The Econometrics of Income and Social Mobility

Inference

Michel Lubrano

January 2019

Contents

1 Introduction

We have reported some empirical applications using transition matrices, but we have said nothing on how to make inference for those matrices. There is also the more delicate question of testing for instance the regularity of a transition matrix, an important topic which is preliminary to any other checking of the progressivity of any transition matrix. Finally, transition matrices are descriptive by nature. It relates to the comparison of two income distribution and stochastic dominance. Nothing is said about the determinants of each distribution or of the joint distribution of y_1 and y_2. We shall introduce a specific econometric model which help to shed some light for introducing explanatory variable and thus modelling these dynamic processes.

2 The estimation of transition matrices under usual assumptions

Provided the usual assumptions stated in Shorrocks (1976) are verified, that is:

1. Population homogeneity,
2. First-order Markov,
3. Time homogeneity,

the estimation of a transition matrix is a simple task. It is based on the fact that each row is of a transition matrix defines an independent multinomial process.

2.1 A multinomial model

Anderson and Goodman (1957) or Boudon (1973, pages146-149) among others proved that the maximum likelihood estimator of each element of a transition matrix P is:

$$\hat{P} = [\hat{p}_{ij}] = \left[\frac{n_{ij}}{n_i}\right].$$

This estimator $\hat{p}_{ij}$ is consistent and has variance, using the properties of the multinomial process:

$$\frac{n_i p_{ij}(1-p_{ij})}{n_i^2} = \frac{p_{ij}(1-p_{ij})}{n_i}.$$

When the sample size n tends to infinity, each row P_i of P tends to a multivariate normal distribution with:

$$\sqrt{n_i}(\hat{P}_i - P_i) \xrightarrow{D} N(0, \Sigma_i),$$

where

$$\Sigma_i = \begin{pmatrix} \frac{p_{i1}(1-p_{i1})}{n_i} & \cdots & -\frac{p_{i1}p_{ik}}{n_i} \\ & \ddots & \\ -\frac{p_{ik}p_{i1}}{n_i} & \cdots & \frac{p_{ik}(1-p_{ik})}{n_i} \end{pmatrix}.$$

As each row of matrix P is independent of the others, the stacked vector of the rows P_i verifies:

$$\sqrt{n}(\mathrm{vec}(\hat{P}) - \mathrm{vec}(P)) \xrightarrow{D} N(0, \Sigma),$$

where

$$\Sigma = \begin{pmatrix} \Sigma_1 & \cdots & 0 \\ 0 & \ddots & \\ 0 & \cdots & \Sigma_k \end{pmatrix} \tag{1}$$

is a $k^2 \times k^2$ block diagonal matrix with Σ_i on its diagonal and zeros elsewhere.

2.2 Transition matrices and incomplete panels

When using incomplete panels, there are of course problems and the usual estimation method cannot be used. We have to cope with missing values. Sherlaw-Johnson et al. (1995) propose using the EM algorithm to cope with missing observations. The missing values are predicted, then conditionally on these prediction, the transition matrix is re-estimated. In a Bayesian framework, a Gibb sampler would also be a natural solution.

2.3 Distribution of indices

A mobility index $M(.)$ is a function of the transition matrix P. Thus its natural estimator of an index will be:

$$\hat{M}(P) = M(\hat{P}),$$

i.e. a deterministic function of the estimated transition matrix. A standard deviation for that estimator will be given by a transformation of the standard deviation of the estimators of each elements of the transition matrix P. As the transformation $M(.)$ is most of the time not linear, we will have to use the Delta method to compute it. Let us recall the definition of Delta method in the multivariate case.

Definition 1 *Let us consider a consistent estimator b of $\beta \in \mathbb{R}^m$ such that:*

$$\sqrt{n}(b - \beta) \xrightarrow{D} N(0, \Sigma).$$

Let us consider a continuous function g having its first order derivatives. The asymptotic distribution of $g(\beta)$ is given by:

$$\sqrt{n}(g(b) - g(\beta)) \xrightarrow{D} N(0, \nabla g(\beta)' \Sigma \nabla g(\beta)),$$

where $\nabla g(\beta)$ is the gradient vector of g evaluated at β.

Let's verify that the mobility index $M(.)$ fulfills the Delta method assumptions. First we have shown previously that $\hat{P}$ is a consistent estimator of P. Then, from Trede (1999) we have that the asymptotic distribution of $\hat{P}$ is normal with independent rows: each row follows a multinomial distribution, hence for $n \to \infty$:

$$\sqrt{n}(\text{vec}(\hat{P}) - \text{vec}(P)) \xrightarrow{D} N(0, \Sigma),$$

where Σ is defined in (1).

Therefore the Delta method is applicable and we can derive then that:

$$\sqrt{n}(M(\hat{P}) - M(P)) \to N(0, \sigma_M^2),$$

with

$$\sigma_M^2 = (DM(P))\Sigma(DM(P))'.$$

Moreover,

$$DM(P) = \frac{\partial M(P)}{\partial \text{vec}(P')'}$$

Table 1: Transition matrix mobility measures and their derivative

Index	DM(P)
M_P	$-\frac{1}{m-1}vec(I)'$
M_E	$-\frac{1}{m-1}vec(\sum_i \check{P}_{\lambda}^{i})'$
M_D	$-sign(det(P))vec(\widetilde{P'})'$
M_2	$-vec(\check{P'}_{\lambda_2})'$
M_B	$\left[(\sum_{ij} p_{ij}\pi_s(z_{ti}-z_{mi})\lvert i-j\rvert) + \pi_s(\lvert s-t\rvert - \lvert s-m\rvert)\right]_{s,t=1\ldots m}$
M_U	$-\frac{m}{m-1}\left[(\sum_i \pi_s(z_{ti}-z_{mi})(1-p_{ii})) - (\delta_{st}\pi_s - \delta_{sm}\pi_m\right]_{s,t=1\ldots m}$

$\tilde{P}$ is the matrix of cofactors of P, $\check{P}_{\lambda} = \frac{\partial \lvert\lambda\rvert}{\partial P}$, Z is a fundamental matrix of P, $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$.

is a m^2 vector and vec(P) is the row vector emerging when the rows of P are put next to each other.

Trede (1999) has computed the derivation $DM(P)$ for several mobility indices and has summarized them in Table 2 to make easy asymptotic estimation of these mobility indices.

Obviously, $DM(P)$ and Σ are unknown and need to be estimated using the estimation of the matrix $\hat{P} = [\hat{p}_{ij}]$ and $\hat{p}_i$. Therefore we replace each element p_{ij} in $DM(P)$ and in Σ by its estimator $\hat{p}_{ij}$. Thus an estimation of σ_M would be $\hat{\sigma}_M^2 = (DM(\hat{P}))\hat{\Sigma}(DM(\hat{P}))'$.

3 The dynamic multinomial logit as an alternative

Following the results of Shorrocks (1976), it is evident that the three basic assumptions stated above are rarely fulfilled. We are in a way obliged to follow a different route as that suggested by Shorrocks because we assume that the first order Markov property is verified. The assumptions we are going to relax are the population homogeneity and the time homogeneity. The solution is quite simple as it relies on the dynamic multinomial logit model, which is just an extension of the original multinomial process. The advantage of this model is that it allows to introduce covariates to model observed individual effects and time effects. We shall leave aside the case of unobserved individual effects.

3.1 The model

To introduce observed heterogeneity, we have to consider a dynamic multinomial logit model which explains the probability that an individual i will be in state k when he was in state j in the previous period as a function of exogenous variables. Using the model of Honoré and Kyriazidou (2000) and Egger et al. (2007), but without individual effects, the unobserved propensity to select option k among K possibilities can be modelled as:

$$s_{kit}^* = \alpha_k + x_{it}\beta_k + \sum_{j=1}^{K-1} \gamma_{jk} \mathbb{1}\{s_{i,t-1} = j\} + \epsilon_{kit}. \tag{2}$$

The observed choice s_{it} is made according to the following observational rule

$$s_{it} = k \text{ if } s_{kit}^* = \max_l (s_{lit}^*),$$

which corresponds to the random utility model. If the ϵ_{kit} are identically and independently distributed as a Type I extreme value distribution, then the probability that individual i is in state k at time t when he was in state j at time $t-1$ is given by:

$$p_{jk} = \Pr(s_{it} = k | s_{i,t-1} = j, x_{it}) = \frac{\exp(\alpha_k + x_{it}\beta_k + \gamma_{jk})}{\sum_{l=1}^{K} \exp(\alpha_l + x_{it}\beta_l + \gamma_{jl})}, \tag{3}$$

where x_{it} are explanatory the variables. α_k is a category specific constant common to all individuals. γ_{jk} is the coefficient on the lagged dependent variable attached to the transition between state j to state k. As the probabilities have to sum to 1, we must impose a normalisation. We can chose

$\alpha_K = \gamma_K = 0$, $\beta_K = 0$. Equation (3) is central to our analysis.

The interpretation of the coefficients is rather complex in term of **odd ratios**. These are as follows. We have first that:

$$\frac{\Pr(s_{it} = k|s_{i,t-1} = j)}{\Pr(s_{it} = K|s_{i,t-1} = j)} = \exp(\alpha_k + x_{it}\beta_k)\exp(\gamma_{jk}).$$

We can then compute:

$$\frac{\Pr(s_{it} = k|s_{i,t-1} = K)}{\Pr(s_{it} = K|s_{i,t-1} = K)} = \exp(\alpha_k + x_{it}\beta_k)\exp(\gamma_{Kk}).$$

As $\gamma_{Kk} = 0$, the ratio of the two above expressions is equal to: $\exp(\gamma_{jk})$,

$$\exp(\gamma_{jk}) = \frac{\Pr(s_{it} = k|s_{i,t-1} = j)}{\Pr(s_{it} = K|s_{i,t-1} = j)} \Bigg/ \frac{\Pr(s_{it} = k|s_{i,t-1} = K)}{\Pr(s_{it} = K|s_{i,t-1} = K)}, \tag{4}$$

which gives the interpretation of this coefficient. So $\exp(\gamma_{jk})$ refers to the ratio of the odds of being in status k compared to the baseline status K when having been in status j in the previous period over the same odds when having been in baseline status K in the previous period.

It is simpler to consider **marginal effects** which are defined as:

$$\frac{\partial \Pr(s = k|s = j)}{\partial x} = \Pr(s = k)[\beta_k - \sum_l \Pr(s = l)\beta_l].$$

In the right hand side of this formula, $\Pr(s = k)$ is given by (3). Of course, this probability is a function of the vectors of exogenous variables. As we need a single number for the marginal effect, we have to compute $\Pr(s = k)$ at the mean value of each exogenous variable.

3.2 Implicit transition matrices

The question now is to recover the implicit transition matrix that is contained in the dynamic multinomial logit model. We have to exploit the conditional probabilities given (3) to reconstruct the first $K - 1$ lines of the transition matrix P. Then we use the identification restrictions $\alpha_K = \gamma_K = 0$, $\beta_K = 0$ for the last line of the transition matrix. The last column of the matrix is found using the constraint that each line sums up to 1. Of course, in order to obtain a single probability, we have to take the covariates at their sample mean. We thus obtain an average transition matrix. If we want to illustrate individual effects, for instance the difference between males and females, we have to compute two different transition matrices, one for each gender.

3.3 An illustration

For an application using the BHPS, we have estimated a dynamic multinomial logit model for explaining the transition between three different job statuses: working, unemployed and not working. We explained the transition between these three states by age, gender, education and time dummies. We chose “non-participating” as the baseline.

Table 2: Estimation of a dynamic Multinomial Logit model for job status transitions

			Marginal effects	
Destination status	Working	Unemployed	Working	Unemployed
Origin: Working	4.394 (0.032)	2.014 (0.058)	0.191	-0.064
Origin: Unemployed	1.855 (0.054)	3.085 (0.068)	0.018	0.036
intercept	3.804 (1.726)	19.702 (2.179)		
$\log age$	-2.205 (0.975)	-10.849 (1.242)	0.172	-0.245
$(\log age)^2$	0.332 (0.137)	1.466 (0.176)	-0.021	0.032
high educ	0.751 (0.038)	-0.154 (0.055)	0.047	-0.025
mid educ	0.459 (0.035)	-0.145 (0.048)	0.030	-0.017
gender	-2.113 (0.049)	-2.495 (0.057)	-0.049	-0.013
N. Obs	115 991			
log-likelihood	-31 965			

We used the routine `vglm` of the package `VGAM` of `R` to estimate this model. Observations are pooled. Standard errors in parentheses. Year dummies were included, but are not displayed.

We derived from these estimated coefficients two transition matrices, one for males, one for females, computed at the mean value of the other exogenous variables. We report the results in Table 3. If the average of these

Table 3: Implicit conditional transition matrices

	Working	Unemployed	Non-particip.
		Males	
Working	0.973	0.021	0.005
Unemployed	0.533	0.429	0.038
Non-particip.	0.591	0.038	0.263
		Females	
Working	0.943	0.015	0.043
Unemployed	0.466	0.264	0.269
Non-particip.	0.207	0.269	0.757

two matrices look pretty the same as the marginal matrix estimated in the usual way under the three assumptions of Shorrocks, there are huge differences between males and females for the unemployed and the not working lines. Males are almost always participating. Their most likely alternative is between working or being unemployed. Females mostly do not stay unemployed. They either go back to work or leave the labour market. When they have left the labour market, they have a strong tendency to stay in that state.

3.4 Specification tests

Using a dynamic multinomial logit is also a convenient way of testing two of the three usual assumptions concerning Markov transition matrices.

The first assumption to test is the presence of observed individual heterogeneity (*population homogeneity*). A LR test between a pure dynamic model (Log Lik. = -34 064) and a the same model with exogenous variables, but no time dummies (Log Lik. = -32019) give a statistics of 4090 with 10 degrees of freedom, so the pure dynamic model is rejected with a P value of 0.000. We had an example of individual heterogeneity with the two transition matrices for males and females which are clearly different.

The second assumption is *time homogeneity*. This can be tested by comparing the dynamic model with observed individual effects (Log Lik. = -32019) and the same model with time dummies (Log Lik. = -31965) (one for each year). The LR test has a statistics of 108 with 32 degrees of freedom and again a P value of 0.000. The rejection of time homogeneity can be explained by the business cycle. The transition between working status and unemployed status highly depends on the economic activity in the short term, but is also depends on the effects of globalisation in a longer term.

4 Hart (1976) an alternative model for income dynamics

In a quite old paper Hart (1976) investigates the question of comparing income distributions, both in a static and in a dynamic framework. This type of investigation can be related to the work of Atkinson (1970) in a static framework and that of Atkinson et al. (1992) in a dynamic framework. The model of Hart (1976) is quite simple as it relies on the lognormal distribution. The question is then to know why he did not use the usual tools of stochastic dominance that are available for instance in Levy (1973) for the lognormal process. We shall comment the paper of Hart (1976) at the light of these new tools.

4.1 The lognormal process

If y follows a normal process, then $x = \exp(y)$ follows a lognormal process, so:

$$f(x|\mu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} \exp -\frac{(\log(x) - \mu)^2}{2\sigma^2}.$$

Its mean and variance depends on both parameters

$$\mathrm{E}(x) = \exp(\mu + \sigma^2/2) \qquad \mathrm{Var}(x) = (\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2).$$

4.2 Static criterion

How to rank income distributions? A Pareto criterion would validate the preference for a distribution where possibly only the rich have increased their income. This would not be felt as fair. A Rawlsian criteria would prefer an income distribution where the income of the poor has risen, irrespective to the other incomes. Clearly, we have to take into account both the increase of the mean and the change in inequality.

A lognormal distribution is noted:

$$x_t \sim \Lambda(\mu, \sigma^2),$$

where the two parameters are the mean and the variance of the log observations. The arithmetic mean of x, noted $\bar{x}$ (or α in Hart) is given by $\exp(\mu + \sigma^2/2)$. It can be increased by μ and by σ. Hart distinguishes several cases for ranking lognormals, depending on the ordering of the respective parameters. Its reasoning is based on the arithmetic mean and the coefficient of variation $\eta = \exp(\sigma^2) - 1$. The arithmetic mean can be increased by increasing μ and also σ. The Lorenz curve and the Gini index are a function of only σ, irrespective to the value of μ in the lognormal process.

1. Case I: $\bar{x}_1 = \bar{x}_2$ and $\sigma_1 < \sigma_2$, this implies that $\mu_1 > \mu_2$. There is a clear preference for distribution 1.

2. Case II: $\bar{x}_1 > \bar{x}_2$ and $\sigma_1 < \sigma_2$, so that $\mu_1 > \mu_2$, many economists will prefer x_1 to x_2 even though it is possible that $var(x_l) > var(x_2)$.

3. Case III: If $\bar{x}_1 < \bar{x}_2$ and $\sigma_1 < \sigma_2$, there will be little agreement on ranking of the two distributions.

These three cases raised by Hart are difficult to interpret. It seems more reasonable to use the more recent tools developed in the literature of inequality measuring, Lorenz ordering and stochastic dominance.

1. The Lorenz curve has an analytical expression which depends only on the parameter σ. For two income distributions with parameters σ_1 and σ_2, the first distribution is preferred to the second one in term of Lorenz ordering iff $\sigma_1 < \sigma_2$, whatever the value of μ and $\bar{x}$. Of course this ordering takes into account only the inequality and not the income growth which can decrease with a decrease of σ as shown using the arithmetic mean. Usually this criteria is used to compare distributions having the same mean. So for $\bar{x}$ to be held constant, we must have an increasing μ if σ is decreasing. We are back to case I of Hart (1976). Case II can be preferred by the poor and not by the upper classes.

2. Standard welfare functions have two arguments: mean income and inequality. If we want to take into account the combination of these two parameters to compare income distributions, we have to consider stochastic dominance, following Atkinson (1970). We have access to analytical results for the lognormal process with Levy (1973). The main result is that stochastic dominance at the order one is obtained for the lognormal process when $\sigma_1 = \sigma_2$ and $\mu_1 > \mu_2$. For $\sigma_1 \neq \sigma_2$, the dominance curves may intersect and so there is no longer any general result at the order one for comparing the two income distributions. An income distribution will be thus preferred according to the criteria of stochastic dominance at the order one when $\sigma_1 = \sigma_2$ and $\mu 1 > \mu_2$. This case is not present in Hart (1976).

4.3 Dynamic criteria

The measurement of income inequality is made difficult in a dynamic context because individual at various stages of their life cycle are mixed. And poor and rich individuals have different life expectancy. In order to get rid of these difficulties, Hart (1976) make use of a special sample which is composed of a cohort of 800 individuals of the same age (30 years in 1963) which is observed till 1971. This sample suggests that the increase in the inequality of incomes of the same people is significant. Between 1963 and 1966 inequality increased only by 4 per cent. But between 1966 and 1970, there was an acceleration as the increase went up to 18 per cent. The use of this type of sample is designed to cope with some individual effects. The question is then to investigate who stays in the same income quantile, if the poorest remains the poorest or if there is a reshuffling of the positions.

We supposed that the initial distribution was lognormal. We have to use a dynamic model that preserves this property for the next period, which means that if y_t is lognormal, y_{t+1} will be also log normal. This model is very simple to find as it is simply the Galton-Markov model that was used extensively for instance by Atkinson et al. (1992) for analysing the dynamics of income. This model is based on the multiplicative property of the lognormal process. If $x \sim \Lambda(\mu, \sigma^2)$ then $y = ax^b$ will also be log normal with $y \sim \Lambda(a + b\mu, b^2\sigma^2)$. The Galton-Markov model essentially considers the evolution of the demeaned log income (log income minus the mean of the log) as a function of past income plus a random noise. The model states that:

$$y_t = \beta y_{t-1} + \epsilon_t \qquad \epsilon_t \sim N(0, \omega^2) \qquad y_t = \log x_t - \overline{\log x_t}.$$

The original model does not contain a constant term or drift because the variable is taken in deviation to its mean. When $\beta < 1$, the process is mean reverting and:

$$y_t = \beta^t y_0 + \sum_{i=1}^{t-1} \beta^i \epsilon_{t-i}.$$

We deduce that

$$Var(y_t) = \frac{\sigma_\epsilon^2}{1 - \beta^2}.$$

A key parameter will be the correlation between y_0 and y_1:

$$\rho^2 = 1 - \frac{\omega^2}{Var(y_1)}.$$

Here again, there is a long discussion around different polar cases defined by different parameter configurations.

However, it would be better once again to note that in the second period, the distribution of y_{t+1} is still lognormal with

$$y_{t+1} \sim \Lambda(\log(a) + \beta\mu, \beta^2\sigma^2 + \omega^2).$$

We then have to compare two lognormal distributions using the same tools as before. We modify slightly the Galton model so as to have

$$y_t = \log a + \beta y_{t-1} + \epsilon_t$$

which means adding the drift $\log a$.

1. Mobility is measured by ω^2, the variance of the dynamic system. This will be a key parameter.

2. Inequality is decreasing in the sense of a Lorenz ordering if

$$\sigma^2 < \beta^2\sigma^2 + \omega^2 \Rightarrow \sigma^2(1-\beta^2) < \omega^2.$$

3. We have stochastic dominance at the order one if

$$\sigma^2 = \beta^2\sigma^2 + \omega^2 \Rightarrow \sigma^2(1-\beta^2) = \omega^2 \qquad \mu < \log a + \beta\mu \Rightarrow \mu(1-\beta) < \log a$$

Hart (1976) estimates his parameters using the above described data set of 800 workers of the same age. He comments the evolution of the dynamics of inequality between 1963-66 and 1966-70.

References

Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about Markov chains. *The Annals of Mathematical Statistics*, 28(1):89–110.

Atkinson, A. (1970). On the measurement of inequality. *Journal of Economic Theory*, 2:244–263.

Atkinson, A. B., Bourguignon, F., and Morrisson, C. (1992). *Empirical Studies of Earnings Mobility*, volume 52. Harwood Academic Publishers.

Boudon, R. (1973). *Mathematical Structures of Social Mobility.* Elsevier Scientific Pub. Co., Amsterdam and New York.

Egger, P., Pfaffermayer, M., and Weber, A. (2007). Sectoral adjustment of employment to shifts in outsourcing and trade: evidence from a dynamic fixed effects multinomial logit model. *Journal of Applied Econometrics*, 22(3):559–580.

Hart, P. E. (1976). The comparative statics and dynamics of income distributions. *Journal of the Royal Statistical Society. Series A (General)*, 139(1):108–125.

Honoré, B. E. and Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68(4):839–874.

Levy, H. (1973). Stochastic dominance among log-normal prospects. *International Economic Review*, 14(3):601–614.

Sherlaw-Johnson, C., Gallivan, S., and Burridge, J. (1995). Estimating a Markov transition matrix from observational data. *The Journal of the Operational Research Society*, 46(3):405–410.

Shorrocks, A. F. (1976). Income mobility and the markov assumption. *The Economic Journal*, pages 566–578.

Trede, M. (1999). Statistical inference for measures of income mobility. *Jahrbucher fur Nationaloknomie und Statistik*, 218:473–490.