Intergenerational mobility in the very long run: Florence 1427-2011

Guglielmo Barone and Sauro Mocetti ♣

First and preliminary: May 2015

Abstract. The paper examines intergenerational mobility in the very long run, across generations that are six centuries apart. We exploit a unique dataset containing detailed information at the individual level for all people living in Florence in 1427. These individuals have been associated to their pseudo-descendants living in Florence in 2011, using the informational content of surnames. We find that earnings elasticity is about 0.04, much higher than the one predicted by traditional models of intergenerational mobility. We also find evidence of strong real wealth inheritance and occupation persistence in particular professions. Main findings are confirmed when we test the robustness of the pseudo-links and address the potential selectivity bias due to the heterogeneous survival rates across families.

Keywords: intergenerational mobility, informational content of surnames, Florence.

JEL classification: J62, N33, D31.

♣ Barone: Bank of Italy (Regional Economic Research Division, branch of Bologna) and RCEA, guglielmo.barone@bancaditalia.it; Mocetti: Bank of Italy (Regional Economic Research Division, branch of Bologna), sauro.mocetti@bancaditalia.it. We thank Giovanna Labartino, Peter Lindert and Giovanni Pica who kindly shared data on the distribution of surnames across provinces, data on incomes and earnings in 1427 and estimates on intergenerational income mobility in Florence using recent tax records data, respectively; we also thank the statistical office of Florence for providing us tax records by surnames. The views expressed in this paper have to be referred only to the authors and not to the Institution they are affiliated with.

"Prestige is an accident that affects human beings. It comes into being and decays inevitably. […] It reaches its end in a single family within four successive generations"
Ibn Khaldun

"Almost all the earnings advantages or disadvantages of ancestors are wiped out in three generations"
Gary Becker and Nigel Tomes [1]

1. Introduction

Most theoretical and empirical studies on intergenerational mobility focus on correlation in socioeconomic status between two successive generations[2] – parents and their children – and share a common view that the economic advantages and disadvantages of ancestors vanish in few generations. In this paper we question this view and empirically show the persistence of socioeconomic status across generations that are six centuries apart.

Linking people belonging to generations that are distant each other is difficult because of data limitations. In this paper we exploit a unique dataset (1427 Florentine Census) containing main socioeconomic variables at the individual level for people living in Florence in 1427. These individuals (our ancestors) have been associated to their pseudo-descendants living in Florence in 2011, using the informational content of surnames. From a more technical point of view we use a two-sample two-stage least squares (TS2SLS) approach: first, we use the sample of ancestors and regress log of earnings on a full set of surname dummies (and other socio-demographic controls such as age and gender); second, we observe current taxpayers (living in the city of Florence and present in the 2011 tax records) and regress log of their earnings on those of their ancestors, as predicted by their surname and the coefficients estimated in the first step. The

[1] Ibn Khaldun was the greatest Arab historian and he is considered among the founding fathers of modern sociology, historiography and economics; the citation has been drawn from his influential book *The Muqaddimah* (1377). Becker and Tomes provided, in their seminal contributions, the theoretical framework that represented the main building block of research on intergenerational mobility; the citation has been drawn from Becker and Tomes (1986).

[2] The earnings persistence between generations has been observed in all countries studied so far, although to varying degrees. See Black and Devereux (2011) and Corak (2013) for recent cross-country surveys. Chetty et al. (2014) moved the analysis at the local level, providing evidence across areas within the U.S.

same strategy has been repeated using log of real wealth or dummies for professions instead of log of earnings as dependent variables.[3]

We find that the elasticity of descendants' earnings with respect to ancestors' earnings is around 0.04. Stated differently, a one-standard deviation increase in the ancestors' earnings increases the descendants' earnings by 7% of its standard deviation. Intergenerational persistence in real wealth is even stronger. These results suggest that long run mobility is much lower than previously thought. To reconcile our results with those predicted by traditional models of intergenerational mobility, we provide further evidence showing that intergenerational mobility in the 15th Century was much lower than nowadays; moreover, we also find evidence of dynasties in certain (elite) professions, suggesting the existence of unobservable variables that may be transmitted across generations and that are not necessarily fully captured by earnings.

Our empirical findings have two main weaknesses. First, the strength of the pseudo-links may be questioned as we are working with generations that are six centuries apart. However, it is worth noting that pseudo-links are generated through both surnames and geographical localization because we consider people living in Florence. If the same data were available for all Italian cities, our strategy would entails prediction of ancestors' socioeconomic status using interaction between surnames and cities. This is arguably a more demanding and more precise approach to create links across generations than the one adopted in previous studies. Moreover, a rich set of robustness checks, including placebo regressions where we randomly reassign surnames to the descendants, is largely reassuring on the strength of the pseudo-links. Second, family survival rates – and, therefore, the likelihood of finding descendants of Florentine families in the 15th Century among current taxpayers – may vary to a large extent across families. If variation in the survival rate were correlated with current earnings and wealth, this would bias our estimates. To address this issue, we account for survival through an Heckman approach that confirms our main findings.

[3] Björklund and Jäntti (1997) were the first to apply the TS2SLS approach to intergenerational mobility estimation. Thenceforth the same strategy has been adopted for many country studies – see Corak (2006) for a review – and the variable traditionally used to predicted pseudo-fathers earnings were occupation, education and sector of activity. On the contrary, Aaronson and Mazumder (2008) imputed father's income using state and year of birth while Olivetti and Paserman (2015) exploited the information about socioeconomic status conveyed by first names. These variables, however, are partly endogenous since the choices about first names and/or the state of residence may be related to parental characteristics while surnames are more exogenous markers (and plausibly convey more information).

The main element of novelty of the paper is that we are able to provide evidence on intergenerational mobility in the very long run, linking ancestors and descendants that are six centuries apart (i.e. 20 generations of 30 years). Indeed, linking people through several generations has been done rarely.[4] Lindahl et al. (2015) use a Swedish data set that links individual earnings (and education) for three generations and find that persistence is much stronger across three generations than predicted from simple models for two generations. Chan and Boliver (2013) show a statistically significant association between grandparents' and grandchildren's class positions, even after parents' class position is taken into account. More closely to our paper, Collado et al. (2012) and Clark and Cummins (2014) exploited the distribution of surnames to estimate social mobility in the long run. Collado et al. (2012), using data from two Spanish regions, find that socioeconomic status at the end of the 20th Century still depends on the socioeconomic status of one's great-great grandparents; however, they also suggest that the correlation vanishes after five generations. Clark and Cummins (2014) use the distribution of rare surnames in England and find significant correlation between the wealth of families that are five generations apart.[5]

The empirical analysis also has other prominent strengths. First, we consider different socioeconomic outcomes including earnings, wealth and professions. Second, ancestors' socioeconomic status has been predicted using surnames at the city level, thus generating more precise links across generations with respect to other studies that use names or surnames at the national level. The huge heterogeneity of Italian surnames further strengthens the quality of the pseudo-links. Third, the Italian Renaissance offers a unique background to trace family dynasties and investigate the transmission of inequalities across centuries. Indeed, Florence in the 15th Century was already an advanced and complex society, characterized by a significant level of inequality (Milanovic et al., 2011) and by a rich variety of professions and occupational stratification.

The rest of the paper is structured as follows. Section 2 presents the empirical strategy. Section 3 provides some background information and describes the data and the variables. Section 4 shows the empirical results. Section 5 concludes the paper.

[4] See Solon (2014) for theoretical models accounting for intergenerational mobility across multiple generations.

[5] In the data used by Clark and Cummins (2014), the wealth is estimated at death. This may lead to a mis-measurement of wellbeing by ignoring transfers inter-vivos. Our data, on the contrary, have the advantage of being available when individual is adult. Moreover, we can control for the evolution of the outcome variable in the lifecycle by adding age among the controls.

2. Empirical strategy

The main requirement when analyzing socioeconomic mobility is an appropriate data set that spans over generations. Unfortunately, such a suitable dataset is not easily available and this is even more true if we consider generations that are centuries apart. To overcome the problem, we adopt an approach that combines information from two separate samples (TS2SLS).[6]

In the first sample we have information about ancestors' socioeconomic status (say earnings) and their surnames, and we run the following regression:

$$y_i^a = \delta S_i^a + \gamma X_i^a + \mu_i^a \tag{1}$$

where y_i is the log of earnings of people living in Florence in the 15th Century, X_i^a is a vector of controls such as age and gender and S_i^a is a set of dummies for each surname.

In the second sample we have information about pseudo-descendants, i.e. taxpayers currently living in Florence, and we run the following regression:

$$y_i^d = \beta\left(\hat{\delta} S_i^d\right) + \rho X_i^d + \mu_i^d \tag{2}$$

where y_i^d is the log of earnings of people currently living in Florence, X_i^d is as above a vector of controls for age and gender and $\hat{\delta} S_i^d$ is the log of ancestors' earnings imputed using surnames and the coefficients estimated in equation (1); the β is the TS2SLS estimate of intergenerational elasticity.

3. Data and descriptive analysis

3.1 Data sources and background information

Florence originated as a Roman city, and later, after a long period as a flourishing trading and banking medieval commune, it was the birthplace of the Italian Renaissance. According to the *Encyclopedia Britannica*, it was politically, economically, and culturally one of the most important cities in Europe and the

[6] The properties of the TS2SLS estimator are discussed in Inoue and Solon (2010).

world from the 14th to 16th centuries. During the 15ht Century, in particular, Florence alone had an income higher than that of the whole England, thanks to factories and large banks that had subsidiaries spread in much of Europe. The local currency, the florin, was the strongest currency and the most traded in Europe.[7] See Figure 1 for a map on Italian city-states in that period.

In 1427, in the midst of the fiscal crisis provoked by Florence's protracted wars with Milan, the Priors of the Republic decreed an entirely new tax survey that applied to citizens of Florence and to inhabitants of the Florentine districts (1427 Census, henceforth). The assessments were entrusted to a commission of ten officials, and their staff, and were largely complete within a few months, although revisions continued during 1428 and 1429. It has been acknowledged as one of the most comprehensive tax surveys to be conducted in the pre-modern Western Europe.[8]

The 1427 Census represents the first sample, containing information on socioeconomic status of the ancestors. Indeed, the dataset reports for each household, among other variables, the name and the surname of the head of the household, occupation at 2-digit level, assets (i.e. value of real property and of private and public investments), age and gender. The data were enriched with estimates on earnings, attributed to each person on the basis of occupations and the associated skill group.[9]

The 2011 tax records represent the second sample, containing information on socioeconomic status of the pseudo-descendants. From tax records, we draw information on incomes and main demographic characteristics (age and gender). Income items reported on personal tax returns include salaries and pensions, self-employment income, real estate income, and other smaller income items. In order to comply with the privacy protection rules, tax records has been collapsed at the surname level and only surnames with a frequency equal to 5 or above has been included. We define as earnings the total income net of real estate incomes while real wealth has been estimated from real estate incomes.[10]

---

[7] Moreover, the Medici – the most renowned rulers – gathered to court the best artists, writers, philosophers and humanists of the time such as, among the most famous, Leonardo da Vinci, Michelangelo, Botticelli, Dante, Machiavelli and Galileo.

[8] The documentary sources are fully described in Herlihy and Klapisch-Zuber (1985).

[9] Data on earnings were kindly provided by Peter Lindert (University of Davis). See the document gpih.ucdavis.edu/files/BLW/Tuscany_1427.doc for further information. The same data were also used in Milanovic et al. (2011) for an analysis on inequality in pre-industrial societies.

[10] Specifically, from the Bank of Italy Survey of Household Income and Wealth we select individuals living in the province of Florence, we regress the log of real assets on age, gender and incomes from building (actual and imputed rent) and we store the coefficients. Then we predict real wealth for

Pseudo-links between ancestors and their descendants are generated using geographical localization – only people living in Florence were included in both samples – and exploiting the informational content of surnames.

Italians have one of the largest collections of surnames of any ethnicity in the world. Surnames were inherited from one generation to the next, through the patriline, and most Italians began to assume hereditary surnames in the 15th Century. Some surnames derived from one's father names (patronymics), through the use of the Latin genitive (e.g. Mattei means son of Matteo) or formed by the preposition of "di"/"de" followed by the name (e.g. Di Matteo or De Matteo is the son of Matteo).[11] The origin or residence of the family gave rise to many surnames such as the habitat – Della Valle (i.e. "of the valley") – specific places – Romano (i.e. "Roman") – or nearby landmarks – Piazza (i.e. "square"). Ancestors' occupation (or object associated to the occupation) was also a widespread source of surnames, such as Medici ("medical doctors"), Martelli ("hammer") or Forni ("ovens"). Finally, also nicknames, typically referring to physical attributes, also gave rise to some family names, e.g. Basso ("short"). The huge variety of surnames was also amplified by the extraordinary linguistic diversity. Therefore there are surnames' ending that are region specific. For example, in Veneto many surnames end with "n" (e.g. Zonin), in Campania with "iello" (e.g. Borriello), in Sardinia with "u" or "s" (e.g. Piccinnu and Marras) and in Tuscany with "ai" or "ucci" (e.g. Bollai and Balducci). Unsurprisingly, the surnames present in our samples are highly Florence-specific: on average, the ratio between the surname share in Florence and the corresponding figure at the national level is nearly 6.

3.2 Descriptive evidence

[Here descriptive statistics]

[Table 1] [# surnames, corresponding to # taxpayers in Florence]

[Table 2]

[Here something on the economy, the occupational structure and inequality in Florence]

Banking, in the modern sense of the word, can be traced to the early Renaissance Italy, to the rich cities of Florence, Venice and Genoa. The Bardi,

---

individuals included in the tax records using age, gender, real estate incomes and the coefficients estimated and stored above.

[11] The large number of Italian surnames ending in "i" is also due to the medieval habit of identifying families by the name of the ancestors in the plural (which have an "i" suffix in Italian.

Medici and Peruzzi (Florentine) families were among the most famous and prominent bankers in Europe. The oldest pharmacy in Europe was presumably set up in Florence. Goldsmiths, in turn, were already organized into guilds and were among the wealthiest people in the city.

Anecdotal evidence highlights the strength of family dynasties across centuries. For example, the Frescobaldi was a prominent Florentine noble family that have been involved in the political, sociological, and economic history of Tuscany since the Middle Ages.

4. Results

4.1 Main results

In the first stage we regress log of ancestors' earnings on surnames using 1427 Census data. We find that surnames accounts for about [12%] of the total variation in log earnings and [22%] of the total variation in log wealth. These results support the hypothesis that surnames carry information about the father's socioeconomic status.[12] We use coefficients for surnames estimated in the first stage to predict ancestors' earnings and wealth for taxpayers included in the 2011 tax records, as shown in equation (2).

Table 3 presents our TS2SLS estimates of the intergenerational earnings elasticity.[13] We consider three different empirical specifications, with the first including only the predicted ancestors' earnings, the second and the third adding age and age and gender, respectively. The earnings elasticity is fairly stable across specification, with a magnitude around 0.04, and it is statistically significant at 5% level. Table 4 replicates the estimation outlined in equation (2) with respect to the wealth elasticity. Again, we detect a positive and highly significant parameter that ranges from 0.02 to 0.03. This first set of results documents a surprisingly high persistence of earnings and wealth across six centuries.

We can't directly compare the two elasticities because the size of the coefficients partly depends on the mean and the variance of the independent variable. To address this issue, in Table 5 we compare the magnitude of the two elasticities by estimating equation (2) on the same sample and computing the standardized beta coefficients. It turns out that that the size of the wealth elasticity

[12] Further evidence on this point will be discussed later.

[13] Standard errors have been bootstrapped with 1,000 replications in order to that take into account the fact that the key regressor is generated.

largely exceeds that of earnings elasticity either without controlling for sex and age (columns 1-2) or including those controls (columns 3-4). According to our preferred specification, a one-standard deviation increase in the ancestors' earnings increases the descendants' earnings by 6.7% of its standard deviation. The corresponding figure for real wealth is 9.7%. Therefore wealth persistence is higher than earnings' persistence and this is an expected result as real wealth can be transmitted across generations more easily and directly.

Table 6 provides a first set of robustness checks. First we address tax evasion. Our dependent variables are based on tax records that, as well known, may suffer from a severe underestimation due to tax evasion. In the first two columns we upwardly revise the variables from tax records with the correction factors suggested by Marino and Zizza (2011).[14] Results are unchanged and this may be explained by the fact that tax evasion might influence our results only if it is correlated with pseudo-ancestors' earnings (or wealth), which is clearly a very unlikely possibility. Second we address outliers as the distributions of earnings and wealth have long tails that might drive the results. In the last two columns we trim both the dependent variable and the key regressor at the 1% and the 99% level and we re-estimated equation (2): again, the estimates of positive and significant intergenerational elasticities are fully confirmed.

4.2 Robustness of pseudo-links

Our empirical strategy relies on the assumption that the probability that one taxpayer (randomly) taken from the 2011 tax record is a descendant of one taxpayer (randomly) selected from the 1427 Census is strictly higher if the two share the same surname. It is possible to support this assumption with a number of tests. At least two facts challenge our working hypothesis. First, people sharing the same surname may well not belong to the same family. Our test is then based on the idea that the more a surname is common the less sharing the surname is informative about the actual kinship. In the first two columns of Table 7 we re-estimate equation (2) by weighting observations with the inverse of the relative frequency in 1427, so giving more weight to rare surnames. Our results are largely

[14] Marino and Zizza (2011) compares incomes from tax records with those collected through the Survey of Household Income and Wealth. This approach is based on the hypothesis that as the survey questionnaire is multipurpose and replying is not compulsory, it is likely that respondents do not feel threatened or suspicious and would hence reply truthfully. On this basis, they provide for each income types a proxy of tax evasion (as measured by the difference between income from the survey and income from the fiscal source).

confirmed. The second threat to our assumption rests on the fact that the city of Florence is not a closed system. For instance, take a surname in 1427, say Bardi. Even if Bardi was a rare surname in 1427 it may well happen that in recent years a (non-descendant) immigrant named Bardi settled in Florence from outside. Our methodology treats erroneously the latter as a pseudo-descendant of the former. We minimize such a risk in the last two columns of Table 7 where we split our key parameters by interacting them with a dummy variable that equals 1 for more typical Florentine surname and 0 otherwise.[15] The results are reassuring: elasticities are larger for more Florence-specific surnames.

The two exercises discussed above *indirectly* test the robustness of pseudo-links. We complement them with a *direct* test that goes as follows. We randomly reassigned surnames to taxpayers in 2011 and re-estimate the TS2SLS intergenerational elasticities. If the positive correlations we detect are not related to the lineage (whose measurement might be affected by error) but emerge by chance, we should find that our estimates are not statistically different from those stemming from a random reshuffling of surnames. Figure 2 shows the distribution of estimated earnings elasticity for 1 million replications. The two dashed vertical lines are the 95th and the 99th percentiles while the red line indicates the position of the estimate based on real surnames. These results provide a clear graphical representation of the informational content of surnames and the goodness of the pseudo-links: the simulated p-value in this exercise is lower than 1%. Figure 3 shows the corresponding results for wealth where the outcome of the check is even more telling.

4.3 Selectivity bias due to families' survival rate

The analysis of intergenerational mobility in the long run points out to a demographic issue since families' survival rate depends on basic demographic processes that transform populations from one generation to the next. Moreover, reproduction, marriage, fertility, migration, and mortality may differ across people with different socioeconomic background.

As far as migration is concerned, some of the families recorded in the 1427 Census might have decided to migrate in the following centuries. Since they are not necessarily a random sample of the original population, this might bias our

[15] The measure for Florence-specific surnames is given by the ratio between the surname share in Florence and the corresponding figure at the national level. In the table we consider more typical Florentine surnames those with a value above the median.

estimates.[16] Analogously, dynasty's reproduction rate (i.e. fertility/mortality rate) may be correlated with income and/or wealth. Jones et al. (2010) show a strong and robust negative relationship between income and fertility, though they also argue that in the agrarian (pre-industrialization) economies the reverse could have been possible, as documented for example in Clark and Cummins (2009).[17] On the other side, it is reasonable to expect that the wealthiest families were those better equipped to survive across centuries (and therefore those that can be matched to the current tax records). How do we address these issues?

First, we provide descriptive evidence of the distribution of earnings and wealth in 1427, between the families who are still present in the tax records of 2011 and those who are not. The first row of Table 8 shows that the earnings of matched and unmatched surnames are not statistically significant (columns 1-3), nor that their distribution significantly differ. On the other hand this result does not apply to real wealth: matched surnames are healthier and differences are not limited to the mean but extend to the whole distribution. This is a quite expected results since inheritance of real wealth (including housing ownership) may be an obstacle to geographical moves. In order to take into account that surviving surnames might be a non-random sample and that selection bias might affect our results – a concern that seems more relevant for real wealth – we adopt a two stage Heckman correction. In the first stage we exploit further information recorded in the 1247 Census. Namely, we estimate a probit model with survival rate as a function of two dummies for migrant status (from other Italian cities and from abroad) and of the family size. The migrant status might influence the surname survival because migrants might display a higher propensity to a new mobility episode; family size, on the other hand, has a direct positive effect on the survival rate. The identifying assumption is that these three variables to not have a direct effect on earnings and wealth in 2011. Table 9 shows that being a migrant from abroad and larger family size influence the probability to be included in the sample and enter with the expected signs. We then compute the inverse Mills ratio

[16] Borjas (1987) provided a theoretical model to predict whether migrants are drawn mainly from the upper or lower tail of the skill (i.e. income) distribution. Migrants are the most able of the sending region (i.e. Florence) if the income distribution is more equal than in the destination region (and there is a strong positive correlation between the earnings they may expect at home country and those they may expect in the destination region; on the contrary, the migrants are the less able if the income distribution in Florence is more unequal.

[17] According to Clark and Cummins (2009), preindustrial families chose high gross fertility rates in order to maximize the chances of at least one surviving child. However, because of poorer health and nutrition, the poor were unable to match the rich in gross fertility. Thus, a large fraction of the poor died childless.

to correct the elasticity estimates: Table 10 indicates that selectivity is an issue only for wealth elasticity though our coefficients of interests are very close to the baseline results (if any, slightly upwardly revised).

4.4 Discussion of long term persistence

Intergenerational mobility scholars typically presume that correlations across generations decline geometrically (i.e. the correlation between grandparent and child is the square of the parent–child correlation, that the correlation between great-grandparent and child is the cube, etc.). If it was true, our estimates, which are referred to about 20 generations, would be not consistent with the prevailing estimates on earnings and wealth mobility. For example, assume the following deterministic law of motion for earnings: $y_t = \beta y_{t-1}$ where y is log of earnings and β the earnings elasticity between two successive generations. In Italy, according to the existing evidence, β is equal to 0.5.[18] Therefore, our earnings elasticity estimate, equal to 0.04, is consistent with a 4-5 generation span, much less than our case. Different plausible values for β cannot remove this inconsistency. In the following we discuss why we observe stronger persistence than that predicted by traditional models and we propose three possible explanations.[19]

First, traditional models rely critically on the assumption that the intergenerational transmission process of human capital has a memory of only one period. But this is an implausible assumption. Grandparents can directly transmit their cultural capital to their grandchildren through childrearing or other forms of interactions.[20] Moreover, the transmission mechanism might also work for more than two generations through a shared a persistent family culture. What matters to our aims is that in a simple two-generation transmission mechanism, income convergence will take longer. Recalling the example above, if we modify the law of motion as follows: $y_t = \beta y_{t-1} + \delta y_{t-2}$ and we assume $\beta = 0.5$ and $\delta = 0.25$ then we find that our earnings elasticity estimate is consistent with a 8-9 generation span. However, this explanation continues to be, by itself, not enough to fully explain the inconsistency.

[18] See Mocetti (2007).
[19] Our discussion is made with earnings elasticity but can be straightforwardly applied also to wealth elasticity.
[20] See Lindahl et al. (2015).

Second, intergenerational mobility in the past might have been lower than nowadays. Considering again our earnings elasticity across 20 generations, this is consistent with a β across successive generations equal, on average, to 0.85. In the pre-industrial era, the persistence in social standing across generations has been perceived as large, while some scholars tend to believe that industrialization and the rise of capitalism would bring a more fluid society.[21] Here we can provide some empirical support to this claim. Guell et al. (2014) developed a novel measure of intergenerational mobility that needs only cross-sectional data and is based on the informational content of surnames (ICS). They also show that ICS is a monotonic increasing function of the traditional intergenerational mobility measure. Following this methodology, our estimate for ICS in the 1427 data is 0.124.[22] This estimate turns out to be much larger than that found by Guell et al. (2015) for the province of Florence (0.021).[23] Hence, we argue that intergenerational mobility in the past was (much) lower than nowadays. Moreover, it is reasonable to assume that this immobile society was prevailing from 15th to 19th century.

Third, earnings elasticity might not fully capture the dynamics of the intergenerational persistence process and might not decline geometrically as commonly thought. Indeed, many social institutions contribute to status inheritance over multiple generations, especially at the bottom (e.g. due to ethnic or other social discrimination) and at the top (e.g. membership of exclusive clubs and/or elite professions) of hierarchies. In a world of absolutely perfect status inheritance – for example, a pure caste system – children, parents, grandparents, and earlier ancestors are identical in their social and economic positions. The perfect correlations between each generation make alternative types of intergenerational effects (e.g. children-parents, children-grandparents, etc.) indistinguishable. In a similar vein, Zylberberg (2014) underlines the existence of unobservable variables that are transmitted by parents and are not captured by

[21] See Erikson and Goldthorpe (1992) and Piketty (2000) for a discussion between the liberal and Marxist theory about the degree of intergenerational mobility in the industrial society.

[22] ICS is defined as $ICS \equiv R_D^2 - R_F^2$. The first R-squared (R_D^2) is obtained from the regression: $y_{i,s} = D + \mu_{i,s}$ where $y_{i,s}$ is log of incomes of individual i with surname s and D is an S-vector of surname-dummy variables with $D_S = 1$ if individual i has surname s and $D_S = 0$ otherwise. The second R-squared (R_F^2) is obtained from the regression $y_{i,s} = F + \mu_{i,s}$ where F is an S-vector of "fake" dummy variables that randomly assign surnames to individuals in a manner that maintains the marginal distribution of surnames. Therefore, the second regression mixes up the surnames so that they cannot be informative.

[23] From this exercise we can also draw qualitative results and the magnitude of the ICS should be compared with caution.

their son's earnings; sons of successful families may preserve the high prospects for their descendants even when their own earnings are not very high. In his theoretical framework, dynasties moves across careers rather than across income levels and a society can modelled as a Markov process in which the transition matrix is block-diagonal: only within-block mobility is allowed. This third explanation is consistent with an earning elasticity that do not decline geometrically and with a society characterized by dynasties in professions.

On the empirical side, we show suggestive evidence that some form of dynastic transmission of profession underlies our empirical case. Namely we show that the probability to be employed in a certain elite or niche occupation today is higher the more pseudo-ancestors were employed in the same occupation. We selected 4 professions: lawyers, bankers, medical doctors and pharmacists, and goldsmiths. We consider only these professions for several reasons. First, data availability since we need profession already existing in 1427 and for those we have access to publicly available data nowadays. Second, they should be elite or a niche profession, consistent with the fact that there should be unobservable variables (such as specific human capital or guilt privileges) that favored the career following and that are not fully captured by earnings. Third, available empirical evidence on career following focused exactly on the same professions: Lentz and Laband (1989) for doctors, Laband and Lentz (1992) for lawyers and Mocetti (2014) for pharmacists.

We proceeded as follows. The 1427-29 database on taxpayer include detailed information on the occupation (e.g. banker, innkeeper, etc.). For each profession we computed at the surname level a profession intensity measure as the percentage of people in that profession. For example, 2 out of 6 persons named Baroncelli worked as bankers, so that the banker-intensity of the surname "Baroncelli" is 0.333.

As far as contemporaneous data are concerned, we obtained the names of all people that has been working as managers in Tuscan banks in recent years from proprietary Bank of Italy supervisory reports (OR.SO. database). Data on lawyers working in the Florence area publicly available (http://www.consiglionazionaleforense.it) as well as those on medical doctors (http://www.ordine-medici-firenze.it/) and on pharmacists (http://www.ordinefarmacisti.fi.it/). Data on goldsmiths are taken from the Italian National Business Register database: we selected all members of governing bodies of firms in some applicable sectors.[24] We also have access to the list of the names of all Florentine taxpayers in 2004. Then, we were able to allocate them to the 4 professions thanks to the availability of name data on professions, with a residual category "other occupation". For example, there are 231 taxpayers named "Alessi". From the other sources we know that 1 "Alessi" works as lawyer, 1 as physician and 2 as goldsmith. 227 "Alessi" will be employed in "other occupation". Without loss of generality, we can impute the lawyer status to the first "Alessi", the physician status to the second and so on. After merging this dataset with the 1427-29 profession intensity we run the following regression:

$$y_{ips} = \beta intensity_{ps} + \varepsilon_{ips} \qquad (3)$$

where the dependent variable is a dummy that equals 1 if taxpayer *i* with surname *s* is employed in profession *p* and 0 otherwise; *intensity* is computed on the 1427 Census; a positive and significant estimate for β will signal the existence of long run intergenerational transmission of professions. Results are reported in Table 10. In the first column we detect a positive and significant estimate for β that survives after the inclusion of occupation fixed effects (column 2), while surname fixed effects do not significantly change the point estimate. In the last four columns we split the sample by professions. It turns out that the overall positive correlation can be separately found for lawyers, bankers and goldsmiths but not for medical doctors and pharmacists for which it is imprecisely measured.

5. Conclusions

[To be done]

[24] Nace Rev. 2 codes 461892, 464800, 477700, 321210.

References

Aaronson, D. and B. Mazumder (2008), Intergenerational economic mobility in the United States, 1940 to 2000, *Journal of Human Resources*, 43, pp. 139-172.

Björklund A. and M. Jäntti (1997), Intergenerational income mobility in Sweden compared to the United States, *American Economic Review*, 87, pp. 1009-1018.

Black, S.E. and P.J. Devereux (2011), Recent developments in intergenerational mobility, in O. Ashenfelter and D. Card (eds.), *Handbook in Labor Economics*, North Holland.

Borjas, G.J. (1987), Self-selection and the earnings of immigrants, *American Economic Review*, 77, pp. 531-553.

Chan, T.W. and V. Boliver (2013), The grandparents effect in social mobility: Evidence from British birth cohort studies, *American Sociological Review*, 78, pp. 662-678.

Chetty, R., N. Hendren, P. Kline and E. Saez (2014), Where is the land of opportunity? The geography of intergenerational mobility in the United States, *Quarterly Journal of Economics*, 129, pp. 1553-1623.

Clark, G. and N. Cummins (2009), Urbanization, mortality, and fertility in Malthusian England, *American Economic Review: Papers & Proceedings*, 99, pp. 242-247.

Clark, G. and N. Cummins (2014), Intergenerational wealth mobility in England, 1858-2012: surnames and social mobility, *Economic Journal*, 125, pp. 61-85.

Collado, M.D., I. Ortuño-Ortin and A. Romeu (2012), Long-run intergenerational social mobility and the distribution of surnames, working paper, Universidad de Alicante.

Corak, M. (2006), Do poor children become poor adults? Lessons for public policy from a cross country comparison of generational earnings mobility, *Research on Economic Inequality*. 13, pp. 143-188.

Corak, M. (2013), Income inequality, equality of opportunity, and intergenerational mobility, *Journal of Economic Perspectives*, 27, pp. 79-102.

Erikson, R. and J.H. Goldthorpe (1992), *The constant flux: A study of class mobility in industrial societies*, Oxford: Clarendon Press.

Güell, M., M. Pellizzari, G. Pica and J.V. Rodriguez Mora (2015), Correlating social mobility and economic outcomes, unpublished paper.

Güell, M., J.V. Rodriguez Mora and C.I. Telmer (2014), Intergenerational mobility and the informational content of surnames, *Review of Economic Studies*, forthcoming.

Herlihy, D. and C. Klapisch-Zuber (1985), *Tuscans and their families: A study of the Florentine Catasto of 1427*, New Haven: Yale University Press.

Inoue, A. and G. Solon (2010), Two-sample instrumental variables estimators, *Review of Economics and Statistics*, 92, pp. 557-561.

Jones, L.E., A. Schoonbroodt and M. Tertilt (2010), Fertility theories: Can they explain the negative fertility-income relationship? In J.B. Shoven (ed.), *Demography and the economy*, University of Chicago Press.

Laband, D.N. and B.F. Lentz (1992), Self-recruitment in the legal profession, *Journal of Labor Economics*, 10, pp. 182-201.

Lentz, B.F. and D.N. Laband (1989), Why so many children of doctors become doctors, *Journal of Human Resources*, 24, pp. 396-413.

Lindahl, M., M. Palme, S. Sandgren Massih and A. Sjögren (2015), Long-term intergenerational persistence of human capital: an empirical analysis of four generations, *Journal of Human Resources*, 50, pp. 1-33.

Long, J. and J. Ferrie (2013), Intergenerational occupational mobility in Great Britain and the United States since 1850, *American Economic Review*, 103, pp. 1109-1137.

Marino, M.R. and R. Zizza (2011), Personal income tax evasion in Italy: an estimate by taxpayer's type, in M. Pickhardt and A. Prinz (eds.), *Tax evasion and the shadow economy*, Cheltenham: Edward Elgar.

Milanovic, B., P.H. Lindert and J.G. Williamson (2011), Pre-industrial inequality, *Economic Journal*, 121, pp. 255-272.

Mocetti, S. (2007), Intergenerational earnings mobility in Italy, *B.E. Journal of Economic Analysis and Policy*, 7, (2).

Mocetti, S. (2014), Dynasties in professions: The role of rents, Temi di Discussione n. 995, Banca d'Italia.

Olivetti, C. and D.M. Paserman (2015), In the name of the son (and the daughter): intergenerational mobility in the United States, 1850-1940, *American Economic Review*, forthcoming.

Piketty, T. (2000), Theories of persistent inequality and intergenerational mobility, *Handbook of Income Distribution*, 1, pp. 429-476.

Solon, G. (2014), Theoretical models of inequality transmission across multiple generations, *Research in Social Stratification and Mobility,* 35, pp. 13-18.

Zylberberg, Y. (2014), Dynastic income mobility, unpublished paper.

Tables

Table 1. Descriptive statistics

Variable:	Mean	Standard deviation
Tax records 2010		
Earnings	24,234	4,929
Real wealth	59,225	26,148
Sex	0.52	0.05
Age	58.4	3.0
Catasto of 1427-29		
Earnings	36.2	44.8
Real wealth	291.2	705.0
Sex	0.15	0.36
Age	45.9	16.9

Source: tax records from Florence statistical office (fiscal year 2010) and 1427 Catasto of Florence, provided by Brown University; monetary variables are in Euro in the tax records and in Florentine florin in the 1427 Catasto archive.

Table 2. Persistence in families' socioeconomic status

Surname	Income (2010)	Occupation (1427)	% earnings (1427)	% wealth (1427)
5 poorest:				
DI SIMONE	7,528	Dealers in linen cloth, second-hand clothing	37%	49%
FERNANDEZ	9,369	Maestro (title applied to many professions but most commonly to medical doctors)	61%	44%
FUCCI	11,358	Sewers	43%	34%
LUCA	12,013	Workers in combing, carding and sorting wool	30%	39%
BARTOLO	12,287	Workers in combing, carding and sorting wool	34%	53%
5 richest:				
STEFANO	81,339	Brick layer, flag-stone worker, sculptor, workers in processing and finishing stone	38%	45%
FEDERIGHI	85,862	Messer (lawyer)	94%	93%
DI FILIPPO	95,881	Wool manufacturer or merchant; members of the wool guild	69%	65%
ANTINORI	99,254	Silk merchant or weaver; members of the silk guild	97%	97%
BONAMICI	149,547	Shoemaker; members of the shoemakers' guild	90%	89%

Table 3. Earnings mobility: baseline

Dependent variable:	Log of earnings	Log of earnings	Log of earnings
Log of ancestors' earnings	0.039**	0.043**	0.036**
	(0.017)	(0.020)	(0.018)
Female		-0.482***	-0.467***
		(0.116)	(0.122)
Log of age			-0.037
			(0.118)
Observations	806	806	806
R-squared	0.007	0.026	0.024

Bootstrap standard errors in parentheses (1,000 replications); *** p<0.01, ** p<0.05, * p<0.1

Table 4. Real wealth mobility: baseline

Dependent variable:	Log of wealth	Log of wealth	Log of wealth
Log of ancestors' wealth	0.027***	0.026***	0.019**
	(0.008)	(0.008)	(0.008)
Female		0.325	-0.320
		(0.286)	(0.299)
Log of age			2.293***
			(0.283)
Observations	679	679	679
R-squared	0.018	0.019	0.101

Bootstrap standard errors in parentheses (1,000 replications); *** p<0.01, ** p<0.05, * p<0.1

Table 5. Comparison between earnings and wealth mobility

Dependent variable:	Log of earnings	Log of wealth	Log of earnings	Log of wealth
Log of ancestors' earnings/wealth	0.042**	0.027***	0.038**	0.019**
Standardized beta coefficient	0.096	0.134	0.067	0.097
	(0.017)	(0.008)	(0.019)	(0.008)
Controls	NO	NO	YES	YES
Observations	679	679	679	679
R-squared	0.009	0.018	0.030	0.101

Bootstrap standard errors in parentheses (1,000 replications); controls are female and log of age; *** p<0.01, ** p<0.05, * p<0.1

Table 6. Income and real wealth mobility: robustness

Dependent variable:	Log of earnings	Log of wealth	Log of earnings	Log of wealth
Log of ancestors' earnings/wealth	0.048**	0.019**	0.055**	0.018**
	(0.023)	(0.008)	(0.024)	(0.008)
Controls	YES	YES	YES	YES
Model	correction for tax evasion	correction for tax evasion	trimming	trimming
Observations	806	679	790	667
R-squared	0.025	0.101	0.028	0.093

Bootstrap standard errors in parentheses (1,000 replications); controls are female and log of age; *** p<0.01, ** p<0.05, * p<0.1

Table 7. Mobility for rare and Florence-specific surnames

Dependent variable:	Log of earnings	Log of wealth	Log of earnings	Log of wealth
Log of ancestors' earnings/wealth	0.040*	0.018*		
	(0.022)	(0.009)		
× Less typical Florentine surnames			-0.002	0.017*
			(0.037)	(0.010)
× More typical Florentine surnames			0.051**	0.020*
			(0.022)	(0.012)
Controls	YES	YES	YES	YES
Specification	More weights to rare surnames in 1427		Differences by low- high- Florence-specific surnames	
Observations	806	679	806	679
R-squared	0.042	0.110	0.026	0.100

Bootstrap standard errors in parentheses (1,000 replications); controls are female and log of age; *** p<0.01, ** p<0.05, * p<0.1

Table 8. Earnings and wealth distribution by survival rate

	matched	unmatched	difference	Kolmogorov-Smirnov
Log of ancestors' earnings	36.2	35.9	0.305	0.009
			(0.721)	(0.679)
Log of ancestors' wealth	291.8	271.9	19.915**	0.022***
			(9.643)	(0.004)

Matched surnames are those present in both 1427 Census and 2010 tax records; unmatched surnames are those existing in 1427 Census but not in 2010 tax records; standard errors in parenthesis when testing differences in means; p-values in parenthesis for the Kolmogorov-Smirnov test for equality of distributions.

Table 9. First stage: survival rate

Dependent variable:	=1 if survive
=1 if migrants from other Italian cities in 1427-29	0.029
	(0.055)
=1 if migrants from abroad in 1427-29	-0.206***
	(0.037)
Size of the family in 1427-29	0.011***
	(0.003)

Bootstrap standard errors in parentheses (1,000 replications); *** p<0.01, ** p<0.05, * p<0.1

Table 10. Second stage: selection corrected estimates

Dependent variable:	Log of earnings	Log of earnings	Log of wealth	Log of wealth
Log of ancestors' earnings/wealth	0.040**	0.037**	0.030***	0.023***
	(0.017)	(0.019)	(0.008)	(0.008)
Controls	NO	YES	NO	YES
Inverse Mills' ratio	0.036	-0.002	0.504*	0.534*
	(0.128)	(0.122)	(0.295)	(0.285)
Observations	806	806	679	679
R-squared	0.007	0.024	0.022	0.106

Bootstrap standard errors in parentheses (1,000 replications); controls are female and log of age; *** p<0.01, ** p<0.05, * p<0.1

Table 10. Probability to belong to a given profession

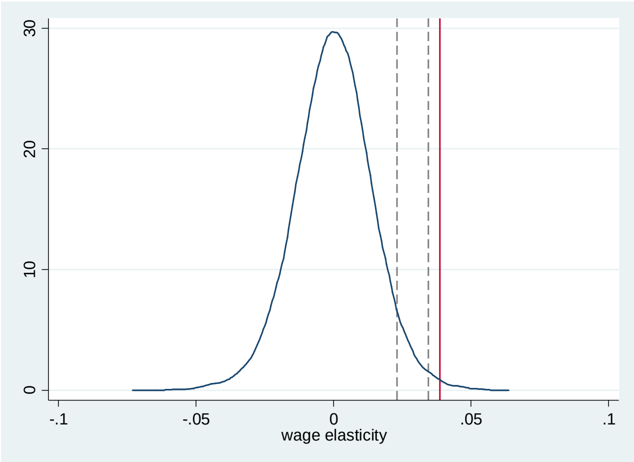
Dependent variable:							
Surn. inten.	0.007***	0.003**	0.003*	0.005**	0.001*	0.001	0.009**
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.004)
Profession	ALL	ALL	ALL	Lawyer	Banker	Medical doctor or pharmacist	Gold-smith
Prof. FE	NO	YES	YES	NO	NO	NO	NO
Surn. FE	NO	NO	YES	NO	NO	NO	NO
Obs.	532,772	532,772	532,772	133,193	133,193	133,193	133,193
R-squared	0.000	0.003	0.007	0.000	0.000	0.000	0.000

Robust standard errors clusterized at the surname level in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Figures

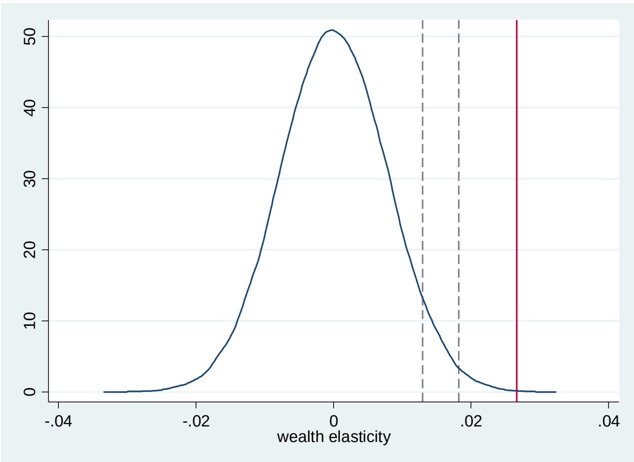
Figure 1. Italian city-states in the 1400

Figure 2. Earnings mobility with randomly assigned surnames



Distribution of estimated earnings elasticity randomly matching ancestors' and descendants' earnings; dashed lines represent 95° and 99° percentile, red line represents the earnings elasticity properly matching ancestors and descendants through surnames.

Figure 3. Wealth mobility with randomly assigned surnames



Distribution of estimated wealth elasticity randomly matching ancestors' and descendants' wealth; dashed lines represent 95° and 99° percentile, red line represents the wealth elasticity properly matching ancestors and descendants through surnames.