The Econometrics of Income and Social Mobility

Social mobility and Income Mobility

Michel Lubrano

January 2019

Contents

1 Introduction

We have studied income mobility is a very special way that could be inferred and understood easily. We just required the availability of household survey. Those are quite numerous. However, we could not follow the same household over time. We displayed the graph of the evolution of quantiles of the income distribution, but there is no guaranty that the same household remain in a given quantile. So, this was some kind of a crude estimation of income mobility.

In this chapter, we jump back in the past at a large distance, because the original paper we shall examine is Prais (1955). This paper is not about income mobility, but about social mobility in term of social classes or social occupations ranked in term of prestige. We need the definition of a scale defining the prestige of professions or social classes. There are several classifications, we can quote for instance that of Erikson et al. (1979).

The tools used are more complex than those built around the Lorenz curve and less familiar to economist: Markov transition matrices in relation to stochastic processes. So we shall define Markov processes and their properties. And of course, in order to summarize and compare transition matrices, we shall use mobility indices.

2 Social mobility and transition matrices

2.1 Informal examples

Social mobility concerns the passage between different social states over a given period of time. Social mobility is mathematically characterized by:

- There are k different possible social states.
- i is the starting state, j the destination state
- p_{ij} is the probability to move from state i to state j during the reference period.
- n_i is the number of individuals in state i

We are in fact introducing the Markov process of order one. It can be used to model:

- changes of social status between father and son: Prais (1955)
- change in occupational status
- changes in voting behaviour
- change in geographical regions
- Income mobility between different income classes over one or several years

2.2 Formal definition of Markov Chains

Let us consider k different states (job status, occupational status, etc...) such that an individual is assigned to only one state at a given time period. We let n_{ij}, $i, j = 1...k$ be the number of individuals initially in state i moving to state j in the next period. We can define a first matrix which describes the number of individuals in each case:

$$N = \begin{pmatrix} n_{11} & ... & n_{1j} & ... & n_{1k} \\ ... & ... & ... & ... & ... \\ ... & ... & n_{ij} & ... & ... \\ n_{k1} & ... & ... & ... & n_{kk} \end{pmatrix}$$

From this matrix, we define:

$$n_{i.} = \sum_{j=1}^{k} n_{ij},$$

the initial number of people in state i and of course $n = \sum_{i=1}^{k} n_{i.}$ the total number of individuals in the sample.

We now normalize this matrix so as to obtain a transition matrix P with independent lines which sum up to one,

$$P = [p_{ij}],$$

where p_{ij} represents the conditional probability for an individual to move from state i to state j in the next period. We have:

$$\sum_{j} p_{ij} = 1.$$

Let us call $\pi^{(0)}$ the row vector of probabilities of the k initial states at time 0. It indicates the probability of an individual to belong to a particular state i at the starting point of the Markov process. The probability to belong to any of the states at time 1 is described by the Markov process of order 1 as being given by $\pi^{(1)}$. The relation between $\pi^{(0)}$ and $\pi^{(1)}$ is given by:

$$\pi^{(1)} = \pi^{(0)} P, \tag{1}$$

by definition of the transition matrix.

The **Stationarity** assumption for a Markov process says that P is time invariant. Consequently, we can iterate (1) so as to obtain the distribution π at time t:

$$\pi^{(t)} = \pi^{(0)} P^{t}.$$

We suppose that the transition matrix has k distinct eigenvalues:

$$|\lambda_1| > |\lambda_2| > .... > |\lambda_m|.$$

Since P is a row stochastic matrix, its largest left eigenvalue is 1. Consequently, P^t is perfectly defined and converges to a finite matrix when t tends to infinity.

The stationary distribution $\pi* = (\pi_1^*, ..., \pi_k^*)^*$ is a row vector of non negative elements which sum up to 1 such that

$$\pi^* = \pi^* P.$$

It can be shown, see e.g. Guilbaud (1977) that the equilibrium vector is given by

$$\pi^* = [(P - I)(P - I)' + ii']^{-1} i,$$

where I is the identity matrix and i the column vector of ones.

This distribution vector is a normalized (meaning that the sum of its entries is 1) left eigenvector of the transition matrix associated with the eigenvalue 1. If the Markov chain is irreducible (it is possible to get to any state from any state) and aperiodic (an individual returns to state i can occur at irregular times), then there is a unique stationary distribution π^* and in this case P^t converges to a rank-one matrix in which each row is the stationary distribution π^*, that is

$$lim_{t\to\infty} P^t = \begin{pmatrix} \pi_1^* \cdots \pi_k^* \\ \cdots \\ \pi_1^* \cdots \pi_k^* \end{pmatrix} = i(\pi^*)'.$$

More details on the Markov model can be found for instance in Feller (1968, Chap 15).

3 Social mobility: Prais (1955)

This is the first paper in economics to study social mobility using a Markov model according to the prevailing literature, see e.g. Bartholomew (1973, 1982). From a random sample of 3500 males aged over 18 from the Social Survey in 1949, Prais (1955) explores the mobility between father and son in term of profession, using a Markov process of order one.

Let us first present the transition matrix obtained in this paper, using the notations developed above, where each row sum to one. *In Prais (1955), it is the columns that sum to one.* This is Table 1, where each row sums to unity.

Table 1: The Social Transition Matrix in England, 1949
The j^{th} element of row i^{th} gives the proportion of fathers in the i^{th} class whose sons are in the j^{th} social class. Transition from i^{th} class to j^{th} class

	1	2	3	4	5	6	7
1 High Administrative	**0.388**	0.146	0.202	0.062	0.140	0.047	0.015
2 Executive	0.107	**0.267**	0.227	0.120	0.206	0.053	0.020
3 Higher grade supervisory	0.035	0.101	0.188	0.191	**0.357**	0.067	0.061
4 Lower grade supervisory	0.021	0.039	0.112	0.212	**0.430**	0.124	0.062
5 Skilled manual	0.009	0.024	0.075	0.123	**0.473**	0.171	0.125
6 Semi skilled manual	0.000	0.013	0.041	0.088	**0.391**	0.312	0.155
7 Unskilled manual	0.000	0.008	0.036	0.083	**0.364**	0.235	0.274

We shall adopt the usual notation from now on. The rows are independent.

$$\pi'_t = \pi'_{t-1}P$$

π_t represents the distribution over classes at time t. After n generations the distribution will be:

$$\pi'_{t+n} = \pi'_t P^n,$$

provided P remains constant over time. The equilibrium distribution is given by

$$\pi'_* = \pi'_* P.$$

Once this distribution is reached, it will be kept for ever, provided of course that P remains constant. In fact sociologists are mainly interested in the changes inside P. Thus the equilibrium distribution is independent of the starting distribution. It is also independent of the time span. As if P relates the status of sons to that of fathers, the matrix relating that of grandsons to grandfathers is P^2. In Table 2, are computed the Actual and equilibrium distributions of social classes in England, 1949. Of course the columns are summing to 1.

3.1 Average time spent in a social class

There is perfect immobility if a family always stays in the same class. This would correspond to $P = I$. The more mobile is a family, the shorter the period it would stay in the same class.

Let us call n_i the number of families in class i at the beginning of the period. In the second generation, there will be $n_i p_{ii}$, then $n_i p_{ii}^2$ and so on.

Table 2: Actual and equilibrium distributions of social classes in England, 1949

Class	Fathers π_t	Sons π_{t+1}	Equilibrium π^*
High Administrative	0.037	0.029	0.023
Executive	0.043	0.046	0.042
Higher grade supervisory	0.098	0.094	0.088
Lower grade supervisory	0.148	0.131	0.127
Skilled manual	0.432	0.409	0.409
Semi skilled manual	0.131	0.170	0.182
Unskilled manual	0.111	0.121	0.129

The average time (measured in number of generations) is given by

$$1 + p_{ii} + p_{ii}^2 + \cdots = \frac{1}{1 - p_{ii}},$$

with standard deviation:

$$\frac{\sqrt{p_{ii}}}{1 - p_{ii}}.$$

In a perfectly mobile society, the probability of entering a social class should be independent of the origin. The matrix P representing perfect mobility has all the elements in each column equal (each row in the notations of Prais). But of course, columns can be different.

We consider a particular society. We compute the equilibrium distribution. The perfectly mobile society that can be compared to it is characterized by a transition matrix that has all its rows (columns in Prais' notations) equal to the equilibrium distribution π. In other words, from the introduction, this matrix is obtained as the limit of P^t when $t \to \infty$. The least mobile families

Table 3: Average number of generations spent in each social class

Class	England today	Mobile Society	Ratio	S.D.
High Administrative	1.63	1.02	1.59	1.02
Executive	1.36	1.04	1.30	0.71
Higher grade supervisory	1.23	1.10	1.12	0.54
Lower grade supervisory	1.27	1.15	1.11	0.58
Skilled manual	1.90	1.69	1.12	1.30
Semi skilled manual	1.45	1.22	1.19	0.81
Unskilled manual	1.38	1.15	1.20	0.72

are those belonging to the top executive (professional) class. The decimal part of the third column indicates the excess of immobility in percentage. Large self recruiting in the top group. The closer to perfect mobility are the Lower grade non-manual.

3.2 Prais index

A mobility index was later given the name of the Prais index, certainly by Shorrocks (1978) and is expressed as:

$$M_P = \frac{k - tr(P)}{k - 1}.$$

The reason is that Prais has shown that the mean exit time from class i (or the average length of stay in class i) is given by $1/(1 - p_{ii})$. Since M_P can be rewritten as $M_P = \sum_i (1 - p_{ii})/(k - 1)$ it is the reciprocal of the harmonic mean of the mean exit times, normalized by the factor $k/(k - 1)$. This is the explanation built in Shorrocks (1978).

3.3 Conclusion-summary

The states are easy to define because they are social classes based on job definition. This paper is a simple application of a Markov process properties. In particular, it is based on the definition of the equilibrium vector π in relation to the rows of limiting matrix P^t when $t \to \infty$. Immobility is defined in relation to the length of the stay in one's own social class. And perfect mobility is specific for one country. It corresponds to the case where all the lines of the transition matrix P are equal to the equilibrium vector π. So for two countries, there are two definitions of perfect mobility.

Some more properties are needed, in particular concerning the adequation between economic theory and Markov transition matrices. This is the object of the paper by Shorrocks (1978).

4 Axioms and properties

Intergenerational mobility. A transition matrix P with rows summing to one. p_{ij} start from state i and goes to state j in the next period. Using the usual notations. Study the properties of existing mobility indices and looks that axioms would be needed. The existing indices cannot satisfy all these axioms. The conflict comes from the definition of what is a perfectly mobile society when confronted to the requirement that a matrix P is more mobile than P' if some of its off diagonal elements are increased at the expense of the diagonal elements. We note $P \succ P'$.

4.1 Indices

Here are the main available mobility indices. This table comes from Formby et al. (2004) as well as the references quoted there. In this table, π^* represents the

Table 4: Main mobility indices

Measures	Sources
$M_1(P) = \dfrac{k - tr(P)}{k-1}$	Prais (1955), Shorrocks (1978)
$M_2(P) = 1 - \lvert\lambda_2\rvert$	Sommers and Conlisk (1979)
$M_3(P) = 1 - \lvert det(P)\rvert$	Shorrocks (1978)
$M_4(P) = k - \sum_i \pi_i^* p_{ii}$	Bartholomew (1973, 1982)
$M_5(P) = \dfrac{1}{k-1} \sum_i \pi_i^* \sum_j p_{ij} \lvert i-j\rvert$	Bartholomew (1973, 1982)

equilibrium vector of probabilities, the equilibrium distribution. P is the transition matrix and $M(\cdot)$ represents a function of P.

4.2 Axioms

Shorrocks (1978) introduces several axioms that could be imposed over mobility indices and the needed restrictions over transition matrices that could help to insure the compatibility of those axioms.

- **N** Normalization: $0 \leq M(P) \leq 1$. The index is at value between 0 and 1 like some inequality indices.
- **M** Monotonicity: $P \succ P' \Rightarrow M(P) > M(P')$
- **I** Immobility: $M(I) = 0$
- **SI** Strong Immobility: $M(I) = 0$ iff $P = I$
- **PM** Perfect mobility: $M(P) = 1$ if $P = ix'$ with $x'i = 1$.

The index of Bartholomew satisfies (**I**) but not (**SI**), (**N**), (**M**), or (**PM**). The reason is that the axioms (**N**), (**M**), and (**PM**) are incompatible.

The basic conflict is thus between (**PM**) and (**M**). This conflict can be removed reasonably by considering transition matrices that are maximal diagonal, which means:

$$p_{ii} > p_{ij}, \qquad \forall i, j$$

or quasi maximal diagonal:

$$\mu_i p_{ii} > \mu_j p_{ij}, \qquad \forall i, j \quad \text{and } \mu_i, \mu_j > 0.$$

With this last restriction, the Prais index satisfies Immobility (**I**), Immobility (**SI**), and Monotonicity (**M**).

We note that this property of maximal diagonal is not verified by the example provided in Prais (1955). It depends very much on the nature of P and the way it is built. The scale used by Prais (1955) is certainly not regular, which means that the distance between two initial positions is not constant. The grid is very fine for skilled and unskilled categories, but the distance between the first two categories is certainly large. When defining income mobility transition matrices, we shall escape this drawback.

4.3 Time consistency

Transitions matrices may be defined for unequal periods due to data restrictions. This case is not very often considered when intergenerational mobility is studied, because a generation is supposed to last something like 30 years. But for economic examples this can be important, because mobility is measured between years. That makes a difference to have most surveys sampled at one year difference and suddenly a survey coming after two or three years. The Chinese case with the CHNS is a good example of this problem. When comparing those matrices, we may have a tendency to say that matrices concerned with a longer period are more mobile. Let us consider an example which leads to well marked contradictions. We have two transition matrices for say two different countries, P and Q.

$$P^1 = \begin{pmatrix} 0.9 & 0.1 & 0.0 \\ 0.3 & 0.4 & 0.3 \\ 0.3 & 0.3 & 0.4 \end{pmatrix} \qquad Q^2 = \begin{pmatrix} 0.44 & 0.28 & 0.28 \\ 0.28 & 0.44 & 0.28 \\ 0.28 & 0.28 & 0.44 \end{pmatrix}$$

The matrix P^1 refers to a one year interval and while the matrix Q^2 refers to a two year interval. A comparison using the Prais index reveals that $M(P^1) = .65 < .84 = M(Q^2)$. So Q seems more mobile than P. However, to make a valid comparison, we could compute for P the transition matrix equivalent for a two year interval, P^2:

$$P^2 = \begin{pmatrix} 0.84 & 0.13 & 0.03 \\ 0.48 & 0.28 & 0.24 \\ 0.48 & 0.21 & 0.25 \end{pmatrix}.$$

We can now compute the Prais index for matrices corresponding to the same time interval. And $M(P^2) = .815 < M(Q^2)$ suggests that the system generating Q^2 is the more mobile, comforting our first conclusion.

Unfortunately, that is not the end of the matter. It could equally well have been argued that Q^2 was a two year transition matrix for a Markov chain process and that we should look for the associated one year matrix Q^1 to compare it with P^1. Such a matrix is given by

$$Q^1 = \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ 0.2 & 0.6 & 0.2 \\ 0.2 & 0.2 & 0.6 \end{pmatrix}$$

and $M(Q^1) = 0.60 < M(P^1) = 0.65$, reversing the rankings previously assigned to the systems generating the matrices under examination.

The problem does not arise if the index satisfies the one of following axioms:

PC Period Consistency: $M(P) \geq M(Q) \Rightarrow M(P^n) \geq M(Q^n)$ for all integers $n \geq 1$.

This is certainly a desirable property, but it is too restrictive as it violates the monotonicity condition (**M**). We can replace (**PC**) by a stronger condition for the desired index

PI Period Invariance: $M(P;T) = M(Pk;kT)$, $k \geq 1$.

The advantage of this formulation is that the index now compensates for the length of the time interval. If the process is a Markov chain, the mobility value obtained for any structure will be independent of the particular observation period prescribed by the data.

There is at least one index satisfying (**PI**) which is

$$M_D(P;T) = 1 - |detP|^{\alpha/T}, \quad \alpha > 0.$$

Measures derived from the determinant have not been well received, on the ground that they give the completely mobile value when any two rows (or columns) of the concerned matrix are identical. This might appear however as a minor drawback, compared to the period invariance property.

5 Income mobility

5.1 Social mobility versus income mobility

Markov processes model the transition between mutually exclusive classes or states. In a group of applications, mainly those coming from the sociological literature, those classes are easily identified because they correspond to a partition of the social space. We have for instance social classes, social prestige, voting behaviour or more simply economics job status such as working, unemployed, not working for instance. In fact those social statuses are directly linked to dichotomous variables. For studying income mobility, the problem is totaly different because income is a continuous variable that has to be discretized. And there are dozen of ways of discretizing a continuous variable.

In order to detail the various aspects of Markov processes used to model social mobility, it is easier to start from the case where the classes are directly linked to a discrete variable.

5.2 How to build transition matrices

This is detailed in Formby et al. (2004), an econometric paper. We consider a joint distribution between two income variables x and y measured for the same sample of individuals at two different periods of time. Let us note $K(x, y)$ their continuous c.d.f. Clearly, the function $K(x, y)$ completely captures the movement between x and y.

In the mobility measurement literature, the movement between x and y is described by a transition matrix, which is a transformation from a continuous c.d.f. of an income regime. To form such a transition matrix from $K(x, y)$, one first needs to determine the number of and boundaries between income classes. Suppose, there are m classes in each income distribution and the boundaries of these classes are, respectively for the two periods,

$$0 < \zeta_1 < \zeta_2 < \cdots < \zeta_{m-1} < \infty,$$

and

$$0 < \xi_1 < \xi_2 < \cdots < \xi_{m-1} < \infty.$$

The resulting transition matrix is denoted $P = \{p_{ij}\}$, and each of its elements is a conditional probability that an individual moves to class j of income y given that she was initially in class i of income x, i.e.,

$$p_{ij} = \frac{\Pr(\zeta_{i-1} \leq x < \zeta_i \text{ and } \xi_{j-1} \leq y < \xi_j)}{\Pr(\zeta_{i-1} \leq x < \zeta_i)},$$

where $\zeta_0 = \xi_0 = 0$ and $\zeta_m = \xi_m = \infty$. The probability that an individual falls into income class i of x is denoted π_i,i.e., $\pi_i = \Pr(\zeta_{i-1} \leq x < \zeta_i)$.

5.3 Definition of income transition matrices

Of course now the crucial point is how to define the boundaries. There are three ways of doing this. Comments come from Formby et al. (2004). Related concepts in term of welfare will be detailed in a next chapter.

5.3.1 Absolute transition matrices

The first approach views mobility as an absolute concept and exogenously sets boundaries between income classes. The resulting transition matrix is referred to as a ***size transition matrix****. Using this approach the boundaries of income classes ζ_i and ξ_i are predetermined and do not depend on the particular income regime or distribution under investigation. A number of writers, including Solow (1951), McCall (1973), Hart (1976a,b, 1983) and Schluter (1998), adopt this approach and construct size transition matrices. The advantage of this type of transition matrix is that it reflects income movement between different income levels; thus both the exchange of positions of individuals and economic growth (the increasing availability of positions at high income levels) are incorporated into mobility. One can draw welfare implications of mobility directly from comparisons of transition matrices of this type. We argue that size transition matrices are necessary for applying both the Atkinson-Dardanoni condition and the Benabou-Ok condition. Welfare implications of these dominance conditions cannot be drawn if income mobility is not associated with absolute income levels.*

5.3.2 Quantile transition matrices

The second approach views mobility as a relative concept. This approach allows the same number of individuals in each class. The resulting matrix is referred to as a ***quantile transition matrix****. The advantage of this approach is that the transition matrix is biostochastic, and the steady-state condition is always satisfied. The disadvantage is that only those movements that involve reranking (i.e. people switching positions) is recorded as mobility. Thus, the quantile matrix approach cannot take into account whether overall income is increasing or decreasing. Thus,the* upward mobility *accompanying economic growth, which Kuznets (1966) studied, is ignored. It follows that studies using this type of transition matrix cannot draw a complete picture of changes in social welfare between different income regimes. Both Hart*

(1983) and Atkinson et al. *(1992) voice concerns about the use of the quantile approach for this reason.*

π^* is a constant vector in this case, because by definition the quantiles have the same probability, 1/10 for deciles.

5.3.3 Mean transition matrices

The third and fourth approaches incorporate elements of both the absolute and relative approaches to mobility. Class boundaries are defined as percentages of mean income or median income of the initial and ending distributions. The resulting matrices are, respectively, referred to as ***mean transition matrix*** *and* ***median transition matrix****. In an early study, Thatcher (1971) uses the mean transition matrix in his analysis of the UK earnings mobility. Atkinson et al. (1992) argue that Thatcher's approach relates income mobility to both income level and the relative positions of individuals. Trede (1998) and Burkhauser et al. (1998) consider the median transition matrix in their investigations of income/earnings mobility in the United States and Germany.*

5.4 A UK example of income mobility measurement

In his Presidential address to the European Society for Population Economics, Jenkins (2000) underlines that the income distribution in the UK has experienced great changes during the eighties, but that since 1991, this distribution seems to have remained relatively stable. If the poverty line were defined as half the mean income, the percentage of poor households would remain relatively stable, while if it were defined as half the mean of 1991 in real terms, this percentage would steadily decrease. The Gini coefficient remains extremely stable around 0.31-0.32 over the period. These figures characterize a cross-section stability in income.

However, since 1991, the UK has collected the British Household Panel Survey (BHPS). This means that the same households are interviewed each year. It then becomes possible to study income dynamics. Jenkins provides an estimation for a transition matrix between income groups at a distance of one year. These groups are defined by reference to a fraction of the mean, a fraction taken between 0.5 and 1.5. In Table 5, we have in the lines groups for wave $t-1$, and in the columns groups for wave t. This transition matrix is quite illuminating for our purpose.

First of all it clearly shows the interest there is in studying the dynamics of inequality. Society is not rigid. Households are moving between income groups. And this despite the fact that the Gini coefficient remained roughly

Table 5: Transition probabilities

Income group	Period t					
	< 0.5	0.5-0.75	0.75-1.0	1.0-1.25	1.25-1.5	> 1.5
Period $t-1$						
< 0.5	**0.54**	0.30	0.09	0.04	0.02	0.02
0.5-0.75	0.15	**0.56**	0.21	0.05	0.01	0.02
0.75-1.0	0.05	0.19	**0.48**	0.20	0.05	0.03
1.0-1.25	0.03	0.06	0.20	**0.44**	0.20	0.07
1.25-1.5	0.02	0.03	0.08	0.25	**0.35**	0.27
> 1.5	0.01	0.02	0.04	0.06	0.12	**0.75**

the same over the period of estimation. Mobility is revealed when considering panel data sets.

The second remarkable point in this matrix is that the probability of moving between groups is not uniform. When you are in the two poorest groups, there is a larger probability to stay in the same group than to climb up ($\simeq 55\%$). Conversely, when you are in the richest group, the probability to stay in that group is even much larger (75%). When you belong to the higher middle class, there is a larger probability to drop in a lower group the next period than to climb up in a higher group.

References

Bartholomew, D. (1973). *Stochastic Models for Social Processes.* Wiley, London, second edition.

Bartholomew, D. (1982). *Stochastic Models for Social Processes.* Wiley, London, third edition.

Erikson, R., Goldthorpe, J. H., and Portocarrero, L. (1979). Intergenerational class mobility in three Western European societies: England, France and Sweden. *British Journal of Sociology*, 30:415–441.

Feller, W. (1950, 1968). *An Introduction to Probability Theory and Its Applications.* Wiley, New-York, 3rd edition.

Formby, J. P., Smith, W. J., and Zheng, B. (2004). Mobility measurement, transition matrices and statistical inference. *Journal of Econometrics*, 120(1):181–205.

Guilbaud, O. (1977). Estimating the probability eigenvector and related characteristics of an ergodic transition matrix. *Scandinavian Journal of Statistics*, pages 97–104.

Jenkins, S. P. (2000). Modelling household income dynamics. *Journal of Population Economics*, 13:529–567.

Prais, S. (1955). Measuring social mobility. *Journal of the Royal Statistical Society. Series A*, 118:56–66.

Shorrocks, A. F. (1978). The measurement of mobility. *Econometrica: Journal of the Econometric Society*, pages 1013–1024.

Sommers, P. and Conlisk, J. (1979). Eigenvalue immobility measures for markov chains. *Journal of Mathematical Sociology*, 6:253–276.