DYNASTIC INCOME MOBILITY*

Yanos Zylberberg

Abstract

Income mobility is usually approached by the father-son earnings elasticity. This measure misses an important point: the intergenerational income process is not an AR(1) and the full dynamics is not captured by its first moment. The intuition is that sons of successful families preserve the high prospects for their descendants even when their own earnings are not very high. Part of the father-son income mobility is not accompanied by long-run income mobility for dynasties. In this note, I discuss the theoretical relevance of this intuition. I also propose an empirical application based on PSID (contemporary US) and transitions of 4-digit occupations from fathers to sons: the implied earnings process is more persistent than an AR(1).

JEL: J24, J31.

Keywords: Income mobility, transitory shocks, dynasties.

Income mobility is often supposed to capture fairness in society. Consequently, a lot of attention has been devoted to the estimation of income mobility in developed economies over the past 30 years (see Becker and Tomes (1986), Solon (1992), Zimmerman (1992), Björklund and Jäntti (1997), Dearden et al. (1997), Haider and Solon (2006), Mazumder (2005), Lee and Solon (2009) or Mayer and Lopoo (2005)). Because of data limitation, the intergenerational income process is typically summarized by one, supposedly sufficient, statistics: the elasticity of sons' income with respect to their father's. However, when earnings of a dynasty do not follow an

*CREI (Universitat Pompeu Fabra) Ramon Trias Fargas, 25-27, 08005-Barcelona, (+34) 93 542 1145, yzylberberg@crei.cat. I thank the seminar participants at CREI, Universitat de Barcelona, FUNDP and in particular Gani Aldashev, Guilhem Cassan, Antonio Ciccone, Basile Grassi, Flavien Moreau, Jean-Philippe Plateau, Giacomo Ponzetto and Vincenzo Verardi for useful comments. Special thanks go to Regis Barnichon for very helpful remarks. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement nº 241114. Any remaining errors are my own.

AR(1), this moment is not sufficient to infer its full dynamics. In this paper, I argue that the process is very likely to be more persistent than an AR(1) in any society. The idea is that income does not capture the full perspectives of a dynasty and sons of a richer-than-average dynasty may have average-income jobs (think about university professors) even though the perspectives of their children remain above average.

The first objective of this note is to determine under what theoretical conditions the AR(1) process is a good approximation of the real intergenerational process. The second objective is to measure the difference between (i) the AR(1) implied by the use of the son-father income elasticity and (ii) the real dynamics of earnings, i.e. the correlations of current income with any future generation. To this purpose, the theoretical section describes an accounting framework where dynasties moves across *careers* rather than across income levels. Think about *career* as a general label that encompasses the relevant information on the future perspectives of a dynasty in the current period (occupation, education, income, location...). The career generalization relaxes the usual assumption that *current income* embeds all the information on future expected income flows. For instance, individuals in different careers but with the same earnings may have different long-run perspectives. I define in this context an equivalent of AR(1) processes, i.e. a set of transition matrices between careers that would generate exponentially decreasing intergenerational correlations. I then focus on societies in which convergence will be slower. When blocks of careers can be defined such that there is low porosity across blocks and high mobility within each block, I show that the intergenerational earnings process is more persistent than AR(1) processes.

In a second part, I provide a simple empirical illustration in the contemporary United States where careers are approached by the 4-digit occupation of the head of households. The results of this exercise point to a far more rigid society than suggested by the father-son income mobility only. After 5 generations, the correlation is higher than what an AR(1) would give after 3 generations.

The contribution of this note is (i) to describe an approach that allows to infer the higher moments of the intergenerational earnings process from the observation of father-son pairs only, (ii) to estimate these moments, using disregarded information (4-digit occupations) in PSID. The results indicate that occupations at a high level of disaggregation[1] embed more information on the dynasty perspectives than just income. We can think that it is a better predictor of cognitive skills, education or preferences/culture. Nonetheless, the majority of papers in economics using an

[1]The analysis that I perform here uses more than 500 categories.

occupational approach do so because of data limitation (see the historical studies of Ferrie (2005), Long and Ferrie (2005), Clark (2012) on income mobility before the 20th century) and construct occupational transition matrices with less than 10 categories.

The analysis developed here is new, but the intuition that the earning process could not be summarized in its first moment is evoked in two recent theoretical contributions (Solon (2013) and Stuhler (2012)). As regards the empirical estimation of the intergenerational income process, the economic literature has essentially tried to improve on the seminal paper of Becker and Tomes (1986), where the correlation between father and son's income is estimated to be low.[2] Solon (1992), Zimmerman (1992), Björklund and Jäntti (1997), and Dearden et al. (1997) have proposed methods to alleviate reporting biases and measurement errors. Some temporary shocks may move the reported contemporary income away from the permanent level of earnings. Haider and Solon (2006) and Mazumder (2005) have reconstituted life-course earnings, which naturally led to an upward revision of the intergenerational income elasticity. A second strand of the economic literature has decomposed the intergenerational correlations into cultural, genetic and bequest components, showing in particular the importance of human capital transmission.[3] Bowles and Gintis (2002) provides a good overview of these issues.

Section I. presents an illustrative example that helps to understand the argument. Section II. describes the accounting framework and the theoretical conditions under which the AR(1) approximation is reasonable. Section III. provides an empirical application to the United States and infers the higher moments of social mobility, i.e. the correlations of current income with any future generation. Finally, section IV. discusses some robustness checks and concludes.

I. A simple example

Consider in this section a simplistic society evenly divided between 4 time-unvarying careers: 1. bankers earning \$60000/year, 2. anthropologists: \$40000/year, 3. plumbers: \$40000/year, 4. truck drivers: \$20000/year.

Suppose that dynasties move across careers in the following way: the sons of bankers and anthropologists become bankers with probability 1/2 and anthropologists with probability 1/2. Symmetrically, the sons of plumbers and truck drivers become plumbers with probability 1/2 and truck drivers with probability 1/2. Σ,

[2] In this paper, there is an estimation of the grand-father/grand-son elasticity.

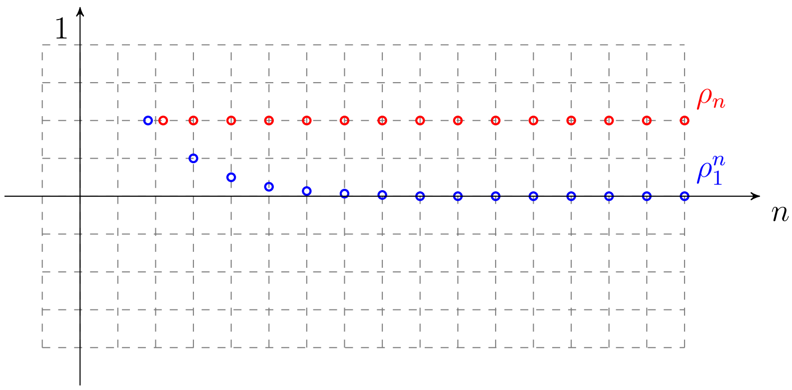
[3] See Lefgren et al. (2012) or Dahl and Lochner (2012) for recent contributions.

the transition matrix associated to this society, is shown below.

$$\Sigma = \left(\begin{array}{cc|cc} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ \hline 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{array}\right)$$

The transition matrix between generation n and generation 0 is Σ^n but, in this example, $\forall n, \Sigma^n = \Sigma$. For any initial career, the expectations of generation n are exactly the same as the prospects of sons: they both have a probability 1/2 to be in one of the two careers of the same block. The two blocks (top left: banker and anthropologist, bottom right: plumber and truck driver) remain completely hermetic.

In this society, the correlation between the earnings of sons and fathers can be easily computed: it is .5. However, since temporary shocks on income are not accompanied by long-term shocks on the dynasty perspectives, the father-son income elasticity of .5 does not translate into a rapid convergence to the mean on the long term. In fact, the correlation between earnings of the current generation and *any* future generation n is $\rho_n = .5$.[4] The social mobility is quite normal if you consider the father-son elasticity ρ_1, but there is never any kind of convergence. Figure I. displays the true intergenerational elasticities against the process implied by an AR(1).



This example gives the intuition of what will be developed in the theoretical part: societies exhibiting hermetic blocks within which there is some mobility do not absorb exponentially an initial shock.

[4]In this example, intergenerational earnings do not follow a Markov process in income because two individuals with the same current earnings (anthropologists and plumbers) have very different ancestors and descendants.

II. The accounting framework

A. Environment

Consider a society populated by a mass 1 of dynasties. Each dynasty has only 1 representant at each period giving birth to only 1 offspring. During each period, the representants of dynasties embrace one of the I available careers. Careers are labels that contain all the relevant information in the current period on the future perspectives of a dynasty: education, cognitive skills, preferences and the current income. A career perfectly characterizes the future expected income flows (including the current one). In practice, one would think of a career as "financial analyst with a PhD in economics from Princeton, working in Manhattan". Denote y_i the income over the life course earned by an individual in career i and assume that it is fixed across generations. Denote n_i the mass of workers in the career i.

Assumption 1. *Assume that the stochastic process composed of the random variables $C_0, \ldots, C_n, \ldots$ standing for the careers of links $0, \ldots, n, \ldots$ is a time-homogeneous Markov stochastic process, i.e.*

$$P\left(C_{n+1} = c_{n+1} | C_n = c_n, \ldots, C_0 = c_0\right) = P\left(C_{n+1} = c_{n+1} | C_n = c_n\right)$$

As already discussed, this assumption is weaker than the usual assumption stating that *current income* embeds all the information on future expected income flows. Any career Markov process can be summarized with a transition matrix Σ, which will be the central object of the analysis.

$$\Sigma = \begin{bmatrix} \sigma_{1,1} & \cdots & \sigma_{1,I} \\ \vdots & \ddots & \vdots \\ \sigma_{I,1} & \cdots & \sigma_{I,I} \end{bmatrix}$$

where

$$\sigma_{i,j} = P\left(C_{n+1} = j | C_n = i\right)$$

To finish with notations, denote

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_I \end{pmatrix}, N = \begin{pmatrix} n_1 \\ \vdots \\ n_I \end{pmatrix}$$

the vector of earnings over the I possible careers and the initial allocation of workers.

Earnings are normalized around the mean: $N^{'}Y = 0$. Suppose that the economy is stationary: the proportion of individuals in each career is constant over generations $\Sigma N = N$. Finally, assume without loss of generality[5] that $n_i = \frac{1}{I}, \forall i$.

The measures of intergenerational mobility in this framework are the correlations between the earnings of generation n and generation 0 conditional on the observation of the career of generation 0.

$$\rho_n = \frac{Y^{'}\Sigma^n Y}{Y^{'}Y}$$

In the series of measures $(\rho_n)_{n>0}$, ρ_1 is the classic father-son measure of intergenerational mobility. A notable difference implied by this framework is that the correlation between a grandfather's earnings and his grandson's would not be inferred through $\rho_1 \times \rho_1$, but instead computed as $\rho_2 \neq \rho_1^2$.

To understand why ρ_n is generally different than ρ_1^n, it is useful to compute the repeated 0-1 mobility ρ_1^n.

$$\rho_1^n = \frac{Y^{'}(\Sigma Z)^{n-1}\Sigma Y}{Y^{'}Y}$$

where $Z = YY^{'}/Y^{'}Y$ is the $I \times I$ cross-covariance matrix of vector Y. The transition matrix associated to ρ_1^n is $(\Sigma Z)^{n-1}\Sigma$ while the transition matrix associated to ρ_n is Σ^n. These two matrices are not necessarily equal and the rest of the analysis will consist in determining when they differ.

B. The AR(1) equivalent

In this discrete framework, the equivalent of AR(1) processes are processes for which the transition matrix Σ commute with Z. The real 0-n mobility and the repeated 0-1 mobility then coincide.

Lemma 1. *Under the assumption (H) that Σ and Z commute, $\rho_n = \rho_1^n$. In addition, (H) is equivalent to the assumption that y follows a discrete equivalent of an AR(1), i.e. there exists ρ such that*

$$\exists \rho, \forall i, \sum_{k=1}^{I} y_k \sigma_{k,i} = \sum_{k=1}^{I} y_i \sigma_{i,k} = \rho y_i \tag{C}$$

And such a ρ naturally coincides with the father-son elasticty captured with careers, i.e. $\rho = \rho_1$.

Proof. See the appendix. □

[5] As there is a finite number of careers, it is always possible to redefine subcareers out of a single career such that each sub-career has the same weight in the population.

Condition (C) can be thought as a characterization of AR(1) processes in this discrete case. This equivalence is not a formal equivalence because errors are not discrete equivalent of white noise, they can follow here a more general distribution. However, condition (C) implies:

- that the excess wage y_{n+1} of generation $n+1$ conditional on the excess wages y_n of generation n is equal to ρy_n. Overall, the whole information on the future of a dynasty is enclosed in current earnings.
- that the excess wage y_{n-1} of generation $n-1$ conditional on the observation of excess wages y_n of generation n is equal to ρy_n. This second condition means that ρ also represents the contribution of past generations shocks (enclosed in y_{n-1}) to a current excess wage relatively to the contemporary shock.

C. A block-diagonal approach

We have identified a sufficient condition (condition (C)) for the intergenerational elasticities to decrease exponentially and a set of processes for which the AR(1) approximation works. I cannot provide a simple necessary condition on the structure of the matrix but it is possible to exhibit structures for which the AR(1) approximation fails. To do so, the initial intuition with the banker/plumber example proves useful. In this example, persistence arises from the fact that there exist blocks in the transition matrix. These blocks differ by their long-term prospects (the average income in the subset of careers represented by the submatrices) but there is some within-block mobility. The rest of the theoretical analysis will generalize the example and focus on block-diagonal matrices.

Consider the set of processes for which there exists a permutation ϕ such that the transition matrix associated with the re-ordered careers $\left(C_{\phi(i)}\right)_i$ is block diagonal. In the appendix, I relax this assumption and extend the analysis to the wider set for which there are small transitions between blocks. The outline of the analysis will be as follows: for the set of block-diagonal matrices, the intergenerational elasticities can be derived as a function of within-block intergenerational elasticities and a between-block term (see proposition 1 below). It is then possible to deduce that the real process is always more persistent than an AR(1) process with the same father-son income elasticity (corollary to proposition 1).

Consider

$$\Sigma = \left(\begin{array}{c|c|c} \Sigma_1 & 0 \dots 0 & 0 \\ \hline 0 & \ddots & 0 \\ \hline 0 & 0 \dots 0 & \Sigma_K \end{array}\right)$$

the block-diagonal matrix where each block Σ_k is a square matrix of size I_k. Denote Y_k the subvector of Y associated with the careers in the block Σ_k. $\bar{y} = 0$ (by assumption, we normalize income over its national mean) and $\bar{y}_k$ are the average national excess income and the average excess income in careers Y_k. Define V_k the within-Σ_k income variance and 1_i the vector composed of 1 and of size i. For each block matrix, we can define the within-Σ_k 0-n intergenerational elasticity:

$$\rho_{n,k} = \frac{\left[Y_k - \bar{y}_k 1_{I_k}\right]' \Sigma_k^n \left[Y_k - \bar{y}_k 1_{I_k}\right]}{\left[Y_k - \bar{y}_k 1_{I_k}\right]' \left[Y_k - \bar{y}_k 1_{I_k}\right]}$$

Proposition 1. *The intergenerational elasticities $\rho_1, \dots, \rho_n, \dots$ can be written as follows:*

$$\forall n, \rho_n = \frac{\bar{\rho}_n + \mu}{1 + \mu}$$

where $\mu = \frac{\sum_{k=1}^K I_k (\bar{y}_k - \bar{y})^2}{\sum_{k=1}^K V_k}$ is the ratio of between-block and within-block variances and $\bar{\rho}_n = \frac{\sum_{k=1}^K \rho_{n,k} V_k}{\sum_{k=1}^K V_k}$ is the weighted average of 0-n intergenerational elasticities within each block.

Proof. See the appendix. □

The proposition states that ρ_n is not the simple average of within-block correlation $\bar{\rho}_n$. The expression includes a term μ that accounts for the differences in long-term perspectives between blocks. This term makes the intergenerational income transmission more persistent than an AR(1) process with the same father-son income elasticity ρ_1.

Corollary 1. *Under the assumption that processes within each block are AR(1) processes of parameter $r \geq 0$, $\bar{\rho}_n = r^n$, the intergenerational elasticities $\rho_1, \dots, \rho_n, \dots$ verify the following properties:*

$$\forall n \geq 1, \rho_n \geq \rho_1^n$$

$$\forall n > 2, (\rho_n = \rho_1^n) \Leftrightarrow (\mu = 0 \quad or \quad r = 1)$$

Finally,

$$\varepsilon > 0, r < 1 \Rightarrow lim_{n\to\infty}\rho_n = \frac{\mu}{1+\mu} > 0$$

Proof. See the appendix. □

The only cases in which the process is at least as persistent as an AR(1) are corner cases in which either there are no average differences between blocks (the long-term perspectives of all blocks are the same), or the society is completely rigid. Otherwise, the process will be more persistent and there will never be full mean-reversion. When you allow for small transitions between blocks,[6] the process is still more persistent than an AR(1) but ends up reverting to the mean. In the empirical application, I will first describe some elements that indicate a block-diagonal structure of transition matrices. Second, I will estimate the moments of the intergenerational income process implied by the transition matrix and show that it is not geometrically decreasing.

III. An empirical application

The first step of the empirical application is to define the empirical equivalent of a career. For simplicity, I will capture it by the main 4-digit occupation of the head of the household. Education would be a good candidate but the categories reported in surveys are too rough to capture elements that are missed by the observation of income.

I associate to each 4-digit occupation (more than 500 in-sample) precise transition probabilities as well as earnings and their weight in the population. The first part of this section describes the data collection and the construction of a career-transition matrix. I provide an attempt to identify blocks in the diagonal of the transition matrix. The second part presents the counterfactual income correlations between generations using the information contained in the transition matrix.

A. Data sources and construction

In line with other studies (see Lee and Solon (2009) and Mayer and Lopoo (2005) in particular), I use the PSID, a nationally representative study of 5,000 families (initially households) in the United States and the new households formed by the descendants of the initial head. As the study is a pretty long panel, it is possible to

[6]See the extension in the appendix.

compare sons and fathers at a working age and to smooth income over the different waves. Accordingly, I create two separate datasets that are ultimately merged.

First, I create a *son* dataset consisting in the households headed by children of the initial households (interviewed in 1968-1970). To this purpose, I merge the latest waves of PSID, i.e. 2003, 2005, 2007, 2009. Households are followed and interviewed every 2 years, giving not only a very large set of information on the composition of the household, its wealth, the earnings from the previous year but also the current 4-digit occupation of working members. Crucially, the 4-digit occupation of fathers and mothers is reported[7]. As regards the father occupation, I aggregate reports over the different waves and keep the most frequently reported 4-digit occupation for the father. I double check that their brothers and sisters, if any of them forms another household, report the same occupation for their parents. As regards the son's occupation, for each wave, the current (main) job is documented. I create a wave-specific transition matrix using each wave-specific occupation and the smoothed father occupation. I then average the 4 wave-specific transition matrices, which creates the transition matrix $\widehat{\Sigma}$ from the father's career to the potential states for the sons over the four waves. In parallel, I create the wealth, capital earnings, labor income for the son household and its members over the period 2003-2009.

To complement this data and perform some robustness checks, I also recreate a *father* dataset from the initial surveys of the father households in 1968, 1969, 1970. At this time, occupations are documented at the 1-digit level but income, household wealth are well reported. Accordingly, the *father* dataset allows me to recreate earnings and wealth but does not improve on the measure of father's occupation (it only allows to check the quality of sons' report).

I restrict the sample to working-age heads, i.e. between 25 and 60, and drop the households in which the labor income is smaller than half of the total income of the household (to get rid of rentiers essentially). In the end, the final dataset is composed of around 4000 father-son pairs.

In order to associate earnings and a population weight to each occupation, I will construct two sets of measures, (I) an in-sample set, with the weights and average earnings of respondents in the PSID, (II) an out-of-sample set with the weights and earnings computed nationally (in 2007) by the Bureau of Labor Statistics. Wages for each occupation[8] are given (average, 10-, 25-, 50-, 75-, 90-percentile) together with the education requirements and the average education of the holders of this job. Occupations are there also decomposed at the finest level (4-digit).

[7]The question specifies that it should be the main occupation of their father/mother; how respondents exactly interpret this term *main* is unclear.

[8]In the appendix, I detail the translation from the PSID code to the BLS code.

Before estimating the counterfactual measures of mobility, I provide a back-of-the-envelope procedure in order to visualize blocks in the transition matrix $\widehat{\Sigma}$. The idea is to transform the occupational transition matrix through a permutation of the labels of occupations. To to this, I rely on the sparse reverse Cuthill-McKee algorithm that reorders sparse matrices around the diagonal and exploit the observation that a block diagonal matrix is a matrix which remains close to the diagonal even when put to the square, cube... Consequently, I create $\widehat{\Sigma}^2, \ldots, \widehat{\Sigma}^n$ and let the sparse reordering algorithm extract the permutation of labels that ensure that the matrix $I + \widehat{\Sigma} + \ldots + \widehat{\Sigma}^n + \ldots = \left(I - \widehat{\Sigma}\right)^{-1}$ is close to the diagonal. I then relabel the occupations thanks to the permutation and display the resulting matrix in figure 3. The transition matrix does not exhibit very salient blocks, a dynasty with any occupation has a small but positive probability of accessing any other occupation. However, a block can be identified in the top left corner. This block is uniquely composed of high-skill occupations, among which there is high porosity. Despite similar and very high educational requirements (these occupations require on average 17 years of studies), there is some variability in the wage of these high-skill occupations: the average annual earnings are around $63,000, with a very high standard deviation of $25,000. Consequently, fathers in some well-paid occupations (surgeons) are very likely to have sons in average-salary occupations (teachers), without reneging on the long-term perspectives of the dynasty. This high-skill block remains quite different from the rest of occupations: the two following blocks are both composed of relatively low-skill occupations (wages around $35,000, less than 13 years of education). The rest of the matrix is composed of medium-skill occupations (wages around $39,000, above 15 years of education) and does not exhibit salient blocks. There may be a higher mobility in these medium-skill careers.

B. Counterfactual measures of mobility

Having created the matrices $\widehat{\Sigma}$, N and Y from the PSID subsample and the BLS statistics, it is possible to compute the series of measures $(\rho_n)_{n>0}$. As only the transition from fathers to sons is observed, those measures are counterfactual correlations. It corresponds to the creation of a virtual dynasty based on the father-son transition matrix and these statistics are the correlations between the earnings of those virtual dynasty members.

Before presenting the results of the counterfactual mobilities between generation n and generation 0, I present some results on real correlations, i.e. correlations between the declared income of fathers between 1968 and 1970 and the sons in the waves 2003, 2005, 2007 and 2009. Table 1 gives the results on different subsamples and with different controls (many controls can be added without changing the results, I only include here the age of the fathers in 1968 and the age of the sons in 2003 and state fixed-effects). Remark that those estimates are around .3 which is low compared to the literature. The reason is that only the labor income of the head is considered excluding the spouse labor earnings and capital gains.

Estimates of the same father-son correlation can also be produced from (i) the estimated occupational transition matrix $\widehat{\Sigma}$, (ii) the weights N and the earnings Y (in the tables here, earnings will be out-of-sample earnings, i.e. the BLS average wage) associated with each occupation. Table 2 provides these results with 4 different specifications, with PSID weights or the national weights, and a transition matrix computed with 3- or 4-digit occupations. The first two columns display the specifications that are the closest from the direct income correlations. The father-son income correlations generated by this indirect procedure are a bit lower than when directly computed: .25 and .23 against .3. An explanation is that an above-average father in his own occupation may produce an above-average son in his own occupation. Fathers and sons both working in New York may earn more than the national average of their occupations. Giving to fathers and sons the average national wage instead of their real earnings biases downward the elasticity by neglecting those within-occupation persistence. The addition of state fixed-effects in table 1 seems to bridge the gap between the two estimates. In conclusion, the father-son correlation implied by the transition between occupations is very close to the actual father-son income correlation.

Table 2 also presents the inferred elasticities at a higher order, i.e. the inferred generation n/generation 0 income correlation implied by the transition matrix $\widehat{\Sigma}$. Both coefficients and standard errors correspond to the outcomes of a regression of the average wage in the potential occupations of a virtual generation n with the average wage of the occupation of generation 0. Focusing on the first two columns (with the same weights as in PSID), the grandfather-grandson income elasticity ρ_2 is around .10, the next moment ρ_3 is between .035 and .06, the following one ρ_4 between .02 and .04. Those numbers need to be compared to the geometrically decreasing moments implied by $\rho_1 = .25$, $\rho_2 = (.25)^2 = .0625$, $\rho_3 = (.25)^3 = .0156$, $\rho_4 = (.25)^4 = .004$.

The analysis indicates higher correlations than what would predict an AR(1) process. A way to visualize the amplitude of the persistence of the income transmission process is to represent it graphically. Figure 1 shows the differences between the true (red) elasticity between the n-th generation and the 0 and the elasticity computed repeatedly (blue) computed with 4-digit occupation and the PSID weights. The computed correlations are further and further from what would be implied by an AR(1) process. In particular, while an AR(1) with $\rho \approx .25$ shows full mean reversion after 4 periods, it is not the case with our implied correlations. There seems to be a slowdown of convergence to the mean, in line with the theoretical results in the case of quasi-block-diagonal transition matrices.

IV. Discussion

A. Robustness checks

Some issues arise with the analysis presented above. I discuss them below (all the results that are mentioned can be provided upon request).

The first issue concerns the choice of 4-digit occupation as *careers*. An analysis based on education may also deliver the same insights. Two stumbling blocks prevent me from providing education-based estimates. There are no survey data giving a sufficiently precise label for the degree of a father and his son. In parallel, the average wage corresponding to such a label would be very hard to compute. In short, an analysis based on education would have few bins and noisy earnings associated to each bin.

The second issue relates to the choice of a wage for each occupation. First, the relative wage of an occupation may have evolved during the past 40 years. Second, there is some heterogeneity in the earnings of individuals within the same occupation. Regarding the first point, I cannot go back as far as in 1968, but attributing to the fathers the average wage of their occupation in 1990 rather than the average wage today does not change the results. The reason is that the time variations are small compared to the between-occupation heterogeneity in wages. As regards the second point, it is also possible to draw wages from the distribution of earnings in the population and attribute it to fathers and sons (independent draws). Given the number of observations and the high variation between-occupation, doing so provides similar results than when giving directly the average wages. It could be argued however that father's and son's idiosyncratic shocks are correlated. In other words, high-performing fathers in their own occupation may have high-performing sons in theirs. As long as the shocks are expected to be positively correlated, the present study gives a lower bound of the persistence of the earnings process. Any within-job persistence (through skills, preferences, location, work ethics...) would add to the estimated persistence.

The third issue is that most of the estimates in the literature include capital earnings in addition to labor earnings. Accordingly, the estimates produced here are consistently lower than the benchmark in the literature. The present paper does not have much to say about capital transmission. It could be noted however that a similar intuition as the one developed in this paper could be applied to capital earnings. If returns to capital are very volatile, the capital earnings would fluctuate a lot across generations even for a constant dynastic wealth and the intergenerational earnings process would be more persistent than an AR(1).

Finally, a fourth issue may be the sample selection. The intergenerational process is estimated on fathers and sons at a working age, excluding rentiers or individuals marrying wives with high earnings. Upon request, I can provide the results when including single daughters, rentiers, or retired individuals/students. The results are very close to the ones presented here. On the same note, the separate estimates using any single wave for the sons (2003, 2005, 2007 or 2009) gives the same patterns and very close numbers.

B. Concluding remarks

The present study points to the estimation of social mobility as being a more complicated exercise than the estimation of the correlation between two consecutive generations. I propose a simple methodology to approach the higher moments without observing them directly. The idea is to rely on some information that embeds more of the perspectives of a dynasty than income only. I use here occupations at a 4-digit level, but precise indicators of education, preferences and cognitive skills would be good candidates.

The results indicate that, in the contemporary US, the process is more persistent than an AR(1). Some short-term mobility does not affect the long-term perspectives of dynasties. There are no obvious reasons why this general insight would not hold in different countries. I would expect the effect to be even more salient in societies where many talented agents accept to work in average-salary jobs. In France, the most successful students in a cohort often end up with high-level administrative jobs or in academia.

Finally, the argument that the observation of income is not sufficient to infer social mobility is often advanced in social sciences. In sociology, the estimation of social mobility has consisted in the creation of categories (social classes) and the construction of a transition matrix between these classes. The usual criticisms are that categorization is somewhat arbitrary and the estimates cannot directly be converted into a tax policy. For those reasons, this approach has generally been disregarded by economists. I find here some support for approaches not centered on income: the observation of other factors helps us to refine the estimation of the intergenerational income process.

References

Becker, G. S. and Tomes, N.: 1986, Human capital and the rise and fall of families, *Journal of Labor Economics* **4**(3), pp. S1–S39.

Björklund, A. and Jäntti, M.: 1997, Intergenerational income mobility in sweden compared to the united states, *The American Economic Review* **87**(5), pp. 1009–1018.

Borjas, G. J.: 1992, Ethnic capital and intergenerational mobility, *The Quarterly Journal of Economics* **107**(1), 123–150.

Bowles, S. and Gintis, H.: 2002, The inheritance of inequality, *Journal of Economic Perspectives* **16**(3), 3–30.

Charles, K. K. and Hurst, E.: 2003, The correlation of wealth across generations, *Journal of Political Economy* **111**(6), 1155–1182.

Clark, G.: 2012, Surnames and the laws of social mobility, *Technical report*.

Cunha, F. and Heckman, J. J.: 2007, The evolution of inequality, heterogeneity and uncertainty in labor earnings in the u.s. economy, *Nber working papers*, National Bureau of Economic Research, Inc.

Dahl, G. B. and Lochner, L.: 2012, The impact of family income on child achievement: Evidence from the earned income tax credit, *American Economic Review* **102**(5), 1927–56.

Dearden, L., Machin, S. and Reed, H.: 1997, Intergenerational mobility in britain, *Economic Journal* **107**(440), 47–66.

Erikson, R. and Goldthorpe, J. H.: 2002, Intergenerational inequality: A sociological perspective, *The Journal of Economic Perspectives* **16**(3), pp. 31–44.

Erikson, R. and Goldthorpe, J. H.: 2009, Social class, family background, and intergenerational mobility: A comment on mcintosh and munk, *European Economic Review* **53**(1), 118 – 120.

Ferrie, J. P.: 2005, History lessons: The end of american exceptionalism? mobility in the united states since 1850, *The Journal of Economic Perspectives* **19**(3), 199–215.

Haider, S. and Solon, G.: 2006, Life-cycle variation in the association between current and lifetime earnings, *American Economic Review* **96**(4), 1308–1320.

Ichino, A., Karabarbounis, L. and Moretti, E.: 2011, The political economy of intergenerational income mobility, *Economic Inquiry* **49**(1), 47–69.

Lee, C.-I. and Solon, G.: 2009, Trends in intergenerational income mobility, *The Review of Economics and Statistics* **91**(4), 766–772.

Lefgren, L., Lindquist, M. J. and Sims, D.: 2012, Rich dad, smart dad: Decomposing the intergenerational transmission of income, *Journal of Political Economy* **120**(2), pp. 268–303.

Long, J. and Ferrie, J.: 2005, A tale of two labor markets: Intergenerational occupational mobility in britain and the u.s. since 1850, *NBER Working Papers 11253*, National Bureau of Economic Research, Inc.

Long, J. and Ferrie, J.: 2007, The path to convergence: Intergenerational occupational mobility in britain and the us in three eras*, *The Economic Journal* **117**(519), C61–C71.

Mayer, S. E. and Lopoo, L. M.: 2005, Has the intergenerational transmission of economic status changed?, *Journal of Human Resources* **40**(1).

Mazumder, B.: 2005, Fortunate sons: New estimates of intergenerational mobility in the united states using social security earnings data, *The Review of Economics and Statistics* **87**(2), 235–255.

McIntosh, J. and Munk, M. D.: 2009, Social class, family background, and intergenerational mobility, *European Economic Review* **53**(1), 107 – 117.

Nakao, K. and Treas, J.: 1989, Computing 1989 occupational prestige scores.

Piketty, T.: 2011, On the long-run evolution of inheritance: France 1820–2050, *The Quarterly Journal of Economics* **126**(3), 1071–1131.

Solon, G.: 1992, Intergenerational income mobility in the united states, *The American Economic Review* **82**(3), pp. 393–408.

Solon, G.: 2013, Theoretical models of inequality transmission across multiple generations, *Working Paper 18790*, National Bureau of Economic Research.

Stuhler, J.: 2012, Mobility across multiple generations: The iterated regression fallacy, *Technical report*, IZA Discussion Paper No. 7072.

Zimmerman, D. J.: 1992, Regression toward mediocrity in economic stature, *American Economic Review* **82**(3), 409–29.

A Tables and figures

Table 1: Intergenerational earnings correlations

SPECIFICATION			OLS			
ρ_1	0.31523	0.30558	0.32451	0.26523	0.25713	0.27305
	(0.01592)	(0.01589)	(0.01704)	(0.01746)	(0.01742)	(0.01864)
Sample	Male	Male	Male 25-60	Male	Male	Male 25-60
Controls	-	Age	Age	-	Age	Age
Fixed effects	-	-	-	State	State	State
Observations (pairs)	4,091	4,091	3,720	4,091	4,091	3,720

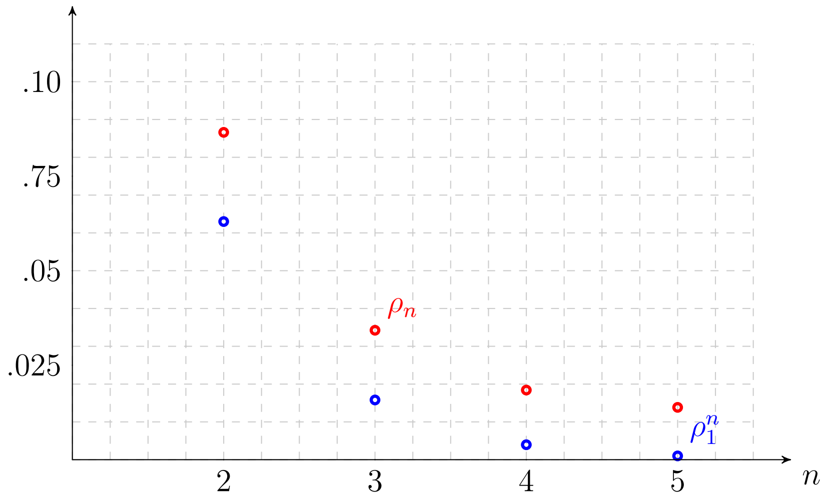
The standard errors for the OLS (between parentheses) are robust.

Table 2: Dynastic earnings correlations

SPECIFICATION	Indirect estimation through the occupational matrix			
ρ_1	0.25173	0.23779	0.14895	0.19648
	(0.01264)	(0.01288)	(0.01438)	(0.01358)
ρ_2	0.08660	0.10324	0.05681	0.066584
	(0.01543)	(0.01515)	(0.01594)	(0.01577)
ρ_3	0.03424	0.05858	0.02713	0.02875
	(0.01632)	(0.01591)	(0.01644)	(0.01641)
ρ_4	0.01843	0.03852	0.01572	0.01511
	(0.01659)	(0.01625)	(0.01663)	(0.01664)
ρ_5	0.01384	0.02707	0.01283	0.00951
	(0.01666)	(0.01644)	(0.01668)	(0.01674)
Sample			All 30-60	
Occupations	4-digit	3-digit	4-digit	3-digit
Weights	PSID	PSID	National	National
Observations (pairs)	3,500	3,500	3,500	3,500

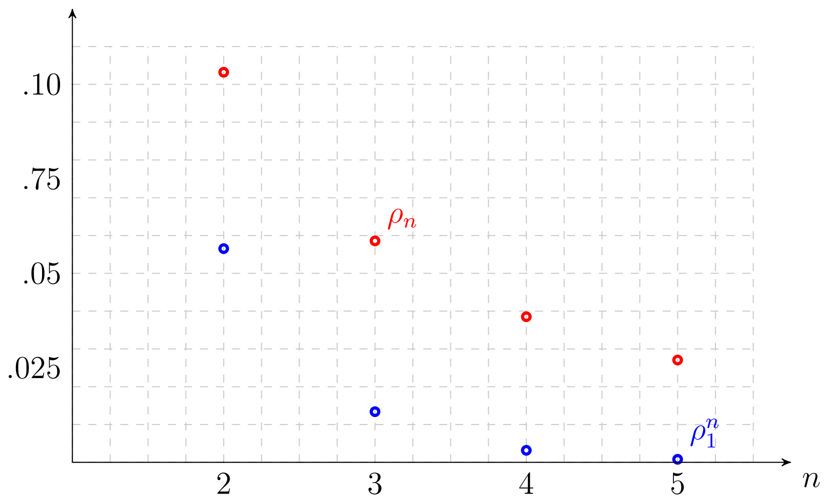
See the appendix for the computational details. The standard errors between parentheses are computed as if it was an OLS regression of the national average wage of the son's occupation against the national average wage of the father's occupation. Standard errors between brackets are computed with Monte-Carlo draws of the occupational matrix.

Figure 1: Dynastic earnings correlations - 4-digit computation, $\rho_1 = .25$



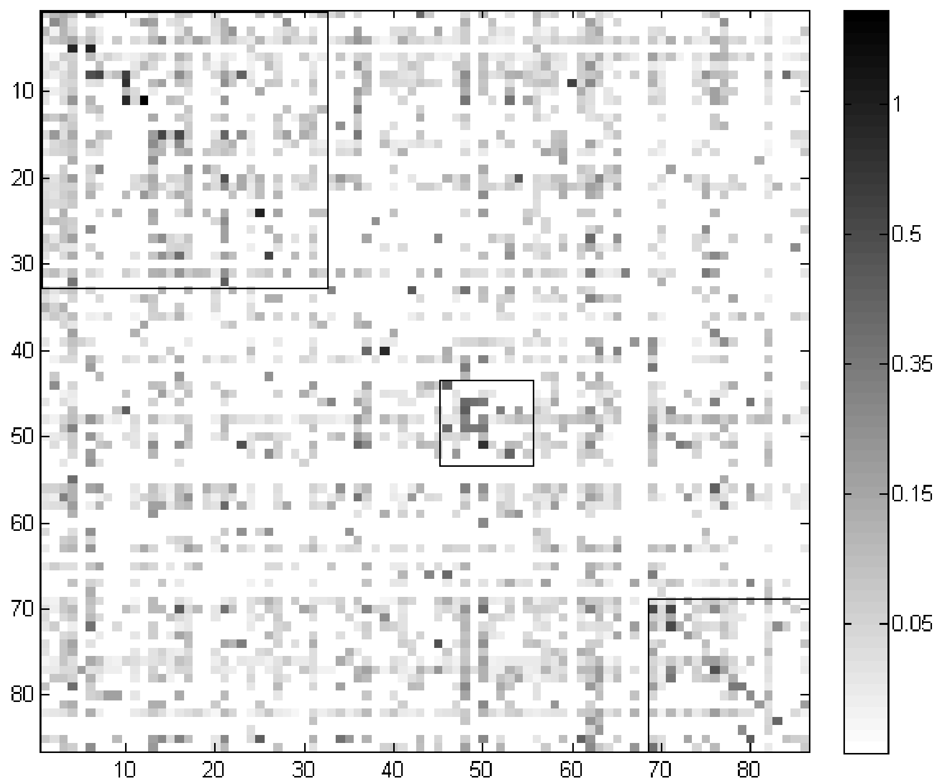
Red: $\rho_2, \rho_3, \ldots$ computed with the transition matrix, blue: $\rho_1^2, \rho_1^3, \ldots$ computed as the AR(1) benchmark.

Figure 2: Dynastic earnings correlations - 3-digit computation, $\rho_1 = .23$



Red: $\rho_2, \rho_3, \ldots$ computed with the transition matrix, blue: $\rho_1^2, \rho_1^3, \ldots$ computed as the AR(1) benchmark.

Figure 3: Re-ordered transition matrix for 3-digit occupations



B Appendix

A. Violation of the AR(1) hypothesis - continuous case

In line with the empirical strategy, I use a discrete approach in the theoretical framework but the argument can be made in a continuous framework (see Solon (2013) and Stuhler (2012) for a more developed analysis). Denote y_n the earnings of the link n of a certain dynasty, and X_n a vector of k underlying processes transmitting from one period to the other. Assume that X_n follow an AR(1) process:

$$X_{n+1} = X_n\beta + \varepsilon_n$$

where $\beta = diag(\beta_1, \ldots, \beta_k)$. Finally, suppose that

$$y_n = X_n\alpha + \nu_n$$

Under those assumptions, the correlation between y_n and y_{n+m} is:

$$\rho_m^y = \frac{\alpha^{'}\beta^m Var(X_n)\alpha}{\alpha'\beta Var(X_n)\alpha + Var(\nu_n)}$$

The empirical computation of ρ_1^y cannot allow us to distinguish what comes from X_n and from ν_n. Accordingly, computing ρ_1^y is not sufficient to infer the next correlations. The distance between an AR(1) process and the process deduced from this approach depends on the weight of father-son income mobility ν_n that is independent of the future perspectives of dynasties.

Remark that the previous model could also encompass the case in which the vector (y_n, X_n) follows an AR(1) process and the conclusion would be the same.

B. Proofs

Proof. Lemma 1. Assume that Z and Σ commute (H), $\rho_1^n = \frac{Y'\Sigma^n Z^{n-1}Y}{Y'Y}$.

Remark that $Z^{n-1}Y = Y$ brings $\rho_n = \rho_1^n$.

In addition, (H) implies that the elements of matrices $Z\Sigma$ and ΣZ are equal:

$$\forall i, j, y_i \sum_{k=1}^{I} y_k\sigma_{k,j} = y_j \sum_{k=1}^{I} y_k\sigma_{i,k}$$

This implies immediately that

$$\forall i,j, \frac{y_i}{\sum_{k=1}^{I} y_k \sigma_{i,k}} = \frac{y_j}{\sum_{k=1}^{I} y_k \sigma_{k,j}}$$

Denote ρ the value of these fractions (constant over i and j)

$$\forall i, \sum_{k=1}^{I} y_k \sigma_{k,i} = \sum_{k=1}^{I} y_i \sigma_{i,k} = \rho y_i$$

□

Proof. Proposition 1.

The proof of this proposition is straightforward. Consider n given, the correlation can be written as follows:

$$\rho_n = \frac{\sum_{k=1}^{K}(Y_k - \bar{y}1_{I_k})' \Sigma_k (Y_k - \bar{y}1_{I_k})}{\sum_{k=1}^{K}(Y_k - \bar{y}1_{I_k})'(Y_k - \bar{y}1_{I_k})}$$

Introducing $\bar{y}_k$ in the previous equation to make $\rho_{n,k}$ appear:

$$\rho_n = \frac{\sum_{k=1}^{K}(Y_k - \bar{y}_k 1_{I_k} + (\bar{y}_k - \bar{y})1_{I_k})' \Sigma_k (Y_k - \bar{y}_k 1_{I_k} + (\bar{y}_k - \bar{y})1_{I_k})}{\sum_{k=1}^{K}(Y_k - \bar{y}_k 1_{I_k} + (\bar{y}_k - \bar{y})1_{I_k})'(Y_k - \bar{y}_k 1_{I_k} + (\bar{y}_k - \bar{y})1_{I_k})}$$

Developing the expression, all the cross terms $\sum_{k=1}^{K}(Y_k - \bar{y}_k 1_{I_k})' \Sigma_k 1_{I_k}(\bar{y}_k - \bar{y}))$ disappear (because $Y_k' \Sigma_k 1_{I_k} = \bar{y}_k$), which brings:

$$\rho_n = \frac{\sum_{k=1}^{K}(Y_k - \bar{y}_k 1_{I_k})' \Sigma_k (Y_k - \bar{y}_k 1_{I_k}) + \sum_{k=1}^{K}(\bar{y}_k - \bar{y})^2 1_{I_k}' \Sigma_k 1_{I_k}}{\sum_{k=1}^{K}(Y_k - \bar{y}_k 1_{I_k})'(Y_k - \bar{y}_k 1_{I_k}) + \sum_{k=1}^{K}(\bar{y}_k - \bar{y})^2 1_{I_k}' 1_{I_k}}$$

Since $1_{I_k}' 1_{I_k} = 1_{I_k}' \Sigma_k 1_{I_k} = I_k$,

$$\rho_n = \frac{\sum_{k=1}^{K}(Y_k - \bar{y}_k 1_{I_k})' \Sigma_k (Y_k - \bar{y}_k 1_{I_k}) + \sum_{k=1}^{K}(\bar{y}_k - \bar{y})^2 I_k}{\sum_{k=1}^{K}(Y_k - \bar{y}_k 1_{I_k})'(Y_k - \bar{y}_k 1_{I_k}) + \sum_{k=1}^{K}(\bar{y}_k - \bar{y})^2 I_k}$$

Let us define

$$\bar{\rho}_n = \frac{\sum_{k=1}^{K} \rho_{n,k} V_k}{\sum_{k=1}^{K} V_k}$$

Then, dividing by $\sum_{k=1}^{K} V_k = \sum_{k=1}^{K}(Y_k - \bar{y}_k 1_{I_k})'(Y_k - \bar{y}_k 1_{I_k})$, and denoting $\mu = \frac{\sum_{k=1}^{K} I_k (\bar{y}_k - \bar{y})^2}{\sum_{k=1}^{K} V_k}$,

$$\rho_n = \frac{\bar{\rho}_n + \mu}{1 + \mu}$$

□

Proof. Corollary 1.

The assumption that each block is an AR(1) process of parameter r implies that $\bar{\rho}_n = r^n$. Once replaced in the formula found in the previous proposition,

$$\forall n, \rho_n = \frac{r^n + \mu}{1 + \mu}$$

Consider $n > 0$ and μ given. Define $f(.)$:

$$\forall r, f(r) = \rho_n - \rho_1^n = \frac{(r^n + \mu)(1 + \mu)^{n-1} - (r + \mu)^n}{(1 + \mu)^n}$$

f is a function defined on $[0, 1]$ verifying the following properties: $f(0) = \frac{\mu(1+\mu)^{n-1} - \mu^n}{(1+\mu)^n} \geq 0$, $f(1) = 0$ and f is decreasing in the segment$[0, 1]$. The last point comes from the fact that f is differentiable and the derivative verifies:

$$\forall r, f^{'}(r) = \frac{nr^{n-1}(1 + \mu)^{n-1} - n(r + \mu)^{n-1}}{(1 + \mu)^n} = n\frac{(r + r\mu)^{n-1} - (r + \mu)^{n-1}}{(1 + \mu)^n} \leq 0$$

Consequently, $\forall r, f(r) \geq 0$. If $\mu > 0$, then f is strictly decreasing, otherwise $f = 0$. This gives us the condition under which the quantities ρ_n and ρ_1^n coincides.

$$(f(r) = 0) \Leftrightarrow (\mu = 0 \quad or \quad r = 1)$$

Finally, the result at the limit is obvious. □

C. Extension of the block diagonal analysis

The block-diagonal analysis can be extended to quasi-block-diagonal matrices, with some porosity ε between blocks. Consider for this purpose a simpler case where blocks have the same size i (it does not change the qualitative results to relax this constraint), i.e.

$$\Sigma = (1 - \varepsilon)\Sigma_d + \frac{\varepsilon}{I - i}U_{ab}$$

where $0 < \varepsilon < 1$, Σ_d the previous block-diagonal matrix and U_{ad} is the associated anti-block diagonal matrix, i.e. the matrix with 1 everywhere except in the blocks of Σ_d:

$$\Sigma_d = \left(\begin{array}{c|c|c} \Sigma_1 & 0\dots0 & 0 \\ \hline 0 & \ddots & 0 \\ \hline 0 & 0\dots0 & \Sigma_K \end{array}\right), \quad U_{ad} = \left(\begin{array}{c|c|c} 0 & 1\dots1 & 1 \\ \hline 1 & \ddots & 1 \\ \hline 1 & 1\dots1 & 0 \end{array}\right)$$

In this setup, since $\Sigma_d U_{ad} = U_{ad}\Sigma_d = 0$,

$$\Sigma^n = (1-\varepsilon)^n \Sigma_d^n + \left(\frac{\varepsilon}{I-i}\right)^n U_{ab}^n$$

It is quite easy to derive that

$$U_{ab}^n = [I-i]^n \frac{1}{I} U + (-i)^n U_{ab}$$

where U is the square $I \times I$ matrix full of 1. From this, we can derive the expression of ρ_n:

$$\rho_n = (1-\varepsilon)^n \frac{\bar{\rho}_n + \mu}{1+\mu} + \varepsilon^n + \left(-\varepsilon\frac{i}{I-i}\right)^n A$$

where

$$A = \frac{\left[Y - N'Y\right]' U_{ab} \left[Y - N'Y\right]}{\left[Y - N'Y\right]' \left[Y - N'Y\right]}$$

The intergenerational elasticity is now composed of two exponentially decreasing terms: (i) the pure block diagonal process weighted by $(1-\varepsilon)^n$, (ii) the residual mobility between blocks $\varepsilon^n + \left(-\frac{i}{I-i}\right)^n A$. In such a framework, the society does revert to mediocrity:

$$\mu > 0, r < 1, \mu > 0 \Rightarrow lim_{n\to\infty}\rho_n = 0$$

The idea is that the small movements between blocks ensure a sufficient porosity for the expected future perspectives of dynasties to converge. The presence of the first term ensures however that the process will be more persistent than an AR(1) as long as ε is not too large.

D. Translation from one occupational code to the other

A problem that arises empirically is that occupational codes need to be translated from 1990-codes to 2000-codes. This translation is not completely obvious, it accounts for the grouping of some 1990 occupations into one banner and reciprocally, the creation of several jobs that were given the same code in 1990. I will create the $i_{90} \times i_{00}$ matrix $T_{90,00}$, giving the weight of each 1990 occupations in each 2000 occupation.

In the same vein, it can be useful to group some occupations at a higher level than the 4 digit occupational code, i.e. 3 digit or 2 digit. In this regard, I create the $i_{00}^4 \times i_{00}^3$ matrix $T_{4,3}$, giving the weight of each 4-digit 2000 occupations in each 3-digit occupation.