TMD DISCUSSION PAPER NO. 26

ESTIMATING INCOME MOBILITY IN COLOMBIA USING MAXIMUM ENTROPY ECONOMETRICS

Samuel Morley
Economic Commission for Latin America and the Caribbean

Sherman Robinson
Rebecca Harris
International Food Policy Research Institute

Trade and Macroeconomics Division
International Food Policy Research Institute
2033 K Street, N.W.
Washington, D.C. 20006 U.S.A.

May 1998

TMD Discussion Papers contain preliminary material and research results, and are circulated prior to a full peer review in order to stimulate discussion and critical comment. It is expected that most Discussion Papers will eventually be published in some other form, and that their content may also be revised.

ESTIMATING INCOME MOBILITY IN COLOMBIA USING MAXIMUM ENTROPY ECONOMETRICS

by

Samuel Morley
Economic Commission for Latin America and the Caribbean

Sherman Robinson
Rebecca Harris
International Food Policy Research Institute

ABSTRACT

Income mobility can be viewed as a first-order Markov process, with a matrix of transition probabilities which measure how individuals move from an income status in time t to a new status in time $t+1$. Direct estimation of transition matrices is difficult, since time series panel data are unavailable and limited data on the distribution of income do not suffice to determine the coefficients mathematically, let alone provide enough degrees of freedom for estimation. In this paper, we show that maximum entropy econometrics offers a feasible way to estimate transition matrices using distributional data from Colombia. Using a cross-entropy estimation method, we make efficient use of prior information about the structure of the transition matrices and how they vary with age. The approach is very flexible, allowing the use of "information" in a variety of forms such as inequality constraints, errors in measurement, and prior estimates. Under weak assumptions about the error generation process, we can derive test statistics based on the likelihood ratio measuring the significance of the estimation. The model fits the data well in that the predicted and actual distributions for period $t+1$ are close. The results show that there is a large degree of upward mobility in Colombia, especially at the bottom of the income distribution and for the younger age cohorts.

Key words: Maximum entropy econometrics, income mobility, Colombia

TABLE OF CONTENTS

ESTIMATING INCOME MOBILITY IN COLOMBIA USING MAXIMUM ENTROPY ECONOMETRICS

Introduction

Latin America has long been the focus of unfavorable international attention because of its high level of income inequality. Recent evidence indicates that while the resumption of economic growth in recent years has begun to reduce poverty, that has not yet had positive effects on the distribution of income. Indeed there are many who believe that the adoption of the neoliberal model has, if anything, made the distribution more unequal.

While the statistical details of the income distribution are by now quite well documented and studied, less attention has been paid to the interpretation of the observations. Interpreting changes in the distribution in terms of welfare is inherently difficult in an economy where both population and income are changing over time. Morley (1981) showed the significant effect of population growth on measures of poverty and income of those at the bottom of the distribution. With income growth, one can have substantial reduction in the extent of poverty, and presumably improvements in welfare, even if the distribution becomes substantially less equal. And, even if both aggregate income and the population are constant over time, the fact that age-earnings profiles rise and then fall (indicating "circular mobility") will yield more equally distributed lifetime incomes of individuals, even though the distribution at any point in time is highly unequal.

Standard studies of the distribution of income lack a connection between the distribution at a point in time and lifetime distributions. Our distribution statistics are snapshots taken at a point in time which do not tell us what happens to those at the bottom or the top of the distribution *over* time. Thus, we need measures of income mobility over time to connect these two concepts and to make any interpretation of the distribution in welfare terms. However, a scarcity of panel data which would permit us to follow particular individuals over time often prohibits us from estimating such measures.

In this paper, we estimate mobility in Colombia using successive household surveys adapting a procedure first suggested by Adelman *et al.* (1994) and by Telser, (1963) and others in another context. Household surveys typically give us the distribution of income earners across income classes at different points in time. We want to estimate the transition matrix; that is, a matrix showing the percentage of individuals who move from income class i at time t to class j at time $t+1$. This estimated transition matrix will provide the raw material for an analysis of income mobility for Colombia between 1988 and 1995.

Direct estimation of transition matrices is made difficult by the fact that for any N income groups, we have only N observations but must estimate N·N parameters (more precisely, (N-1)·N since the rows must sum to one). The problem is inherently underdetermined, with inadequate degrees of freedom to support parameter estimation using standard regression analysis. By making very strong assumptions about the dynamic process, Adelman *et al.* (1994) specified a model with enough degrees of freedom to support estimation of transition matrices for Brazil using an approach that approximated a SUR regression model. In order to estimate, they used a lot of "prior information" by assuming specific functional forms, constancy of parameters across disparate groups, and a particular distribution for the errors.

In this paper, we propose a more flexible "maximum entropy" approach to estimating transition matrices that is based on information theory (Golan, Judge, and Miller, 1996). Using a "cross entropy" estimation method, we make efficient use of prior information about the structure of the transition matrices and how they vary with age. The approach is very flexible and sparse in its assumptions, allowing the use of "information" in a variety of forms such as inequality constraints, errors in measurement, and prior estimates of various parameters. The estimation approach makes almost no assumptions about the error generation process. Under weak assumptions about the distribution of the errors, the method can be shown to have maximum likelihood properties and we can derive test statistics based on the likelihood ratio to measure the significance of the estimation.

The Markov Process

We have income distribution data from Colombia for three three-year transitions (1988-1991, 1991-1994, and 1992-1995) and for sixteen two-year age groups (from ages 20-21 to 50-51). Our estimation problem is to find the transition probabilities for each age group. With these estimated parameters, we can explore how income mobility varies with age and also project, for each age group, what will be the distribution in time *t+1* given the distribution in time *t*.

The Markov process for each transition can be expressed as follows:

$$Y_{i,a,t} = \sum_{j} P_{j,i,a,t} \cdot X_{j,a,t} \qquad (1)$$

where *X* is the income distribution for the first year in the transition, *Y* is the income distribution in the second year, *i* and *j* refer to income classes, *a* refers to age group, and *t* refers to time period.

With 16 age groups and 5 income categories, there are 80 observations for each transition. Since our data are given in relative terms, the *X* and *Y* vectors sum to one, so there are really only 64 (*ie.* 4·16) independent observations for each transition. On the other hand, *P* is a 5-by-5 matrix for each of the 16 age groups and, given that the rows must sum to one, 20 cells in each transition matrix need to be estimated, or 320 parameters in all for each transition. The problem is greatly underdetermined — the observations do not suffice even to uniquely determine a set of parameters mathematically, let alone provide degrees of freedom for statistical estimation.

Maximum Entropy Estimation

This is a classic problem for using an entropy estimation approach based on information theory (see Golan, Judge, and Miller, 1996). Following Shannon (1948) and Jaynes (1957), define the entropy measure of the information content of the transition matrix (ignoring the age and time subscripts for the moment):

$$H(P) = -\sum_i \sum_j P_{i\rangle j} \ln P_{i\rangle j} \qquad (2)$$

where

$0 < P_{i\rangle j} < 1$, $\sum_j P_{i\rangle j} = 1$, and $P_{i\rangle j} \ln P_{i\rangle j} = 0$ if $P_{i\rangle j} = 0$.This function reaches a maximum when $P_{i\rangle j} = \frac{1}{N}$ for all j (where N is the number of income classes), in which the transition probabilities are uniformly distributed for each row. In information theory, the uniform distribution has maximum uncertainty.

The philosophy of maximum entropy (ME) estimation can be stated in two principles: (1) use all the information you have, and (2) do not assume (or use) any information you do not have. The first principle will lead to an approach that seeks to use a variety of information in a flexible way. The second implies an estimation criterion that can be described as the principle of maximum uncertainty or maximum entropy. Kapur and Kesavan (1992, p. 7) state the principle: “Out of all probability distributions consistent with a given set of constraints, choose the one that has maximum uncertainty.” These principles lead to estimation as a constrained maximization problem — find a set of parameters that are as uncertain as possible, using the entropy metric, but that are consistent with all the information you have in the form of constraints.

Since the uniform distribution maximizes uncertainty (measured by entropy), the second principle says that if we have no information other than each probability is positive and they sum to one, then we should choose the uniform distribution. We have no reason to choose any other distribution. This concept is known as “Laplace's principle of insufficient reason.” The estimation approach can then be restated as (Kapur and Kesavan, 1992, p. 12): “Out of all probability distributions satisfying given constraints, choose the distribution that is closest to the uniform distribution.” This restatement requires a measure of the “distance” between two probability distributions.

Kullback and Leibler (1951) present a "cross entropy" measure that is widely used which allows the incorporation of information in the form of "prior" knowledge about the distribution.

Assume that we have some degree of "prior" knowledge of the values of P, which is given by $\bar{P}$. The estimation approach is to minimize the Kullback-Leibler "cross entropy" measure of the distance (*I*) between the estimates of the elements of *P* and the prior:

$$I(P:\bar{P}) = \sum_i \sum_j P_{i,j}(\ln P_{i,j} - \ln \bar{P}_{i,j}) \qquad (3)$$

subject to the constraints on the transitions expressed in equation (1), plus the standard constraints on the probabilities. If the prior distribution is specified as the uniform distribution, this estimation method is equivalent to maximizing entropy in equation (2). In the ME case, the approach is to "push" the estimated parameters toward maximal uncertainty, while in the CE case the approach is to find estimates that "pull" the least distance away from the (maximally uncertain) prior.

The cross entropy approach provides a flexible framework for incorporating prior information into the process of parameter estimation. If we have no prior knowledge of $\bar{P}$, then we should choose the most "uncertain" distribution, namely the uniform distribution. In this case, minimizing cross entropy is equivalent to maximizing entropy. If we can specify a (Bayesian) prior, then we can use it in the cross entropy approach. If we have specific information in the form of constraints on the values of *P*, often in the form of known moments or information about specific elements, then that information can easily be incorporated into the estimation by specifying additional constraint equations. The transition equation (1) is expressed as a weighted sum of the row probabilities, and hence can be seen as defining known moments of the row distributions.

Cross Entropy Estimation of Income Mobility in Colombia

For the Colombia case study we are using data from a national household labor force survey. We use the data to establish three-year observed transitions for 1988-91, 1991-94 and 1992-95 for sixteen two-year age cohorts starting at age 20-21. The study is confined to males only. We have divided the distribution into five income classes. Class one is zero income earners. Class two ranges from zero to \$98 per month (in 1995 dollars), class three from \$98 to \$164, class four from \$164 to \$328, and class five above \$328. In 1995 the minimum wage was about \$130 per month, and thus fell midway in our third income class.

We have the following *a priori* notions about the form of the transition matrix:

1. There is likely to be a systematic or functional relationship between age and mobility.

2. The transition matrices should be very similar, if not identical, across the three transitions for which we have data.

3. Mobility will be higher for younger age groups.

4. There will be little or no downward mobility for the upper income groups.

5. The transition matrix will be highly diagonal. In one transition period, individuals are unlikely to move more than one income class away from their starting class.

6. The transition matrix will be dominantly diagonal for older, upper income groups (*i.e.*, the diagonal element will exceed the sum of off-diagonal elements). It is unlikely to be dominantly diagonal for younger, lower income groups.

7. There is likely to be measurement error in the observed data on income distributions.

To capture the systematic relationship between age and mobility (items 1 and 3 above), we specify that the transition probabilities are a quadratic function of age, cell by cell. The function is given by:

$$P_{i,j,a,t} = \beta_{0,i,j} + \beta_{1,i,j}(AGE) + \beta_{2,i,j}(AGE)^2 \qquad (4)$$

where a subscript t is added to denote the transition period. This equation imposes a particular functional form and assumes that it applies to all three transitions. We explore the sensitivity of the estimates to this specification, trying alternative forms, and describe the results below.

To incorporate our knowledge about the nature of the transition probabilities, we constructed a prior that reflects the information described in items 2 to 6 above. One suggestion is to use a prior of all ones down the diagonal (a "spiked" prior) instead of a uniform distribution. This prior is certainly as diagonal as possible, but does not reflect our knowledge about the likely upward mobility of lower income, younger, groups. It also gives no preference for estimated parameters to be larger the closer they are to the main diagonal. Instead, we construct a "minimum mobility" prior that is consistent with the data we have. Define a measure of mobility, or the degree to which the transition matrix is off-diagonal, as follows:

$$M = \sum_i \sum_j P_{i,j} \cdot (i - j)^2 \qquad (5)$$

This measure will equal zero if the transition matrix is a diagonal matrix of ones. For each transition, we constructed a “minimum mobility” prior that minimizes *M* subject to the constraints of the distribution data for that transition and the definition of the probabilities (that they be between zero and one and sum to one by row). These priors were then used in the cross-entropy estimation.

In a Bayesian specification, the prior should reflect “prior” information that is not contained in the estimation procedure, which is being used to estimate a posterior distribution that “revises” the prior. In this case, we construct a prior that uses the data, but only to capture some of the information about the transition probabilities. The minimum mobility criterion reflects items 4 to 6, while the distribution data impose a degree of upward mobility reflecting item 3. With the construction of the prior, the mobility metric is not used in any other way in the estimation procedure.

We experimented with imposing constraints on the estimated parameters to capture prior information. For example, we specified inequality constraints on the transitions to ensure that they would decline monotonically moving away from the main diagonal. We also experimented with imposing some zeros for cells far away from the diagonal and also with imposing diagonal dominance in the upper income groups (*i.e.*, the main diagonal element is larger than the sum of the off diagonals). With the minimum mobility prior, these constraint equations turned out to be redundant — they were consistent with the data, had zero shadow prices, and had no effect on the estimated parameters. The minimum-mobility prior incorporates all the prior information in items 1 to 6.

To incorporate item 7, that the distributional data may be observed with error, we rewrite equation (1), including both age and time subscripts, to add an error term:

$$Y_{i,a,t} = \sum_{j} P_{j,i,a,t} X_{j,a,t} + e_{i,a,t} \qquad (6)$$

Here, unlike in a classical regression model, we make almost no assumptions about the distribution of the error terms. Following Golan, Judge, and Miller (1996), we write the errors as a convex weighted average of known constants that represent plausible extreme values, defining a range within which the errors will fall:

$$e_{i,a,t} = \sum_{w} W_{i,a,t,w} \bar{v}_{i,a,t,w} \qquad (7)$$

The weights, *W*, are bounded between zero and 1 and must sum to one, the *v*'s are constants (defining the support of the errors), and *w* is the index for the set of weights. If we assume that *w* represents only two weights, then the error is assumed to be drawn from a uniform distribution that ranges between the two values of *v*. If we increase the number of supporting constants, *v*, and hence possible weights, *W*, then we can potentially recover additional information about the errors, *e*, such as variance, skewness, or kurtosis. If we assume a symmetric error distribution centered on zero, then it is convenient to specify a symmetric set of supporting constants, *v*, centered on zero. In our estimation, we specified three values of *v*, supporting a distribution with both mean and variance to be estimated.

As Golan, Judge, and Miller (1996) show, this approach adds a disturbance term that is expressed in natural units, but then transforms the estimation problem into finding a set of probability weights, *W*, that can easily be incorporated into the entropy formalism. Assuming a symmetric set of supports centered on zero, if the weights are uniform, then the errors are all zero. This approach provides a bridge between the maximum entropy approach and "standard" regression models. We simply add an additional term to the cross entropy minimand in equation (3):

$$: \bar{P}) = \sum_{i\rangle j\rangle a\rangle t} P_{i\rangle j\rangle a\rangle t}(\ln P_{i\rangle j\rangle a\rangle t} - \ln \bar{P}_{i\rangle j\rangle a\rangle t}) + \sum_{i\rangle a\rangle t\rangle} W_{i\rangle a\rangle t\rangle w} \ln \qquad (8)$$

and minimize this cross entropy measure subject to constraint equations (6) and (7), as well as bounds and summing constraints on *P* and *W*.

Results

Even a cursory examination of the successive distributions for each cohort indicates that there must have been a tremendous amount of upward mobility in Colombia, particularly for younger workers. Table 1 shows the raw data for two of these cohorts over the time periods in the sample to indicate the scale of upward movement present in the data.

From the table, about 22% of 24-25 year olds were in the bottom two income classes and about 39% in the top two classes in 1988. Three years later, the percentage in the bottom had fallen to 16.2% while the percentage in the top had expanded to over 51%. If we assume no downward mobility and minimum upward mobility, this implies that 43% of 24-25 year olds in the bottom class and 47% of those in class two must have moved up by one income class. Note that if there were downward mobility, that would have increased the observed percentages in the bottom classes in 1991 and 1995, and the amount of upward mobility for those who started in the bottom classes would have had to have been higher. To have almost half of those in the bottom two classes moving up by one income class implies that mobility is an important factor to be taken into consideration in any study of the earnings distribution in Colombia.

Table 2 displays minimum cross-entropy estimates of the transition matrices for each age group, using the minimum mobility prior. Figure 1 graphs the main diagonal elements (the share of people remaining in their income group) for the first four income categories. The graph for the fifth category is simply a straight line at the value one.

The first and most important result coming from these tables is the extraordinary amount of upward mobility, particularly at the bottom of the income distribution and for the younger cohorts. Half of the twenty year olds who start in class one move up to class two over three years. Two-thirds of those in class two move up to class three, and so on.

Table 1 Distribution of workers across income classes (in percentages)

Income class		*24-25 years olds*		
	1988	**1991**	**1992**	**1995**
1	9.84	5.61	8.80	5.71
2	12.01	10.63	15.98	11.09
3	39.14	32.67	39.74	29.86
4	30.97	37.07	27.80	34.67
5	8.04	14.02	7.68	18.67
		30-31 year olds		
	1988	**1991**	**1992**	**1995**
1	4.74	3.28	5.02	3.46
2	10.11	8.98	12.81	9.55
3	29.92	28.24	32.31	25.54
4	36.63	38.73	32.31	36.38
5	18.6	20.77	17.54	25.06

Source: Special tabulations of household surveys

Note that these estimates are conservative in the sense that we have specified a minimum mobility prior. If the estimates are biased, they are biased toward less mobility. If some of those who started in the higher income classes went down over time, more of those at the bottom would have had to have moved up to take their places. For the younger cohorts, the estimates imply that the income group in which individuals begin their working life is unlikely to be where they end up.

As one might expect, upward mobility decreases systematically with age — *i.e.* the older the cohort, the closer the main diagonal is to one. Thus, while there is a lot of mobility for those in their twenties, the estimates indicate that the income class in which individuals find themselves by age 32 is more or less where they are going to stay for the remainder of their working life. 85-90% of all the individuals in the cohorts 32 to 50 remain in their same income class.

Given that we make no assumptions about the underlying error distribution, we cannot use standard measures of statistical significance. We can, however, calculate the correlation coefficient between the actual income distributions ($Y_{i,a,t}$) and the fitted distributions using the estimated P matrices. As seen in table 3, the squared correlation coefficient is 0.95, suggesting that the predictive quality of the estimates is good. This statistic is also equivalent to one minus the ratio of the error variance over total variance of Y, which is the standard measure of the "goodness of fit" of a least squares regression equation.

Golan and Vogel (1997) and Golan (1997) show the equivalence between maximum (and cross) entropy estimators and maximum likelihood estimators under sparse assumptions about the errors. Their approach is to reformulate the maximum entropy problem as its unconstrained dual, assume that the *P's* follow a logistic distribution, express the solution estimators as functions of the dual "shadow prices", and then show that these are maximum likelihood estimators. The entropy results can thus be tested in a likelihood ratio framework. For our case, the log likelihood ratio statistic is calculated as:

$$-2\left(\sum_{i,j,a,t} \bar{P}_{i,j,a,t} \ln \bar{P}_{i,j,a,t}\right)\left(1 - \frac{\sum_{i,j,a,t} P_{i,j,a,t} \ln P_{i,j,a,t}}{\sum_{i,j,a,t} \bar{P}_{i,j,a,t} \ln \bar{P}_{i,j,a,t}}\right) \qquad (9)$$

The intuition behind this statistic is that it measures how different the estimated P is from the prior. The parallel in the maximum likelihood approach is how different the constrained likelihood estimator is from the unconstrained estimator. The likelihood ratio is distributed approximately as a chi-squared distribution, with K-1 degrees of freedom, where K equals the number of income classes.

In the maximum entropy approach, the likelihood ratio indicates how "far" the data pull the estimated parameters away from the uniform, maximally uncertain, prior. A high value of the chi-square statistic indicates that the data are adding a lot of information. In the cross entropy formulation, however, the interpretation of the chi-square statistic differs from that of a standard maximum likelihood test. In this case, a larger value of the chi-square statistic indicates that the data pull the estimated parameters far from the prior values. If the prior is very good, the chi-square statistic should be small, indicating that the data are consistent with the prior. Thinking of the prior as a maintained hypothesis, a low value of the chi-square statistic is equivalent to saying that the maintained hypothesis cannot be rejected. With a chi-square value of 39.45, our estimates are significantly different from the prior. However, as we explain below, this value is lower than the chi-square resulting from a uniform prior, implying that the information content of the minimum mobility prior is valuable.

Table 2 Resulting transition matrix for each age group

age=20-21

	1	2	3	4	5
1	0.509	0.490	0.000	0.000	0.000
2	0.000	0.362	0.638	0.000	0.000
3	0.000	0.000	0.575	0.425	0.000
4	0.000	0.000	0.001	0.642	0.357
5	0.000	0.000	0.000	0.000	1.000

age=22-23

	1	2	3	4	5
1	0.593	0.406	0.000	0.000	0.000
2	0.000	0.470	0.530	0.000	0.000
3	0.000	0.000	0.636	0.364	0.000
4	0.000	0.000	0.000	0.684	0.316
5	0.000	0.000	0.000	0.000	1.000

age=24-25

	1	2	3	4	5
1	0.670	0.330	0.000	0.000	0.000
2	0.000	0.568	0.431	0.000	0.000
3	0.000	0.000	0.691	0.309	0.000
4	0.000	0.000	0.000	0.722	0.278
5	0.000	0.000	0.000	0.000	1.000

age=26-27

	1	2	3	4	5
1	0.738	0.262	0.000	0.000	0.000
2	0.000	0.656	0.343	0.000	0.000
3	0.000	0.000	0.742	0.258	0.000
4	0.000	0.000	0.000	0.757	0.242
5	0.000	0.000	0.000	0.000	1.000

age=28-29

	1	2	3	4	5
1	0.798	0.201	0.000	0.000	0.000
2	0.000	0.735	0.265	0.000	0.000
3	0.000	0.001	0.788	0.211	0.000
4	0.000	0.000	0.000	0.790	0.210
5	0.000	0.000	0.000	0.000	1.000

age=30-31

	1	2	3	4	5
1	0.851	0.149	0.000	0.000	0.000
2	0.000	0.803	0.197	0.000	0.000
3	0.000	0.001	0.830	0.169	0.000
4	0.000	0.000	0.000	0.819	0.181
5	0.000	0.000	0.000	0.000	1.000

age=32-33

	1	2	3	4	5
1	0.895	0.105	0.000	0.000	0.000
2	0.000	0.861	0.139	0.000	0.000
3	0.000	0.001	0.867	0.132	0.000
4	0.000	0.000	0.000	0.845	0.154
5	0.000	0.000	0.000	0.000	1.000

age=34-35

	1	2	3	4	5
1	0.932	0.068	0.000	0.000	0.000
2	0.000	0.908	0.091	0.000	0.000
3	0.000	0.001	0.899	0.100	0.000
4	0.000	0.000	0.001	0.869	0.131
5	0.000	0.000	0.000	0.000	1.000

Table 2, con't

age=36-37

	1	2	3	4	5
1	0.960	0.039	0.000	0.000	0.000
2	0.000	0.946	0.053	0.000	0.000
3	0.000	0.001	0.927	0.072	0.000
4	0.000	0.000	0.001	0.889	0.110
5	0.000	0.000	0.000	0.000	1.000

age=38-39

	1	2	3	4	5
1	0.981	0.019	0.000	0.000	0.000
2	0.001	0.974	0.025	0.000	0.000
3	0.000	0.002	0.950	0.048	0.000
4	0.000	0.000	0.001	0.906	0.092
5	0.000	0.000	0.000	0.000	1.000

age=40-41

	1	2	3	4	5
1	0.994	0.006	0.000	0.000	0.000
2	0.001	0.992	0.008	0.000	0.000
3	0.000	0.002	0.968	0.030	0.000
4	0.000	0.000	0.002	0.921	0.077
5	0.000	0.000	0.000	0.000	1.000

age=42-43

	1	2	3	4	5
1	0.998	0.001	0.000	0.000	0.000
2	0.001	0.999	0.000	0.000	0.000
3	0.000	0.002	0.982	0.015	0.000
4	0.000	0.000	0.003	0.932	0.065
5	0.000	0.000	0.000	0.000	1.000

age=44-45

	1	2	3	4	5
1	0.995	0.004	0.000	0.000	0.000
2	0.001	0.997	0.002	0.000	0.000
3	0.000	0.002	0.991	0.006	0.000
4	0.000	0.000	0.003	0.940	0.056
5	0.000	0.000	0.000	0.000	1.000

age=46-47

	1	2	3	4	5
1	0.984	0.015	0.000	0.000	0.000
2	0.001	0.984	0.015	0.000	0.000
3	0.000	0.003	0.996	0.001	0.000
4	0.000	0.000	0.004	0.946	0.050
5	0.000	0.000	0.000	0.000	1.000

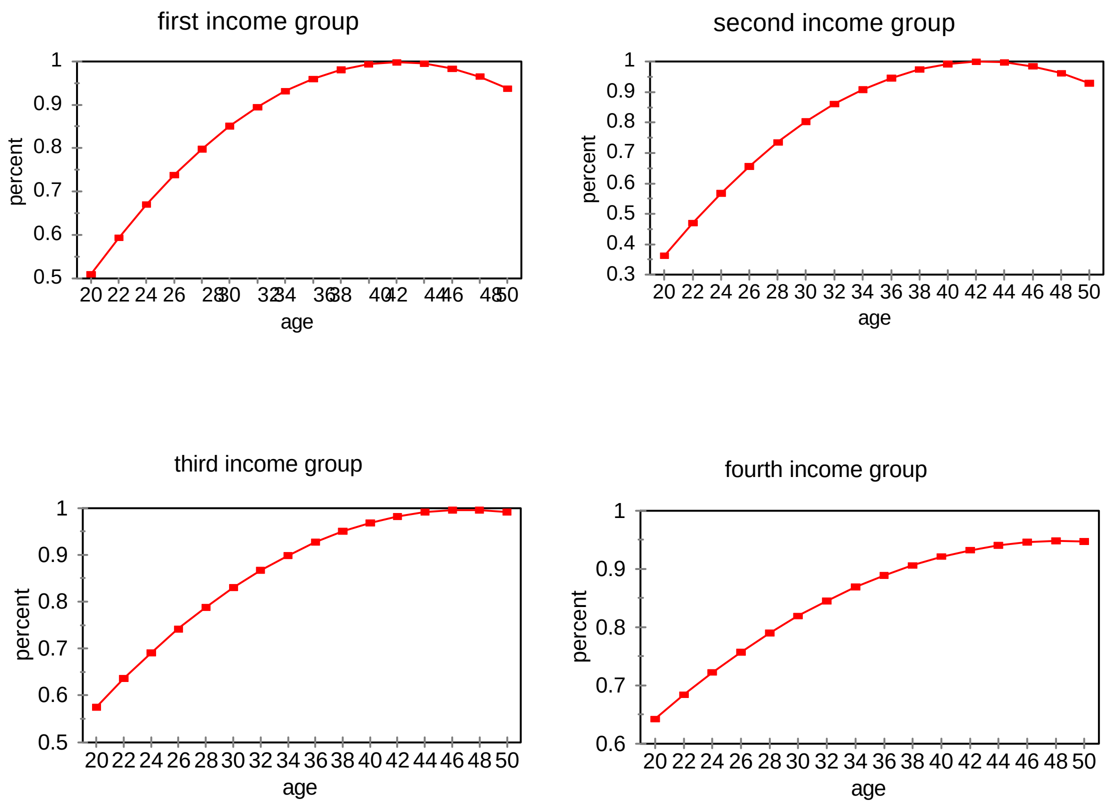
age=48-49

	1	2	3	4	5
1	0.965	0.034	0.000	0.000	0.000
2	0.001	0.961	0.038	0.000	0.000
3	0.000	0.003	0.996	0.001	0.000
4	0.000	0.000	0.005	0.948	0.047
5	0.000	0.000	0.000	0.000	1.000

age=50-51

	1	2	3	4	5
1	0.938	0.061	0.000	0.000	0.000
2	0.001	0.929	0.070	0.000	0.000
3	0.000	0.003	0.991	0.005	0.000
4	0.000	0.000	0.006	0.947	0.046
5	0.000	0.000	0.000	0.000	1.000

Figure 1 Immobility, measured by percentage of people remaining in income group



Sensitivity Analysis

Prediction versus Precision

The objective function in equation (6) accounts for entropies for the distributions of both the parameter *P* and disturbance *W*. In the estimation procedure, we have placed equal weights on these entropy distances, reflecting an equal preference for “precision” in the estimates of the parameters and the “prediction” of the estimated transitions. However, it may be better to put uneven weights on the entropy equation in the form of:

$$= \gamma \sum_{i,j,a,t} P_{i,j,a,t}\left(\ln P_{i,j,a,t} - \ln \bar{P}_{i,j,a,t}\right) + (1-\gamma) \sum_{i,a,w} (W_{i,\epsilon} \qquad (10)$$

in which the weight γ can take on a value between 0 and 1, depending on the importance we wish to attribute to prediction versus precision. It is instructive to consider the extremes. If $\gamma=0$, then the entropy measure is made up solely of the error term and the goal is to maximize “prediction”. In this case, the prior plays no role and the resulting estimates of *P* tend to be odd, with a lot of (unbelievable) mobility, and unstable (given low level of estimation degrees of freedom). On the other hand, if $\gamma=1$, the emphasis is on "precision," minimizing the entropy difference of the estimates from the prior. In this case, the estimated *P* is strongly dominantly diagonal, and there is far more error in the estimated income distributions. [Golan, Judge, and Miller (1996) report Monte Carlo experiments where they explore the implications of changing these weights and conclude that equal weighting of projection and precision is reasonable.]

Form of P as a Function of Age

To test whether the quadratic form of *P* is correct, we added a cubed term in age, yielding:

$$P_{i,j,a,t} = \beta_{0,i,j} + \beta_{1,i,j}(AGE) + \beta_{2,i,j}(AGE)^2 + \beta_{3,i,j}(AGE)^3 \qquad (11)$$

The results were not significantly different. The relationship between mobility and age stayed qualitatively the same; there were no new “peaks” added, which would have suggested that the second order equation was not sufficient. The addition of a cubic term yielded a somewhat steeper slope to the relationship between age and mobility.

The given form of the quadratic equation constrains the estimated transition matrices to be the same for all three transitions. To account for the possibility that the transition matrices differ by transition (*t*), we allowed the constant term to vary by transition. In the estimation, we simply added a t subscript to the constant in equation (4). The estimated P matrices change very little for the first and third transitions, but change significantly for the second transition. In particular, the third and fourth income groups have a significant shift.

It is instructive to compare the likelihood ratios for our different priors to evaluate the results. In Table 3, we report these statistics for estimates using the minimum mobility prior, the minimum mobility prior with the time-varying constant, and a uniform prior. The likelihood ratio when the uniform prior is used is larger than when we specify the minimum mobility prior. This result indicates that the final mobility matrix is being pulled further away from the uniform prior in the first case than from the minimum mobility prior in the second, implying that the minimum mobility prior contains useful information.

Comparing the minimum mobility estimate with and without a time-varying constant, the results support the view that the mobility process is not constant across time. In particular, comparing the different estimated values for the constant, putting a "dummy" for the 1991-94 period significantly reduces the likelihood ratio (as well as raising the R-squared or fit of the estimated distributions). However, there is no strong reason to assume that the constant varies across the samples, especially given that the time periods overlap. Since the qualitative results are similar, we report the results using the minimum mobility prior, which effectively estimates an average value for the constant.

Table 3 Results of Statistical Inference

	Minimum mobility prior	Minimum mobility prior, with time-varying constant	Uniform prior
Likelihood Ratio (chi-square statistic on estimated P)	39.45	20.76	52.2
R-squared (within sample prediction of distribution)	0.946	0.964	0.972

Implications of the Findings

The estimated transition matrices and the observed changes in the cohort distributions suggest a significant amount of upward mobility and very little downward mobility, at least through age 50. Since the younger members of the labor force at the bottom of the income pyramid moved up in large numbers, it is reasonable to ask whether this upward mobility reflected a general improvement in the distribution of income. The answer is that it does not. The Gini coefficient (see Table 4) actually rises between 1988 and 1993, and then falls back between 1993 and 1995, but to a level significantly higher than observed in 1988. The data indicate rising overall inequality coupled with significant upward mobility. How is that possible and what does it imply? We can get a good part of the answer from Table 4.

Table 4 divides the data into cohorts 20-29 and 30-55 in 1988, and then indicates changes in the distributions across income classes of the same cohorts over the period that covers our estimations of the transition matrices (1988-91, 1991-94 and 1992-95). Separating the data into cohorts in this way again makes clear the quantitative importance of upward mobility for the younger cohorts. Twelve percent of the 20-29 year old labor force was in the zero income class. A total of 24% was in the bottom two classes combined. Seven years later, only 15% were in those two classes. Meanwhile, the share of the top class almost tripled, rising from 8% to 23%. The data for the older cohorts confirms what we have already found in the estimations —mobility falls off sharply for those over 30.

Table 4 Distributions across income classes for equivalent cohorts (percent)

	1988	1991	1992	1994	1995
age cohorts	*20-29*	*23-32*	*24-33*	*26-35*	*27-36*
income class					
1	11.8	6.5	6.2	4.3	4.1
2	12.1	11.4	13.5	10.4	10.9
3	38.1	34.1	35.6	31.1	26.9
4	29.9	34.1	30.3	32.0	34.9
5	8.1	13.9	14.4	22.2	23.1
age cohorts	*30-55*	*33-58*	*34-59*	*36-61*	*37-62*
income class					
1	3.4	2.7	3.4	2.9	2.9
2	12.5	11.6	14.8	13.6	14.2
3	26.0	25.2	27.4	26.4	22.4
4	34.3	34.9	30.5	28.2	32.0
5	23.9	25.6	24.0	28.9	28.5
Gini (20-60)	0.464	0.504	0.529	0.534	0.509

Source: Special tabulations of household surveys

Given the amount of upward mobility for younger cohorts, the lack of downward mobility, and the fact that there was not a significant reduction in the total size of the low income labor force, it must be true that this low-income labor force is largely comprised of a revolving group of new entrants or young people. Workers must be entering the labor force at low salaries. As they gain experience, their earnings increase and they move up into higher income classes and are replaced at the bottom by a new group of the young. By breaking down our data by age cohorts, it is possible to see this effect directly.

Consider the 1988 and 1994 labor force from our sample. First we divide the 1988 labor force into three age groups: 20-29, 30-55, and 56-61. For 1994, take the labor force of exactly the same age (20-61). Divide this labor force into the original cohorts, now aged 26-35, 36-61, and 62-67. But now, of course there is also a set of new entrants aged 20-26. Now use this information to determine the composition of the low income labor force in 1988 and 1994. Table 5 shows the results of this calculation.

Table 5 Composition of the low income labor force by age cohort, 1988 and 1994

	1988				1994			
age cohort	20-29	30-55	56-60	total	20-26	26-35	36-61	total
# in lowest two income classes	2317	2284	354	4955	1535	1235	1854	4624
as % of total	46.8	46.1	7.1	100.0	33.2	26.7	40.1	100.0

Source: Special tabulations of household surveys

From Table 5, 47% of the workers in the bottom two income classes in 1988 were in the 20-29 age cohort, 46% in the cohorts 30-55, and the remainder in the 56-61 group. Six years later, the total number of low income workers in the sample has fallen by about 6%. But because of upward mobility, almost half of the original 20-29 cohort has moved up and out of the poverty group and a new group of the young has moved in to take its place. There is also some upward mobility for the 30-55 cohorts, but it is far less significant in either absolute terms or in the amount of upward mobility it represents.

To see this effect more clearly, in table 6 we have decomposed the total change in the number of low income workers between 1988 and 1994. Overall, there was reduction of 331 or 6% in the size of the group. More important, there was an influx of 1535 new entrants, which almost exactly offset the upward mobility of the original low income labor force, particularly of the 20-29 year old group. Thus the data confirm the results from our estimated transition matrices. The low income group is comprised to a significant extent of a group of young new entrants into the labor force. There is a great deal of turnover in the group. It is not an unchanging group of low skilled workers trapped in bad jobs with no hope of improvement. To the contrary, almost half of the 1988 low income 20-29 year olds moved up and out of the low income group in the following six years, and, if the past is any guide, an additional number will do so in the future.

Table 6 Decomposition of changes in the low income labor force, 1988-94

	Change in Number of Workers
new entrants (20-26)	+ 1535
upward mobility	
20-29 cohorts	- 1082
30-55 cohorts	- 430
older workers (out of sample or retired)	- 354
Total change in low income workers	- 331

Source: Special tabulations of the household surveys

There is a great deal of difference between comparing snapshots of the earnings pyramid at two points in time and following particular cohorts within the pyramid, mainly because new entrants to the labor force tend to be clustered at the bottom. Overall, between 1988 and 1994, real income of the 20-55 year old labor force rose by 20%. Real income of the 20-29 cohorts rose by 27%, implying a slight narrowing of the age-wage differential. But if we follow the particular cohorts of 1988 — what are called "survivors" in an earlier paper (Morley 1981) — we get a very different picture. The 20-29 cohort of 1988 had a 64% increase in real income, while the 30-55 cohort gained only 23%. The upward mobility of the young generated a very significant earnings improvement and, since the young were concentrated at the bottom of the income pyramid, there was also a narrowing of inequality among these survivors. But this improvement is hidden by our way of looking at the data, or equivalently by our not paying enough attention to upward mobility.

Conclusion

In this paper, we have shown that maximum entropy econometrics offers a feasible way to estimate income transition matrices using household surveys. It gives us a way to overcome the lack of panel data with which to investigate the question of income mobility. The model fits the data well in the sense that the predicted and actual distributions for period *t+1* are close, applying the estimated transition matrix to the actual distribution in period *t*. According to our estimates, mobility is a positive but non-linear function of age, highest in the younger cohorts and falling quite sharply after age 30. Quantitatively, the most significant finding is the large degree of upward mobility in Colombia, particularly at the bottom of the income distribution and for the younger age cohorts. Half of the twenty year olds in the bottom income group and two thirds of those in income class two move up by one class over three years. This upward mobility was not due to improvement in the overall distribution. Rather it was because the low income labor force is, to a surprising extent, comprised of a constantly revolving group of young workers rather than a fixed group of the low skilled workers trapped in bad jobs. Using data from 1988, we show that 47% of the workers in the bottom two income classes were aged 20-29. Just about half of them move up and out of the bottom group over the following six years, to be replaced by groups of new entrants. This upward mobility needs to be taken into account in any interpretation of point estimates of the distribution of income.

References

Adelman, I., S. Morley, C. Schenzler, and M. Warning. 1994. Estimating income mobility from census data. *Journal of Policy Modeling,* 16(2): 187-221.

Erickson, G.J., ed. *Maximum entropy and Bayesian methods*. Kluwer Academic Publishers. forthcoming.

Goodman, L. 1959. Some alternatives to ecological correlation. *American Journal of Sociology*, 64, 610-625.

Golan, A. 1998. Maximum entropy, likelihood and uncertainty: A comparison. In G. J. Erickson, ed., *Maximum entropy and Bayesian methods*. Kluwer Academic Publishers. forthcoming.

Golan, A., G. Judge, and D. Miller. 1996. *Maximum entropy econometrics: Robust estimation with limited data*. New York: John Wiley & Sons.

Golan, A. and S. J. Vogel. 1997. Estimation of stationary and non-stationary social accounting matrix coefficients with structural and supply-side information. Draft.

Kapur, J. N. and H. K. Kesavan. 1992. *Entropy optimization principles with applications*. San Diego: Academic Press.

Jaynes, E. T. 1957. Information theory and statistical mechanics. *Physics Review*, 106: 620-630.

Jessop, A. 1995. *Informed assessments: An introduction to information, entropy and statistics*. New York: Ellis Horwood.

Morley, S. A. 1981. The effect of changes in the population on several measures of income distribution. *American Economic Review,* 71(3): 285-294.

Shannon, C. E. 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379-423.

Telser, Lester. 1963. Least squares estimates of transition probabilities. In: *Measurement in economics: Essays in honor of Yehuda Grunfeld* (Stanford Press).

Wonnacott, T.H., and R. J. Wonnacott. 1977. *Introductory statistics for business and economics,* 2nd ed. New York: John Wiley & Sons.