granddaughters than grandsons. This evidence suggests that in the South the chain of intergenerational transmission is stronger along the maternal line, while the paternal line seems the more dominant in the Northeast.

[28] In fact, Parman (2011) argues that in the early 20th Century, the wealthy were better able to take advantage of the expansion of public schooling.

We can interpret these differences in the context of the multi-trait matching model and the combination of parameters highlighted in Figure 3. Our results are consistent with π_y and π_x lying in zone II in the Northeast, and in zone III or IV in the South (whether we are in zone III or IV depends on the sign of $\rho_{M,PAT} - \rho_{M,MAT}$, which is not determined empirically).

These patterns can be generated by regional differences in the process by which either the x trait or the y trait is passed along generations. First, fixing π_x, it could be that the probability of having a y trait that is different from your parent is *higher* in the South ($\pi_y^{SOUTH} > \pi_y^{NORTH}$). Alternatively, fixing π_y, it could be that $\pi_x^{SOUTH} < \pi_x^{NORTH}$, meaning that the probability of having an x trait that is different from your parent is *lower* in the South. These are both plausible conjectures, given what we know about these regions during the period under investigation.

One reason for the market trait y to be "stickier" in the Northeast than the South is that the South experienced more industrial upheaval during the early 20th century – the time frame from which all of our G3 samples are drawn – than the Northeast did. In particular, the South experienced a large decline in the prevalence of agriculture between 1900 and 1940. In 1900, approximately 60% of the southern workforce was engaged in agriculture; by 1940, this figure was less than 30%. In contrast, the fraction of the northeastern workforce engaged in agriculture fell from 15% to 5% between 1900 and 1940, a much smaller absolute decline.[29] The South was converging with the rest of the country in terms of industrial composition during this period (Kim and Margo, 2004), which might mean that there was more mobility – in terms of market traits – in the South than in the Northeast. This is especially likely if occupational or industrial knowledge is one of the market traits that fathers pass on to their sons.

The other potential explanation for the differences between the Northeast and the South is that non-market traits (x) – such as kindness, attractiveness, and reproductive or parenting ability – are "stickier" in the South than the Northeast. Historians characterize the South as highly conservative with respect to gender roles. Scott (1970, p. 4) describes the 'ideal' antebellum southern woman as "a submissive wife whose reason for being was to love, honor, obey, and occasionally amuse her husband, to bring up his children and to manage his household." This persisted through the 19th and 20th centuries: southern states were

[29]These figures are based on the authors' calculations using census data (Ruggles et al 2010).

slow to adopt legislation expanding women's property rights during the 19th century (Kahn 1996), and were largely resistant to women's suffrage in the early 20th century (Green 1997). Looking more recently, researchers have found that while southerners' attitudes toward gender roles had started to converge with the rest of country by the late 20th century, there was still a significant gap (Rice and Coate 1995; Hurlbert 1988). In terms of the model, this difference in gender roles may lead to a higher persistence of the non-market trait in the South. If women spent more time "mothering" in the South, and if "mothering ability" is an important non-market trait, then it could be passed along more persistently in that region.

6 Conclusion

In this paper, we have estimated intergenerational elasticities across three generations for the US spanning the late 19$^{\text{th}}$ and early 20$^{\text{th}}$ Century, focusing on the differential role of maternal and paternal grandparents on both granddaughters and grandsons.

We find that both paternal and maternal grandparents have a significant effect on the grandchildren's outcome, above and beyond the direct effect of the fathers. Moreover, the transmission of economic status follows gendered lines: grandsons are more strongly affected by paternal grandfathers, while maternal grandfathers matter more for the outcomes of granddaughters. We interpret these results in light of a matching model where an individual's socioeconomic status is determined by two traits that are inherited from the same sex parent and whose relative importance varies by gender.

Our results can have important implications for our understanding of the persistence of socioeconomic status over the long run. Recent studies have shown that grandfathers, typically paternal ones, have a distinct impact on the outcomes of their grandchildren. Our contribution is to show that maternal grandparents also matter, pointing to the key role of mothers for the transmission of status across multiple generations. The upshot of this result is that stratification in marriage by social class might amplify cross-sectional inequality and lead to lower mobility across generations, even in a context in which married women contributed little to household income.

References

[1] 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.

[2] Acock, Alan C. and Wen Shan Yang (1984). "Parental Power and Adolescents' Parental Identification." *Journal of Marriage and Family.* 46(2), pp. 487-495.

[3] Alston, Lee J. and Morton Owen Schapiro (1984). "Inheritance Laws Across Colonies: Causes and Consequences." *Journal of Economic History.* 44(2): 277-287.

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

[5] 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.

[6] Botticini, Maristella and Aloysius Siow (2003). "Why Dowries?" *American Economic Review*, 93(4), pp. 1385-1398.

[7] Boyd, Carol J. (1989). "Mothers and Daughters: A Discussion of Theory and Research." *Journal of Marriage and Family.* 51(2), pp. 291-301.

[8] Buss, D.M. (1989). Sex differences in human mate preferences: evolutionary hypotheses tested in 37 cultures. *Behavioural and Brain Sciences*, 12: 1-49.

[9] Chen, Natalie, Conconi, Paola and Carlo Perroni (2013). "Multi-Trait Matching and Gender Differentials in Intergenerational Mobility." *Economic Letters*, 102(2), pp. 292-296.

[10] Chetty, Raj, Hendren, Nathaniel, Kline, Patrick, Saez, Emmanuel and Nicholas Turner (2014a). "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." *American Economic Review*, 104(5), pp. 141-147.

[11] Chetty, Raj, Hendren, Nathaniel, Kline, Patrick and Emmanuel Saez (2014b). "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." *Quarterly Journal of Economics*, 129(4), pp. 1553-1623.

[12] Chodorow, Nancy (1978). *The Reproduction of Mothering.* Berkeley: University of California Press.

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

[14] Collado, Dolores M., Ignacio Ortuño Ortín, and Andrés Romeu (2012). "Intergenerational Linkages in Consumption Patterns and the Geographic Distribution of Surnames." *Regional Science and Urban Economics.* 42(1-2): 341-350.

[15] Cox, Donald (2003). "Private Transfers within the Family: Mothers, Fathers, Sons and Daughters," in (Alicia H. Munnell and Annika Sunden, eds.) *Death and Dollars: The Role of Gifts and Bequests in America*, Brookings Institution.

[16] Currie, Janel and Enrico Moretti (2007). "Biology as Destiny? Short- and Long-Run Determinants of Intergenerational Transmission of Birth Weight." *Journal of Labor Economics.* 25: 231-264.

[17] Della Vigna, Stefano and M. Daniele Paserman (2005). "Job Search and Impatience." *Journal of Labor Economics*, 23(3), pp. 527-588.

[18] Duflo, Esther (2003) "Grandmothers and Granddaughters: Old-Age Pensions and Intrahousehold Allocation in South Africa," *World Bank Economic Review*, 17 (1), pp. 1-25.

[19] Eagly, A.H., W. Wood and A. Diekman (2000). "Social Role Theory of Sex Differences and Similarities: A Current Appraisal." In *Psychology of Gender,* Eds. A.H. Eagly, A. Beall, and R.J. Sternberg, New York: Guilford.

[20] Eagly, A.H., W. Wood, and M.C. Johannesen-Schmidt (2004). "The Social Role Theory of Sex Differences and Similarities: Implications for Partner Preference."' In *Psychology of Gender,* Eds. A.H. Eagly, A. Beall, and R.J. Sternberg, 2nd Ed., New York: Guilford.

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

[22] Fafchamps, Marcel and Agnes Quisumbing (2005a). "Assets at Marriage in Rural Ethiopia." *Journal of Development Economics*, 77(1), pp. 1-25.

[23] Fafchamps, Marcel and Agnes Quisumbing (2005b). "Marriage, Bequest, and Assortative Matching in Rural Ethiopia." *Economic Development and Cultural Change*, 53(2), pp. 347-380.

[24] Ferrie, Joseph P. and Jason Long (2015). "Grandfathers Matter(ed): Occupational Mobility Across Three Generations in the U.S. and Britain, 1850-1910." Working Paper.

[25] Fisman, Raymond; Sheena S. Iyengar; Emir Kamenica and Itamar Simonson. (2006) "Gender Differences in Mate Selection: Evidence from a Speed Dating Experiment." *Quarterly Journal of Economics*, 121(2), pp. 673-697.

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

[27] Geddes, Rick and Dean Lueck. "The Gains from Self-Ownership and the Expansion of Women's Rights." *American Economic Review.* 92(4): 1079-1092.

[28] Godoy, Richard, Victoria Reyes-Garcia, Thomas McDade, Susan Tanner, William R. Leonard, Tomas Huanca, Vincent Vadez, Karishma Patel. 2006. "Why do mothers favor girls and fathers, boys? A hypothesis and a test of investment disparity," *Human Nature*, 17(2):169-189.

[29] Green, Elna (1997). *Southern Strategies: Southern Women and the Woman Suffrage Question.* Baltimore: Johns Hopkins University Press.

[30] Güell, Maia, José V. Rodríguez-Mora, and Christopher I.Telmer (2014). "The Informational Content of Surnames, the Evolution of Intergenerational Mobility and Assortative Mating." Working paper.

[31] Hurlbert, Jeanne S. (1988). "The Southern Region: A Test of the Hypothesis of Cultural Distinctiveness." *The Sociological Quarterly.* 30(2): 245-266.

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

[33] Kahn, B. Zorina (1996). "Married Women's Property Laws and Female Commercial Activity: Evidence from United States Patent Records." *Journal of Economic History.* 56(2): 356-388.

[34] Keller, Heidi (2002) "Gender and Birth Order as Determinants of Parental Behavior." *International Journal of Behavioral Development* 26: 177-184.

[35] Kim, Sukkoo and Robert A. Margo (2004). "Historical Perspectives on U.S. Economic Geography." In *Handbook of Regional and Urban Economics,* Eds. J.V. Henderson and J.F. Thisse, Ed. 1, Vol. 4, Ch. 66, pp 2981-3019.

[36] Laibson, David I. (1997). "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, 112(2), pp. 443-477.

[37] Lamb, Michael E., ed. (2004). *The role of the father in child development.* John Wiley and Sons.

[38] Lindahl, Mikael; Marten Pålme; Sophia Sandgren Massih and Anna Sjögren. (2015). "Long-Term Intergenerational Persistence of Human Capital: An Empirical Analysis of Four Generations." *Journal of Human Resources*, 50 (1), pp. 1-33 (Winter).

[39] Macoby, Eleanor E. and Carol N. Jacklin (1974). *The Psychology of Sex Differences.* Stanford: Stanford University Press.

[40] Mailath, George J. and Andrew Postlewaite (1996). "Social Assets." *International Economic Review*, 47(4), pp. 1057-1091.

[41] Mare, Robert D. (2011). "A Multigenerational View of Inequality." *Demography.* 48: 1-23.

[42] O'Donoghue, Ted and Matthew Rabin (1999). "Doing It Now or Later." *American Economic Review*, 89(1), pp. 103-124.

[43] Olivetti, Claudia and M. Daniele Paserman (2015). "In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940." *American Economic Review*, 105(8), pp. 2695-2724.

[44] Parman, John (2011). "American Mobility and the Expansion of Public Education." *Journal of Economic History.* 71(0): 105-132.

[45] Phelps, Edmund S. and Robert A. Pollak (1968). "On Second-Best National Saving and Game-Equilibrium Growth." *Review of Economic Studies*, 35(2), pp. 185-199.

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

[47] Posadas, Josefina and Marian Vidal-Fernandez (2013). "Grandparents' Childcare and Female Labor Force Participation." *IZA Journal of Labor Policy.* 2:14.

[48] Rice, T.W. and D. L. Coates (1995). "Gender Role Attitudes in the Southern United States." *Gender and Society.* 9: 738-750.

[49] Rosenzweig, Mark R. and Oded Stark (1989). "Consumption Smoothing, Migration, and Marriage: Evidence from Rural India." *Journal of Political Economy*, 97(4), pp. 905-926.

[50] 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.

[51] Scott, A.F. (1970). *The Southern Lady: From Pedestal to Politics, 1830 to 1930.* Chicago: University of Chicago Press.

[52] Shammas, Carole, Marylynn Salmon and Michel Dahlin (1987). *Inheritance in America.* New Brunswick: Rutgers University Press.

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

[54] 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.

[55] Solon, Gary R. (2013). "Theoretical Models of Inequality Transmission Across Multiple Generations." NBER Working Paper no. 18790.

[56] Thomas, Duncan (1994). "Like Father, like Son; Like Mother, like Daughter: Parental Resources and Child Height." *Journal of Human Resources.* 29(4): 950-988.

[57] Weinberg, M. Katherine, Tronick, Edward Z., and Cohn, Jeffrey F. (1999). "Gender Differences in Emotional Expressivity and Self-Regulation during Early Infancy." *Developmental Psychology* 35 (1): 175-188.

[58] Zeng, Zhen and Yu Xie (2014). "The Effects of Grandparents on Children's Schooling: Evidence from Rural China." *Demography.* 51(2): 599-617.

7 Tables and Figures

Table 1: Grandsons and Paternal Grandfathers
Right hand side variable: Log occupational income

	(1)	(2)	(3)	(4)	(5)	(6)
	1860-1880-1900			*1850-1880-1910*		
Panel A: Linked data. Occupational Income: G2 actual, G3 imputed						
G1 Paternal		0.2417	0.0829		0.1689	0.0933
		(0.067)	(0.064)		(0.047)	(0.043)
G2	0.5268		0.5237	0.4233		0.4208
	(0.023)		(0.023)	(0.019)		(0.019)
Constant	1.3619	2.1940	1.1289	1.8495	2.6031	1.5846
	(0.069)	(0.196)	(0.185)	(0.057)	(0.137)	(0.130)
Observations	2,763	2,763	2,763	4,007	4,007	4,007
Panel B. 1% sample. Occupational income: G2 and G3 both imputed						
G1 Paternal		0.6022	0.2568		0.2918	0.1152
		(0.101)	(0.097)		(0.089)	(0.072)
G2	0.2905		0.2679	0.3010		0.2918
	(0.035)		(0.037)	(0.031)		(0.032)
Constant	2.1354	1.2434	1.4582	2.2116	2.2560	1.9047
	(0.102)	(0.292)	(0.260)	(0.092)	(0.257)	(0.211)
Observations	77,883	77,902	77,878	82,060	82,070	82,055

Notes. Panel A displays results from OLS regressions of individual G3 log occupational score on G2 log occupational score and G1 log occupational score, imputed as the average G1 log occupational score for each G2 individual's first name. G3 and G2 data come from the IPUMS linked representative samples from 1880-1900 or 1880-1910; G1 data comes from the 1860 or 1850 IPUMS 1% sample. Panel B displays results from OLS regressions of individual G3 log occupational score on G1 and G2 log occupational score, imputed as the average G1 and G2 log occupational score for each G3 individual's first name. All data come from the IPUMS 1% samples for 1850,1860, 1880, 1900 and 1910. In columns (1)-(3), the G3 sample consists of men age 20-35 in 1900; the G2 sample consists of men age 20-35 in 1880 who have children ages 0-15; the G1 sample consists of men in 1860 who have children ages 0-15. In columns (4)-(6), the samples are constructed similarly, using 30-45 year olds in the 1850, 1880, and 1910 censuses.

Table 2: Grandsons and Paternal Grandfathers
Right hand side variable: Percentile rank of log occupational income

	(1)	(2)	(3)	(4)	(5)	(6)
	1860-1880-1900			*1850-1880-1910*		
Panel A: Linked data. Occupational Income: G2 actual, G3 imputed						
G1 Paternal		0.1524	0.0300		0.1619	0.0710
		(0.042)	(0.043)		(0.037)	(0.035)
G2	0.4273		0.4183	0.3682		0.3469
	(0.041)		(0.044)	(0.038)		(0.040)
Constant	2.6519	2.8090	2.6394	2.8845	2.9984	2.8540
	(0.026)	(0.027)	(0.030)	(0.024)	(0.023)	(0.026)
Observations	2,763	2,763	2,763	4,007	4,007	4,007
Panel B. 1% samples. Occupational income: G2 and G3 both imputed						
G1 Paternal		0.1517	0.0690		0.1286	0.0534
		(0.021)	(0.022)		(0.022)	(0.023)
G2	0.2558		0.2185	0.2650		0.2376
	(0.027)		(0.030)	(0.025)		(0.027)
Constant	2.8396	2.8978	2.8204	2.9456	3.0264	2.9301
	(0.015)	(0.012)	(0.016)	(0.014)	(0.013)	(0.016)
Observations	77,883	77,902	77,878	82,060	82,070	82,055

Notes. Panel A displays results from OLS regressions of individual G3 log occupational score on G2 log occupational score and G1 log occupational score, imputed as the average G1 log occupational score for each G2 individual's first name. G3 and G2 data come from the IPUMS linked representative samples from 1880-1900 or 1880-1910; G1 data comes from the 1860 or 1850 IPUMS 1% sample. Panel B displays results from OLS regressions of individual G3 log occupational score on percentile ranks of log occupational scores for G2 and G1 (both imputed). For details on data sources and sample restrictions, see note to Table 1.

Table 3: Intergenerational Elasticities Across Three Generations:
Regressions at 20 Year Intervals

	(1)	(2)	(3)	(4)	(5)	(6)
	1860-1880-1900		*1880-1900-1920*		*1900-1920-1940*	
Panel A: G3 Male						
G1 Paternal	0.0690	-	0.0967	-	0.1391	-
	(0.022)		(0.020)		(0.018)	
G1 Maternal	-	0.0106	-	0.0566	-	0.1205
		(0.024)		(0.021)		(0.017)
G2	0.2185	0.2659	0.3300	0.3642	0.2610	0.2849
	(0.030)	(0.028)	(0.022)	(0.022)	(0.020)	(0.019)
Constant	2.8204	2.8271	2.8172	2.8163	2.8400	2.8301
	(0.016)	(0.017)	(0.017)	(0.016)	(0.012)	(0.012)
Observations	77,878	78,634	106,019	107,047	116,210	117,269
Panel B: G3 Female						
G1 Paternal	0.0891	-	0.0613	-	0.0703	-
	(0.020)		(0.019)		(0.016)	
G1 Maternal	-	0.0661	-	0.0944	-	0.1033
		(0.024)		(0.020)		(0.015)
G2	0.2680	0.3057	0.3642	0.3602	0.2882	0.2705
	(0.025)	(0.024)	(0.020)	(0.020)	(0.015)	(0.015)
Constant	2.8492	2.8333	2.8995	2.8772	2.9686	2.9545
	(0.017)	(0.017)	(0.013)	(0.012)	(0.009)	(0.008)
Observations	44,292	44,930	66,324	67,204	74,857	75,633

Notes. Contains results from OLS regressions of individual G3 log occupational score on the percentile rank of imputed G2 and G1 log occupational score, imputed as the average G2 or G1 log occupational score for each G3 individual's first name. For women, log occupational score is measured as the log occupational score of her husband. Panel A reports the results for G3 males using our three samples constructed at 20 year intervals, and panel B reports similar results for G3 females. The G3 sample consists of adults age 20-35 in the third sample year (1900, 1920 or 1940); the G2 sample consists of adults age 20-35 in the second sample year (1880, 1900 or 1920) who have children ages 0-15; the G1 sample consists of men in the first sample year (1850, 1880 or 1900) who have children ages 0-15. Standard errors are clustered by G3 individual's first name.

Table 4: Intergenerational Elasticities Across Three Generations:
Regressions at 30 year intervals

	(1)	(2)	(3)	(4)	(5)	(6)
	1850-1880-1910		*1870-1900-1930*		*1880-1910-1940*	
	Panel A: G3 Male					
G1 Paternal	0.0534	-	0.0758	-	0.1073	-
	(0.023)		(0.019)		(0.014)	
G1 Maternal	-	0.0411	-	0.0393	-	0.0445
		(0.020)		(0.020)		(0.017)
G2	0.2376	0.2532	0.2679	0.2971	0.2935	0.3355
	(0.027)	(0.024)	(0.019)	(0.019)	(0.018)	(0.019)
Constant	2.9301	2.9244	2.9698	2.9744	2.9395	2.9493
	(0.016)	(0.015)	(0.015)	(0.015)	(0.012)	(0.013)
Observations	82,055	82,179	114,905	114,949	106,458	106,403
	Panel B: G3 Female					
G1 Paternal	0.0227	-	0.0620	-	0.0441	-
	(0.027)		(0.017)		(0.015)	
G1 Maternal	-	0.0859	-	0.0630	-	0.0656
		(0.028)		(0.021)		(0.018)
G2	0.2823	0.2359	0.3466	0.3459	0.3368	0.3233
	(0.027)	(0.025)	(0.018)	(0.019)	(0.015)	(0.017)
Constant	2.9391	2.9269	2.9618	2.9628	2.9888	2.9833
	(0.021)	(0.022)	(0.012)	(0.012)	(0.010)	(0.010)
Observations	55,554	55,631	85,697	85,669	80,612	80,534

Notes. Contains results from OLS regressions of individual G3 log occupational score on the percentile rank of imputed G2 and G1 log occupational score, imputed as the average G2 or G1 log occupational score for each G3 individual's first name. For women, log occupational score is measured as the log occupational score of her husband. Panel A reports the results for G3 males using our three samples constructed at 30 year intervals, and panel B reports similar results for G3 females. The G3 sample consists of adults age 30-45 in the third sample year (1910, 1930 or 1940); the G2 sample consists of adults age 30-45 in the second sample year (1880, 1900 or 1910) who have children ages 0-15; the G1 sample consists of men in the first sample year (1850, 1870 or 1880) who have children ages 0-15. Standard errors are clustered by G3 individual's first name.

Table 5: Summary of G1-G3 Intergenerational Income Elasticities using Different Sample Restrictions and Wage Measures

	(1)	(2)	(3)	(4)	(5)
	Dependent variable: Intergenerational income elasticity: G1-G3				
	All	G3 Male	G3 Female	G1 Paternal	G1 Maternal
G1 Paternal	0.0054	0.0334***	-0.0226**		
	(0.008)	(0.005)	(0.008)		
G3 Male	0.0153*			0.0433***	-0.0127
	(0.008)			(0.008)	(0.008)
Second sample (G3=1920 or 1910)	0.0169*	0.0221***	0.0117	0.0196*	0.0142
	(0.010)	(0.006)	(0.011)	(0.010)	(0.009)
Third sample (G3=1940)	0.0236**	0.0408***	0.0064	0.0259**	0.0213*
	(0.011)	(0.008)	(0.011)	(0.010)	(0.012)
Interval = 30 years	-0.0284***	-0.0306***	-0.0261***	-0.0319***	-0.0248**
	(0.008)	(0.005)	(0.008)	(0.008)	(0.008)
Specification details:					
G2 spouse in same age bracket	-0.0048**	-0.0028	-0.0068**	-0.0041	-0.0055*
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
G3 married	0.0092**	0.0184**		0.0063	0.0121*
	(0.004)	(0.006)		(0.004)	(0.007)
G3 spouse in same age bracket	0.0086**	0.0154**	0.0018	0.0069*	0.0103
	(0.003)	(0.006)	(0.002)	(0.004)	(0.006)
Constant	0.0608***	0.0495***	0.0875***	0.0535***	0.0735***
	(0.012)	(0.011)	(0.012)	(0.014)	(0.012)
Observations	144	72	72	72	72

Notes The dependent variable in each of these regressions is our estimated G1-G3 intergenerational elasticity under different specifications. All G1-G3 elasticities are taken from OLS regressions of G3 log occupational score on the percentile rank of imputed scores of G2 and G1 (see tables 3 and 4 for additional details). These elasticities are estimated for combinations of 2 G2 genders , 2 G3 genders, 3 sample periods, and 2 intervals at which samples are constructed (20 or 30 years), 2 sample restrictions on G2 (baseline, or both spouses in the same age bracket), and 3 sample restrictions on G3 (baseline, married, or married and both spouses in the same age bracket). Column (1) contains elasticities from all specifications ($2 \times 2 \times 3 \times 2 \times 2 \times 3 = 144$ total); the remaining columns contain elasticities for a single G2 or G3 gender. Standard errors are in parentheses.

Table 6: Intergenerational Elasticities Across Three Generations:
Percentile Rank Regressions with Paternal and Maternal Grandfathers

	(1)	(2)	(3)	(4)
	20-year intervals		*30-year intervals*	
	G3 Male	G3 Female	G3 Male	G3 Female
G1 paternal	0.0981***	0.0478***	0.0688***	0.0229**
	(0.013)	(0.011)	(0.011)	(0.012)
G1 maternal	0.0267**	0.0716***	0.0162	0.0537***
	(0.013)	(0.012)	(0.011)	(0.012)
G2	0.2551***	0.2739***	0.2597***	0.3016***
	(0.014)	(0.012)	(0.012)	(0.012)
Observations	298,426	184,468	300,019	219,214
p (G1 paternal = G1 maternal)	0.001	0.167	0.002	0.081
p (G1 pat [G3 male] = G1 pat [G3 female])	0.002		0.0034	
p (G1 mat [G3 male] = G1 mat [G3 female])	0.012		0.026	

Notes Contains results from OLS regressions of individual G3 log occupational score on the percentile rank of imputed scores of G2, paternal G1 and maternal G1; these are imputed as the average for each G3 individual's first name. For women, log occupational score is measured as the log occupational score of her husband. Columns (1) and (2) pool our three samples constructed at 20 year intervals, including decade controls; columns (3) and (4) pool our samples constructed at 30 year intervals. The G3 sample consists of adults age 20-35 (or 30-45) in the third sample year; the G2 sample consists of adults age 20-35 (or 30-45) in the second sample year, who have children ages 0-15 and are married to spouse in the same age bracket; the G1 sample consists of men in the first sample year who have children ages 0-15. Standard errors are clustered by G3 first name - decade groups.

Table 7: Intergenerational Mobility Across Three Generations
Alternative Occupational Wage Measures

	(1)	(2)	(3)	(4)
Panel A: 1900 wage distribution				
	20-year intervals		*30-year intervals*	
	G3 Male	G3 Female	G3 Male	G3 Female
G1 paternal	0.0627***	0.0495***	0.0713***	0.0410***
	(0.010)	(0.010)	(0.010)	(0.010)
G1 maternal	0.0308***	0.0601***	0.0201*	0.0421***
	(0.011)	(0.010)	(0.012)	(0.011)
G2	0.1996***	0.1965***	0.2222***	0.2340***
	(0.010)	(0.010)	(0.011)	(0.010)
Observations	304,261	184,924	303,339	220,357
p (G1 paternal = G1 maternal)	0.060	0.460	0.001	0.940
p (G1 pat [G3 male] = G1 pat [G3 female])	0.356		0.0349	
p (G1 mat [G3 male] = G1 mat [G3 female])	0.049		0.178	
Panel B: Wage distribution based on adjusted average personal property by occupation in 1860 and 1870				
	20-year intervals		*30-year intervals*	
	G3 Male	G3 Female	G3 Male	G3 Female
G1 paternal	0.1533***	0.1074***	0.2200***	0.1187***
	(0.027)	(0.028)	(0.029)	(0.026)
G1 maternal	0.1143***	0.1409***	0.0641*	0.0858***
	(0.026)	(0.025)	(0.033)	(0.028)
G2	0.2518***	0.3054***	0.3357***	0.3621***
	(0.023)	(0.022)	(0.026)	(0.021)
Observations	280,461	177,079	284,666	210,096
p (G1 paternal = G1 maternal)	0.361	0.381	0.001	0.403
p (G1 pat [G3 male] = G1 pat [G3 female])	0.236		0.011	
p (G1 mat [G3 male] = G1 mat [G3 female])	0.463		0.617	

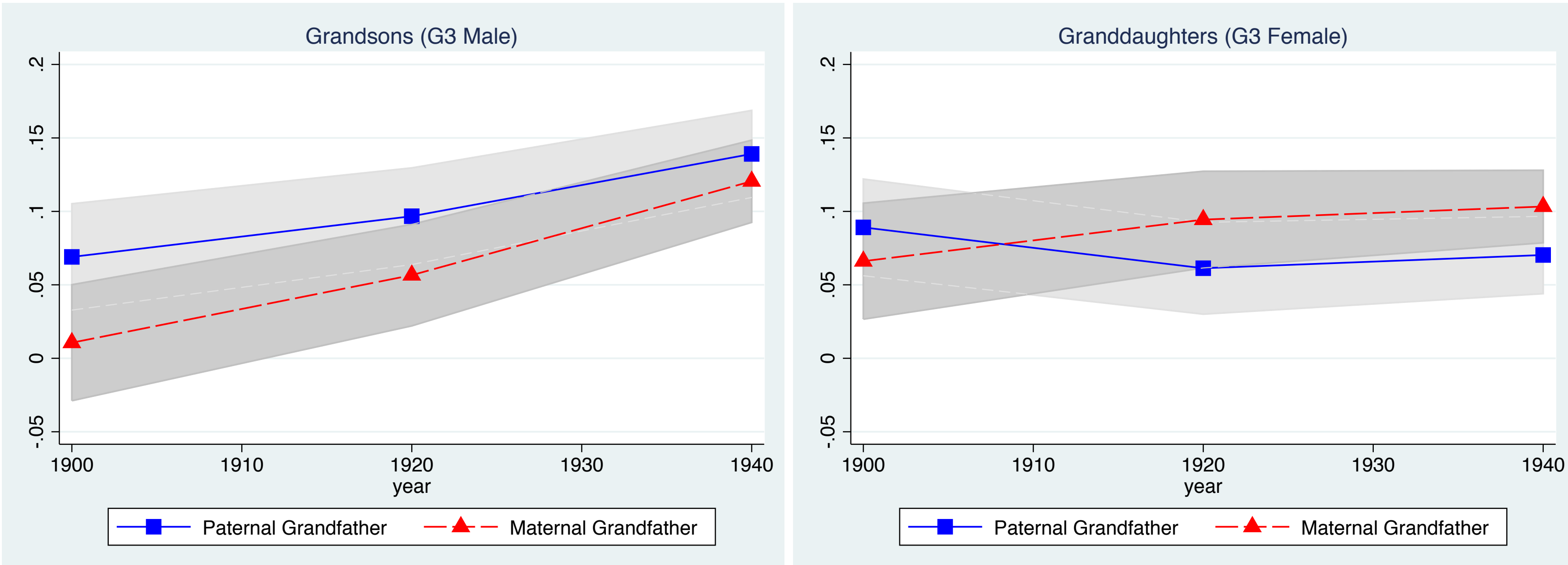
Notes Contains results from OLS regressions of individual G3 log occupational score on the percentile rank of imputed scores of G2, paternal G1 and maternal G1; these are imputed as the average for each G3 individual's first name. For women, log occupational score is measured as the log occupational score of her husband. Columns (1) and (2) pool our three samples constructed at 20 year intervals, including decade controls; columns (3) and (4) pool our samples constructed at 30 year intervals. Panel A measures occupational income using the 1900 wage distribution with an imputed wage for farmers (Preston and Haines 1991; Abramitzky et al 2012; Olivetti and Paserman 2013). Panel B measures occupational income using mean personal wealth by occupation in 1860 and 1870, adjusting the wealth of farmers downward by the average value of farm equipment and livestock (values from Haines and ICPSR 2010). The G3 sample consists of adults age 20-35 (or 30-45) in the third sample year; the G2 sample consists of adults age 20-35 (of 30-45) in the second sample year, who have children ages 0-15 and are married to spouse in the same age bracket; the G1 sample consists of men in the first sample year who have children ages 0-15. Standard errors are clustered by G3 first name - decade groups.

Table 8: Intergenerational Elasticities Across Three Generations:
Percentile Rank Regressions with Paternal and Maternal Grandfathers by Region

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: 20-year intervals					
	Northeast		*Midwest*		*South*	
	G3 Male	G3 Female	G3 Male	G3 Female	G3 Male	G3 Female
G1 paternal	0.0548***	0.0272**	0.0311**	0.0195	0.0772***	0.0467***
	(0.012)	(0.013)	(0.014)	(0.012)	(0.016)	(0.016)
G1 maternal	0.0110	0.0067	0.0239	0.0260**	0.0304*	0.1082***
	(0.013)	(0.011)	(0.016)	(0.013)	(0.017)	(0.016)
G2	0.0749***	0.1028***	0.1939***	0.1821***	0.1817***	0.2233***
	(0.012)	(0.013)	(0.017)	(0.014)	(0.016)	(0.017)
Observations	86,627	49,380	103,555	64,944	68,105	45,941
p (G1 paternal = G1 maternal)	0.017	0.234	0.749	0.723	0.056	0.008
p (G1 pat [G3 male] = G1 pat [G3 female])	0.11		0.544		0.180	
p (G1 mat [G3 male] = G1 mat [G3 female])	0.795		0.918		0.001	
	Panel B: 30-year intervals					
	Northeast		*Midwest*		*South*	
	G3 Male	G3 Female	G3 Male	G3 Female	G3 Male	G3 Female
G1 paternal	0.0362***	0.0142	0.0174	0.0172	0.0484***	0.0332**
	(0.012)	(0.013)	(0.012)	(0.013)	(0.016)	(0.016)
G1 maternal	0.0201	-0.0185	-0.0209	0.0255*	0.0597***	0.1102***
	(0.013)	(0.012)	(0.013)	(0.014)	(0.016)	(0.017)
G2	0.0965***	0.1225***	0.2019***	0.2158***	0.1708***	0.2488***
	(0.014)	(0.012)	(0.014)	(0.015)	(0.016)	(0.017)
Observations	89,758	64,883	102,752	76,107	63,652	48,239
p (G1 paternal = G1 maternal)	0.372	0.039	0.037	0.684	0.639	0.002
p (G1 pat [G3 male] = G1 pat [G3 female])	0.219		0.992		0.506	
p (G1 mat [G3 male] = G1 mat [G3 female])	0.030		0.018		0.033	

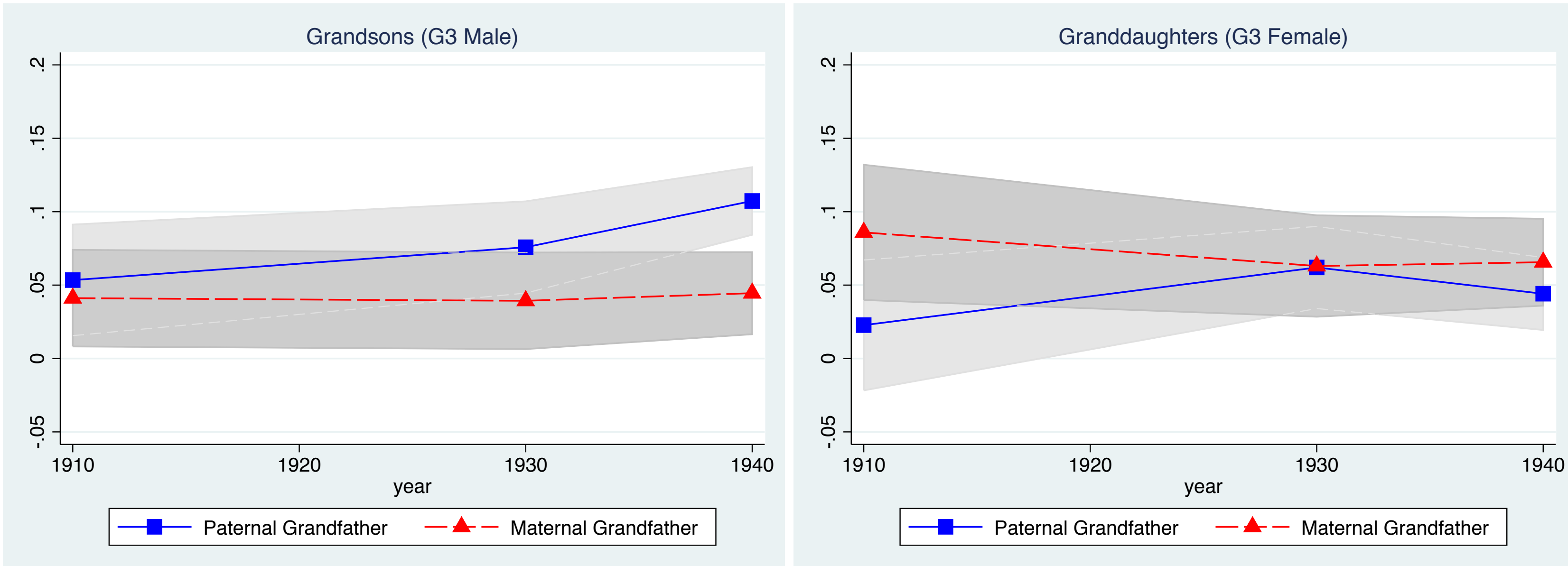
Notes. Contains results from OLS regressions of individual G3 log occupational score on the percentile rank of imputed scores of G2, paternal G1 and maternal G1; these are imputed as the average for each G3 individual's first name. For women, log occupational score is measured as the log occupational score of her husband. Panel A pools our three samples constructed at 20 year intervals, including decade controls; panel B pools our samples constructed at 30 year intervals. Columns (1) and (2) restrict the sample to individuals residing in the Northeast; columns (2) and (3) restrict the sample to individuals residing in the Midwest; columns (5) and (6) restrict the sample to individuals residing in the South. The G3 sample consists of adults age 20-35 (or 30-45) in the third sample year; the G2 sample consists of adults age 20-35 (of 30-45) in the second sample year, who have children ages 0-15 and are married to spouse in the same age bracket; the G1 sample consists of men in the first sample year who have children ages 0-15. Standard errors are clustered by G3 first name - decade groups.

Figure 1: Gender Differentials in G1-G3 elasticities, 20-year intervals



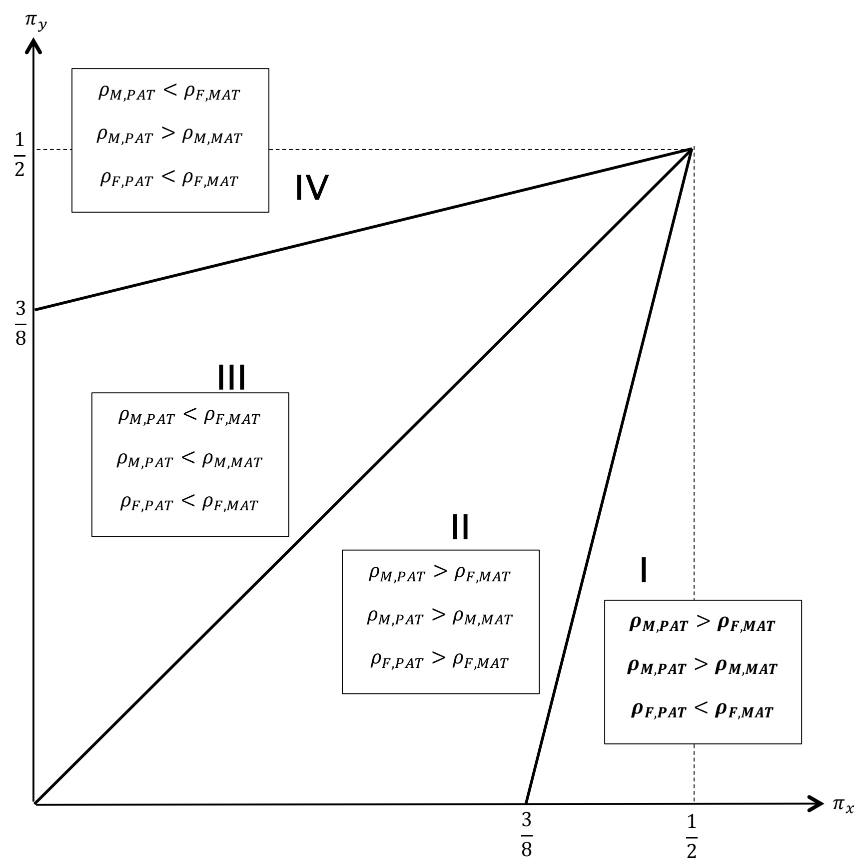
Notes. The left panel displays G1-G3 elasticities as reported in Panel A of Table 3, and the associated 90% confidence intervals. The right panel displays G1-G3 elasticities as reported in Panel B of Table 3, and the associated 90% confidence intervals. See notes to Table 3 for details.

Figure 2: Gender Differentials in G1-G3 elasticities, 30-year intervals



Notes. The left panel displays G1-G3 elasticities as reported in Panel A of Table 4, and the associated 90% confidence intervals. The right panel displays G1-G3 elasticities as reported in Panel B of Table 4, and the associated 90% confidence intervals. See notes to Table 4 for details.

Figure 3: Summary of Parameters of Matching Model and their Implications about the Importance of Paternal vs Maternal Grandfathers



Notes. The figure represents combinations of π_x and π_y (from the matching model described in section 4.2) that predict different relationships between the importance of paternal and maternal grandfathers in determining their grandchildren's outcomes. Here, we assume that $\pi_x^M = \pi_x^F \equiv \pi_x$, and $\pi_y^M = \pi_y^F = \pi_y$. In zones I & IV, the model predicts a gender asymmetry in importance of paternal or maternal grandfathers, with paternal grandfathers mattering more for grandsons and maternal grandfathers mattering more for granddaughters. In zone II, paternal grandfathers matter more for all grandchildren; in zone III, maternal grandfathers matter more for all grandchildren.

A Multi-trait Matching and Inheritance: Details

As shown in the text the two-generation transition probability matrices are Π_M and Π_F:

$$\Pi_M = \begin{bmatrix} (1-\pi_x)(1-\pi_y) & \pi_x(1-\pi_y) & (1-\pi_x)\pi_y & \pi_x\pi_y \\ \pi_x(1-\pi_y) & (1-\pi_x)(1-\pi_y) & \pi_x\pi_y & (1-\pi_x)\pi_y \\ (1-\pi_x)\pi_y & \pi_x\pi_y & (1-\pi_x)(1-\pi_y) & \pi_x(1-\pi_y) \\ \pi_x\pi_y & (1-\pi_x)\pi_y & \pi_x(1-\pi_y) & (1-\pi_x)(1-\pi_y) \end{bmatrix} \tag{3}$$

$$\Pi_F = \begin{bmatrix} (1-\pi_x)(1-\pi_y) & (1-\pi_x)\pi_y & \pi_x(1-\pi_y) & \pi_x\pi_y \\ (1-\pi_x)\pi_y & (1-\pi_x)(1-\pi_y) & \pi_x\pi_y & \pi_x(1-\pi_y) \\ \pi_x(1-\pi_y) & \pi_x\pi_y & (1-\pi_x)(1-\pi_y) & (1-\pi_x)\pi_y \\ \pi_x\pi_y & \pi_x(1-\pi_y) & (1-\pi_x)\pi_y & (1-\pi_x)(1-\pi_y) \end{bmatrix} \tag{4}$$

The three-generation transition matrices are obtained from the product of Π_M and Π_F:

$$\begin{aligned} \Omega_{M,PAT} &= \Pi_M\Pi_M \\ \Omega_{M,MAT} &= \Pi_F\Pi_M \\ \Omega_{F,PAT} &= \Pi_M\Pi_F \\ \Omega_{F,MAT} &= \Pi_F\Pi_F. \end{aligned}$$

One can then use these matrices to calculate the three-generation rank correlations, $\rho_{g,G}$:

$$\rho_{g,G} = \frac{\frac{1}{4}r'\Omega_{g,G}r - E(R)^2}{V(R)},$$

where $r = \begin{pmatrix} 1 & 2 & 3 & 4 \end{pmatrix}'$ and R is the random variable denoting an individual's rank, and has a discrete uniform distribution between 1 and 4.

It is then straightforward to calculate:

$$\begin{aligned} \rho_{M,PAT} &= 1 - \frac{4}{5}\pi_x(1-\pi_x) - \frac{16}{5}\pi_y(1-\pi_y) \\ \rho_{M,MAT} &= \rho_{F,PAT} = 1 - 2\pi_x - 2\pi_y + 4\pi_x\pi y \\ \rho_{F,MAT} &= 1 - \frac{4}{5}\pi_y(1-\pi_y) - \frac{16}{5}\pi_x(1-\pi_x) \end{aligned}$$

It is easy to see that because all the switching probabilities are bounded between 0 and 1/2, all the correlations are necessarily greater than zero.

We are interested in how π_x and π_y affect the relative magnitudes of $\rho_{M,PAT}$, $\rho_{F,MAT}$ and $\rho_{M,MAT}/\rho_{F,PAT}$. Specifically, we are interested in three contrasts: a) $(\rho_{M,PAT} - \rho_{F,MAT})$; b) $(\rho_{M,PAT} - \rho_{M,MAT})$; and c) $(\rho_{F,MAT} - \rho_{F,PAT})$. It is easy to show Proposition 1.

Proposition 1. *In the model described above, the following holds:*

1. $\rho_{M,PAT} - \rho_{F,MAT} > 0$ *if and only if* $\pi_y < \pi_x$.
2. $\rho_{M,PAT} - \rho_{M,MAT} > 0$ *if and only if: a)* $\pi_y < \pi_x$ *or b)* $\pi_y > \frac{3}{8} + \frac{1}{4}\pi_x$.
3. $\rho_{F,MAT} - \rho_{F,PAT} > 0$ *if and only if: a)* $\pi_x < \pi_y$ *or b)* $\pi_y < -\frac{3}{2} + 4\pi_x$.

Proof.

1. Write $\rho_{F,MAT} - \rho_{F,PAT}$ as $\frac{12}{5}\pi_x (1 - \pi_x) - \frac{12}{5}\pi_y (1 - \pi_y)$. Then, because π_x and π_y are smaller than 1/2, the inequality holds if and only if $\pi_x > \pi_y$.
2. Write $\rho_{M,PAT} - \rho_{M,MAT}$ as $\frac{4}{5}\pi_x^2 + \frac{6}{5}\pi_x + \frac{16}{5}\pi_y^2 - \frac{6}{5}\pi_y - \frac{20}{5}\pi_x\pi_y$. This expression can be rewritten as $\frac{2}{5}(\pi_x - \pi_y)(2\pi_x + 3 - 8\pi_y)$. Therefore the expression is positive if and only if the terms in parentheses are either both positive (i.e., $\pi_y < \pi_x$) or both negative ($\pi_y > \pi_x$ and $\pi_y > \frac{3}{8} + \frac{1}{4}\pi_x$).
3. Write $\rho_{F,MAT} - \rho_{F,PAT}$ as $\frac{4}{5}\pi_y^2 + \frac{6}{5}\pi_y + \frac{16}{5}\pi_x^2 - \frac{6}{5}\pi_x - \frac{20}{5}\pi_x\pi_y$. This expression can be rewritten as $\frac{2}{5}(\pi_x - \pi_y)(8\pi_x - 3 - 2\pi_y)$. Therefore the expression is positive if and only if the terms in parentheses are either both positive (i.e., $\pi_y < \pi_x$ and $\pi_y < -\frac{3}{2} + 4\pi_x$) or both negative ($\pi_y > \pi_x$).

□

Taken together, the three parts of Proposition 1 allow us to partition the (π_x, π_y) plane into the four zones shown in Figure 3.

Table A1: Intergenerational Income Elasticities across Three Generations
Different sample restrictions for G2 and G3, 1860-1880-1900 pseudo-panel

		G2: Baseline				G2: Married to Spouse ages 20-35			
	Variable	G3 Male	G3 Female	G3 Male	G3 Female	G3 Male	G3 Female	G3 Male	G3 Female
	G1	0.0690	0.0891	0.0106	0.0661	0.0798	0.0817	0.0072	0.0646
		(0.022)	(0.020)	(0.024)	(0.024)	(0.023)	(0.020)	(0.030)	(0.025)
	G2 Male	0.2185	0.2680			0.2007	0.2768		
		(0.030)	(0.025)			(0.031)	(0.025)		
G3: Baseline	G2 Female			0.2659	0.3057			0.2494	0.2710
				(0.028)	(0.024)			(0.032)	(0.023)
	Constant	2.8204	2.8492	2.8271	2.8333	2.8252	2.8487	2.8397	2.8585
		(0.016)	(0.017)	(0.017)	(0.017)	(0.016)	(0.017)	(0.017)	(0.018)
	Observations	77,878	44,292	78,634	44,930	77,718	44,171	77,761	44,168
	G1	0.0938	0.0891	0.0728	0.0661	0.1051	0.0817	0.0639	0.0646
		(0.022)	(0.020)	(0.022)	(0.024)	(0.023)	(0.020)	(0.027)	(0.025)
	G2 Male	0.2426	0.2680			0.2262	0.2768		
		(0.029)	(0.025)			(0.029)	(0.025)		
G3: Married	G2 Female			0.2759	0.3057			0.2655	0.2710
				(0.027)	(0.024)			(0.029)	(0.023)
	Constant	2.8530	2.8492	2.8481	2.8333	2.8565	2.8487	2.8608	2.8585
		(0.016)	(0.017)	(0.016)	(0.017)	(0.016)	(0.017)	(0.017)	(0.018)
	Observations	35,500	44,292	35,827	44,930	35,434	44,171	35,449	44,168
	G1	0.0959	0.0775	0.0700	0.0739	0.1065	0.0717	0.0581	0.0721
		(0.024)	(0.022)	(0.024)	(0.024)	(0.025)	(0.022)	(0.028)	(0.025)
	G2 Male	0.2302	0.2694			0.2139	0.2776		
		(0.030)	(0.025)			(0.031)	(0.025)		
G3: Married to spouse ages 20-35	G2 Female			0.2617	0.2926			0.2557	0.2608
				(0.028)	(0.024)			(0.030)	(0.024)
	Constant	2.8701	2.8533	2.8690	2.8348	2.8742	2.8522	2.8809	2.8585
		(0.016)	(0.017)	(0.017)	(0.017)	(0.016)	(0.017)	(0.017)	(0.017)
	Observations	29,295	29,550	29,563	29,957	29,238	29,469	29,252	29,453

Table A2: Distribution of Boys' and Girls' Names by Gender of Parent, 1880-1920
Samples used to construct pseudo-panels at 20 year intervals

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Year	Sex of G2 parent	# Children with parent age 20-35	Mean # G3 children per G3 name		Mean # G2 parent names per G3 child name		Mean Herfindahl index of concentration of G2 names per G3 child name		Mean % of parents linked back 20 years, parent names per G3 child name	
			Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
					G3 Male					
1880	Male	24,911	8.10	666.43	4.26	171.99	0.78	0.16	0.81	0.89
1880	Female	38,821	9.25	1007.97	4.78	252.52	0.79	0.14	0.84	0.9
1900	Male	32,553	7.46	407.59	4.18	114.48	0.79	0.17	0.85	0.9
1900	Female	52,774	8.69	678.82	5.01	200.12	0.79	0.13	0.87	0.91
1920	Male	47,031	8.34	498.48	4.81	155	0.8	0.15	0.84	0.91
1920	Female	72,668	9.24	747.76	5.49	240.92	0.8	0.12	0.87	0.91
					G3 Female					
1880	Male	24,016	7.00	381.26	4.13	111.8	0.8	0.17	0.85	0.9
1880	Female	37,505	8.06	588.55	4.61	155.74	0.79	0.15	0.83	0.9
1900	Male	31,204	6.76	248.38	4.17	87.43	0.8	0.17	0.87	0.9
1900	Female	51,064	7.98	412.7	4.92	141.69	0.8	0.14	0.87	0.91
1920	Male	44,684	7.51	371.35	4.83	135.97	0.8	0.15	0.88	0.91
1920	Female	69,728	8.36	569.55	5.25	194.44	0.8	0.13	0.86	0.91

Notes. Summary statistics for boys' and girls' names in 1880, 1900, and 1920, which are the "middle" years for our panels constructed at 20 year intervals. All statistics are computed at the G3 name level; the "weighted" statistics are weighted by the number of G3 individuals with each name. For example, the interpretation of row 1, column (4) is that the average male name in 1880 (among children 0-15 with fathers age 20-35) is given to 8.1 children; the interpretation of row 1, column (5) is that the average male child (among children 0-15 with fathers age 20-35) in 1880 shares a name with 666 children.

Table A3: Distribution of Boys' and Girls' Names by Gender of Parent, 1880-1910
Samples used to construct pseudo-panels at 30 year intervals

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Year	Sex of G2 parent	# Children with parent age 30-45	Mean # G3 children per G3 name		Mean # G2 parent names per G3 child name		Mean Herfindahl index of concentration of G2 names per G3 child name		Mean % of parents linked back 20 years, parent names per G3 child name	
			Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
					G3 Male					
1880	Male	44,942	9.87	1181.2	4.62	253.07	0.79	0.14	0.84	0.9
1880	Female	45,591	10.08	1191.9	4.93	278.05	0.79	0.13	0.83	0.87
1900	Male	65,415	9.51	900.53	4.61	199.53	0.8	0.14	0.84	0.9
1900	Female	64,627	9.52	908.87	5.07	240.19	0.79	0.13	0.85	0.9
1910	Male	72,398	10.36	777.85	5.11	183.87	0.78	0.13	0.84	0.91
1910	Female	73,033	10.62	796.31	5.78	223.25	0.78	0.11	0.86	0.91
					G3 Female					
1880	Male	43,510	8.74	732.53	4.6	174.56	0.79	0.14	0.86	0.9
1880	Female	44,024	8.85	734.1	4.74	180.96	0.79	0.15	0.82	0.87
1900	Male	64,227	8.92	554.92	4.71	153.82	0.8	0.14	0.85	0.89
1900	Female	63,678	8.96	560.35	5.08	169.55	0.8	0.14	0.84	0.9
1910	Male	70,347	9.87	558.81	5.23	160.24	0.79	0.13	0.87	0.91
1910	Female	70,794	10.11	563.58	5.71	176.86	0.79	0.12	0.86	0.91

Notes. Summary statistics for boys' and girls' names in 1880, 1900, and 1910, which are the "middle" years for our panels constructed at 30 year intervals. All statistics are computed at the G3 name level; the "weighted" statistics are weighted by the number of G3 individuals with each name. For example, the interpretation of row 1, column (4) is that the average male name in 1880 (among children 0-15 with fathers age 30-45) is given to 8.1 children; the interpretation of row 1, column (5) is that the average male child in 1880 (among children 0-15 with fathers age 30-45) shares a name with 666 children.