Working Paper Series

WORKING PAPER NO. 534

MEASURING INCOME MOBILITY WITH CENSUS DATA

by

Irma Adelman, Samuel Morley, Christoph Schenzler and Matthew Warnin

University of California
Vanderbilt University

DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS

BERKELEY

CALIFORNIA AGRICULTURAL EXPERIMENT STATION

University of California

DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS
DIVISION OF AGRICULTURE AND NATURAL RESOURCES
UNIVERSITY OF CALIFORNIA

WORKING PAPER NO. 534

MEASURING INCOME MOBILITY WITH CENSUS DATA

by

Irma Adelman, Samuel Morley, Christoph Schenzler and Matthew Warning

University of California
Vanderbilt University

California Agricultural Experiment Station
Giannini Foundation of Agricultural Economics
Berkeley, California
April, 1990

MEASURING INCOME MOBILITY WITH CENSUS DATA

BY

IRMA ADELMAN, SAMUEL MORLEY, CHRISTOPH SCHENZLER AND MATTHEW WARNING*

University of California
Vanderbilt University

Do not quote without permission

---

*We are indebted to George Judge for his helpful comments.

In recent years an increasing amount of attention has been focused on what happens to the distribution of income and to the incomes of key groups such as the poor during the development process. Our knowledge about the links between growth and distribution are based on comparisons of snapshots of income distributions at different points in time.[1]

There are two problems with the snapshot approach to the distribution of income. First, since in actual economies, the number of individuals in the population is growing, everyone's relative position in the income pyramid will be affected by where the new entrants come into the labor force. Thus one cannot track incomes over time for groups such as the rich or the poor by comparing their average incomes at different points in time because one is not dealing with the same group of individuals (Morley 1981). The second problem is that there may be upward mobility for some individuals and downward mobility for others, so that the average earnings over time of the group as a whole may be completely unrepresentative of what happened to the individuals in the group over that time period.

These observational problems stem from the fact that we do not have panel surveys which would enable us to track the incomes of particular individuals over time. In this paper we will describe a method by which such inferences can be made using successive cross section household censuses. The method permits us to estimate mobility over time between different parts of the income pyramid, as well as to determine the influence of economic growth on upward mobility. Obviously the advantage of the method we are suggesting is that panel data are rare in LDC's and very costly to produce whereas censuses are now available for a number of years and a number of countries. In parts one and two we describe the method and the data problems we encountered in using it, in part three we apply the model to Brazil and in part four we draw some conclusions from the exercise.

PART I. THE ESTIMATION PROBLEM

In thinking about the distribution of income or earnings, one must, at the outset, distinguish between two different notions--one is the distribution of income across all the occupations in the society and the other is the distribution across the individuals in the society. Obviously at any point in time those two measures are the same. But they are not the same over time because individuals change jobs. Thus the structure of earnings could well remain constant over time while the earnings of individuals did not. That is particularly true in an economy where the population is growing. This is not to deny the importance of the snapshot approach, but rather to complement it with a measure which comes closer to telling what happened to individuals in the labor force over time because we think that is a key element in the evaluation of income inequality.

One can think of the distribution of income at any point in time as the result of a first order Markov process in which the probability that any individual will be in income class j at time t+1 will depend on which income class he or she was in at time t. Formally our interest in what happens to the income of particular groups over time could then be solved by estimating the transition matrix of the Markov process.[2] What we have from the reported censuses are the row and column sums of transition matrices, whose ijth elements are the number of people in income class j at time t and income class i at time t+1. We are looking for some way of estimating these cell entries, given our observation of the row and column sums. Since there are only 2n data points and n^2 unknowns, we need some additional data or restrictions to make progress.

Telser (1963) addressed this problem in the context of market shares for cigarettes using a time series approach. If one takes a sufficient number of observations of the distribution (in his case of cigarette smokers across brands) and if one assumes these distributions are generated by a first order Markov process, Telser showed how to derive an unbiased regression estimator of the unknown elements of the transition matrix. The method gives the transition matrix which minimizes the difference between the actual distribution at time t+1 and the distribution predicted by applying the transition matrix to the distribution at time t. Lee, Judge and Zellner (1970) propose alternative Bayesian and non-Bayesian approaches to the estimation of transition probabilities from time series data on marginal totals and examine the properties of these estimates.

Unfortunately the time series approach is not practical for the income distribution problem in LDCs because we do not have a sufficient number of censuses. But we can use regional data from the censuses themselves as an alterative. If we have regional data, and can assume either that the same first order Markov mechanism operates in each region, or that it differs across regions in a predictable way, we can proceed, as Telser did, to use regression analysis to find the transition matrix which minimizes the difference between the observed and the predicted regional distributions at time t+1, given the observed distribution at time t.

A similar problem has been addressed in sociology and political science. In 1953 Goodman proposed a simple regression to estimate the interior elements in a four way table of individual characteristics when only the regional row and column sums of the two characteristics are known. His technique made the assumption that the interior conditional probabilities were constant across regions. Crewe and Payne (1976) applied the same general technique to get an estimate of the percentage of different occupational groups voting for the two British political parties. They extended Goodman's technique by assuming that the conditional probabilities were a function of exogenous factors that vary across regions. They derived a best linear unbiased estimator which simultaneously produced an estimate of the transition matrix and of the effect of the exogenous variables on that transition matrix. Their model was applied to a two by two case--two parties and two broad occupational classes. Our model is a simple extension of Crewe and Payne to the n-dimension case, where the n dimensions are income classes and where we are trying to find the proportion of those in income class j in time t who move to class i at time t+1.

Let P be an nxn transition matrix whose ijth element, P_{ij}, is the proportion of those in income class j at time t who move to class i at time t+1. Let X_i and Y_i be the observed fraction of the total population in income class i at time t and t+1 respectively. N is the number of mutually exclusive income classes. By definition, in matrix notation,

$$(1) \qquad Y = P*X$$

$$\text{or } Y_i = \sum_j P_{ij}X_j \qquad (i = 1,\ldots,n)$$

Equation one looks like a regression model where we observe the X's and the Y's and estimate the unknown transition parameters P_{ij}. Clearly only n-1 of these equations are independent. However, rather than dropping one of the equations, we make the equivalent restriction that the sum of each column of P_{ij}'s be equal to one. We further require that each estimated P_{ij} falls between zero and one. The problem with equation (1) is that we do not have enough data to estimate the P_{ij}. In our case we have 5 income classes so we are trying to estimate 25 elements of the transition matrix, but we have only five observations of the marginal totals X_i and Y_i.

We proceed by using regional observations. If the Markov process could be assumed to be the same across regions, we could increase the number of observations by taking regional observed values of the distribution. However it is probably unreasonable to assume that mobility is the same across regions. Instead, one would expect it to vary positively with many variables like income growth and labor force structure that vary across regions. Surely one's chances of moving up the distribution ladder are higher in fast growing or highly industrialized regions. Following Crewe and Payne (1976) it is straightforward to modify equation (1) to take account of regional variations in the transition matrix.

We hypothesize that transition probabilities are a function of observable characteristics Z that differ across regions. Thus in the simplest form with only one Z variable:

$$(2) \qquad P_{ij} = a_{ij} + b_{ij}Z$$

In (2) Z is a variable with region-specific values. In our case, the growth rate of income was used. More complex formulations, in which the P_{ij} depend on more variables, are possible, but were precluded in our estimation by the small number of degrees of freedom we had.

If we now substitute equation (2) into equation (1) we get:

$$(3) \qquad Y_i = \sum_j (a_{ij} + b_{ij}Z)X_j \quad (i = 1, \ldots, n)$$

This is the equation system we will estimate under the two restrictions:

$$(4) \qquad 0 \leq P_{ij} \leq 1 \quad \text{for all } i, j$$

$$(5) \qquad \sum_i P_{ij} = 1 \text{ for } j = 1, \ldots, n$$

Unfortunately, available statistical packages cannot incorporate both restrictions. Packages which allow for estimation of systems of equations will incorporate the cross-equation constraint (5) but not the within-equation inequality constraint (4). Bayesian packages, which can incorporate the inequality constraint, do not allow for estimation of systems of equations and thus prohibit incorporation of cross-equation constraints.

To circumvent this problem, one can estimate a system of equations explicity incorporating restriction (5), and perform a non-linear transformation on the coefficients (transition probabilities) that restricts their values to between 0 and 1, thus incorporating restriction (4). For estimation without a Z variable, such a transformation could take the form:

$$Y_i = \sum_j e^{-a_{ij}^2} X_j$$

Here $P_{ij} = e^{-a_{ij}^2}$ and thus must fall between 0 and 1 for all values of a_{ij}. This method is relatively straightforward for the simple case, where a Z variable is excluded, but proves intractable with the inclusion of such a variable.

The alternative we used was treating the problem as a non-linear programming problem representing the ordinary least squares approach. The objective function minimized is the sum of squared errors and the constraints are given by equations (6), (7) and (8) below. This yields a non-linear programming problem with non-linear inequality constraints.

A representative equation of the constraint set is given by:

$$(6) \qquad Y_i^r = \sum (a_{ij} + b_{ij} Z^r) X_j^r + \epsilon^r$$

where the r superscript indicates regional observations and ϵ^r is the statistical error term.

In our estimation, we required that the inequality constraint hold for all values of Z in the sample and that the cross-equation constraint hold for the mean value of Z in the sample. That is:

$$(7) \qquad 0 \leq P_{ij}^r \leq 1 \qquad \text{for all i, j and r}$$

$$(8) \qquad \sum_i \overline{P_{ij}^r} = 1 \qquad \text{for j} = 1,\ldots,n$$

Where $P^r_{ij} = a_{ij} + b_{ij} Z^r$ and $\overline{\overline{P^r_{ij}}}$ is the sample mean of P^r_{ij}.

The above procedure yields unbiased estimates of the parameters under the usual assumptions that the distribution of the error term is iid.

To estimate standard errors of the estimated coefficient we used the "delete-one jackknife technique" (i.e., subsample size = sample size - 1) by subsampling with replacement from the data and estimating the coefficients for each subsample. Such a technique was required because standard calculations do not incorporate the information contained in the restrictions and so give misleading estimates of the standard errors. Following Efron (1982) the jackknife estimate of the standard error of a paramenter estimate $\hat{\theta}_{ij}$ is calculated according to:

$$s.e._{\hat{\theta}_{ij}} = \sqrt{\frac{n-1}{n} \sum_{k=1}^{n} (\hat{\theta}^k_{ij} - \bar{\hat{\theta}}_{ij})^2}$$

Where $\hat{\theta}^k_{ij}$ is the kth subsample estimate of the parameter, and $\bar{\hat{\theta}}_{ij}$ is the sample mean of the subsample estimates of the parameter.

PART II. GENERATING THE NECESSARY DATA

Central to the procedure we are using here are sets of regional observations of the distribution of income by age and sex. We have chosen Brazil for this purpose, partly because this regional data is available for 1970 and 1980, and partly because of the interest inherent in Brazil's experience as a country which gained notoriety for its rapid but inequitable growth path over the period 1960-80. Our estimate will permit us to make what we think is a crucial distinction between what happened to the distribution of income and what happened to the income of those who were in the distribution at a point in time. We hope that by describing in some detail both the econo-metric procedure and how the necessary data can be obtained from census tapes, we will encourage others to apply the same technique in other countries.

The goal, in the data preparation about to be described, is to obtain an estimate of the income distribution in 1980 of those in the distribution who survived from 1970--a population we label "survivors". That means that we have to remove new entrants from the observed 1980 group in those cohorts where the 1980 labor force is larger than the 1970; and we have to remove those who retire from the 1970 group in those cohorts where the 1970 labor force is larger than in 1980. The problem is that we have no way of knowing which of the 1980 workers are new entrants in expanding cohorts, or which will retire in those age groups which shrank during the 1970's. We can, however, determine many of the characteristics of new entrants and retirees. For each age cohort we first disaggregate by sex education and region with each combination defining a cell. We then subtract the 1970 from the 1980 totals in each cell. If the difference is positive we know there were new entrants between 1970 and 1980, with the particular characteristics of the cell. We assumed that the new entrants had the same distribution of income as the total observed for that cell in 1980. That permits us to estimate the income distribution of survivors in each cell by subtracting, element by element, the vector of new entrants from the 1980 cell totals.

If we then aggregate across the 120 cells (2 sex, 5 for education and 12 regions) in each age cohort, we obtain an estimate of the 1980 income distribution of survivors. It is a vector giving the observed value of the X_i's which we will compare with the Y_i's obtained from the observed 1970 distribution.

For older age cohorts which had retirements instead of new entrants over the 1970's, we use a procedure similar to that described above to estimate the 1970 income distribution of those who would retire during the next decade. In any cell if the 1980 number is smaller than the 1970 number we know there were net retirements. Here we assumed that the retirees had the same income profile in 1970 as the rest of the members of the cell, and we subtract element by element the vector of retirees from the total distribution of the population in the cell in 1970. We then aggregate across the ten cells as before to get a survivors' distribution for 1970. This distribution gives the vector of Y_i's which we will use in our regressions along with the X_i's obtained from the observed 1980 distribution.

The other complication encountered in adapting census data for income-mobility estimates is regional migration. Clearly, in a country like Brazil, there is a substantial amount of interregional migration. Since we assume that mobility differs across region we have the choice of either excluding migrants, placing them in their destination populations or in their originating regions. We chose the last of the three options because it allows us to use the observed regional distributions without making a correction for migrants similar to the one made for retirements. The disadvantage with our procedure is that for regions with substantial outmigration the income growth that we use is not equal to that of the originating region since some part comes from migrants to faster growing areas. An interesting question which we can fairly easily explore in an extension to this work is the effect of migration on mobility. How much do those who migrate contribute to the observed mobility patterns? Did migrants do better than those they left behind? We can get a good answer to both of these questions by either comparing the transition matrices of migrants and non-migrants or by putting migrants into the destination population.

The Brazilian public use census tapes upon which this work is based, are a 1% sample of the demographic censuses of 1970 and 1980. They contain data on earnings, age, sex, occupation, education, current and previous residence, time in present residence, and many other variables. We aggregated the data into the twelve regions shown in the appendix. We treated as migrants all those who had resided in their current region for less than ten years. We then reassigned migrants to the region they reported as their previous residence.

We divided the 1970 population into six age groups: 15-19, 20-29, 30-39, 40-49, 50-59 and over. We created five education groups, classifying individuals according to the last grade passed. They are: no education: elementary (5 years or less): middle school (6-8 years): high school (9-12 years); and university.

Income is reported in current cruzeiros. We converted the 1970 data to 1980 cruzeiros using the Rio de Janeiro cost of living index. We then created the following five real income classes:

No income, 0-3599, 3600-4999, 5000-11999, > 12000

We set the upper limit of class one at 3599 CR$ because the 1970 minimum wage was 3600 measured in 1980 cruzeiros. Over the subsequent decade, the minimum wage rose in real terms reaching 4149 CR$ or $79 in 1980. That means, of course, that any worker holding a less than minimum wage job in 1970 would move from class one to class two by finding a minimum wage paying job in 1980. Note that the income variable that we used ostensibly includes earned income from all sources, but there is a substantial degree of underreporting, particularly of income from capital.

PART III. ESTIMATES OF MOBILITY IN BRAZIL

Before the oil shocks, Brazil was often held up as the quintessential example of inequitable growth. Between 1960 and 1980 it enjoyed one of the worlds highest growth rates with per capita income rising by 3.9% per year. But the benefits of this prodigious boom do not appear to have been distributed at all equally across the working population. The Gini coefficient rose from .50 to .59 and the average income of the top 20% grew half again as fast as that of the bottom 60%. During the 1970's, the period we will concentrate on here, the income share of the bottom 60% shrank from 21.2% to 19.7% while the top 20% rose from 61.7% to 63.3%.[3] There is a long literature suggesting reasons for this unfortunate pattern. (See Bacha and Taylor (1978), Fishlow, (1972), Langoni (1973), Morley and Williamson (1975), Morley (1982), Fields (1977), Pferrerman and Webb (1979), Denslow and Tyler (1983), Hoffman and Kageyama (1986)).

We do not wish to add to this literature here. Instead we wish to look behind the aggregate numbers to get a measure of mobility during the 1970's. Granted that the average income of the worst jobs rose more slowly than that of the best. Granted that wage differentials widened. What happened to the average incomes of those who held these worst jobs in 1970 over the ensuing decade? How did they fare relative to those who were further up the income pyramid in 1970? These are some of the main questions we will attempt to answer here.

One cannot answer these questions with published data because it does not distinguish survivors from new entrants. Since new entrants tend to occupy lower paying jobs, their presence biases downward any comparison based on all respondents. To see this we have displayed three separate distributions in table one: that of the entire observed labor force over 15 years of age, that of survivors, defined as those who were present in both 1970 and 1980 and that of new entrants.

TABLE 1 about here

It is obvious that by any measure there was a lot of upward mobility in Brazil during the 1970's. 44% of the male labor force earned less that the 1970 minimum wage of 3600 cruzeiros. Ten years later only 25% earned that little. But the full extent of income growth for those at the bottom of the income pyramid is hidden by the presence of new entrants, 31% of whom earned less than 3600 Cr$ in 1980. When we look just at male survivors we find that the 46% of the labor force earning less than the minimum wage dropped to 19% ten years later. The difference between the survivors and the new entrants distribution is not so great for females, but is still significant. For survivors as a group, income grew by 9% per year over the decade. That compares to reported growth of 4% for the labor force as a whole. Clearly, while Brazilian style growth led to a widening of income differentials, it generated a substantial amount of upward mobility as well.

We now look more closely at the detailed evidence on mobility by using our model to estimate transition matrices for different sex and age groups. We estimated separately for males and females and for four age groups: 15-19, 20-39, 40-59 and 60 years and older. Each of the middle groups represents aggregation of two of the previously defined age groups. This aggregation increases the degrees of freedom in the estimation process. The aggregation is limited to these groups in order to be consistent with a priori assumptions on mobility patterns.

Table two shows the constrained equation estimates and Table three the transition matrices derived from the coefficients in Table two. To derive Table three we set the regional growth variable at its average level for the relevant age group. Thus, for example, to get the entry (P_{01}) for males in age group one take the coefficient on X1 and subtract from it .008 times the average growth rate of income for this age group (19.73%).

$$(P_{01} = .192 - .008\ (19.73) = .033)$$

Consider now the mobility patterns implied by Table three. For males first, there is very little downward mobility, except for the oldest age group. The upper off diagonal of all the matrices is either zero or a small number.** If you were lucky enough to be in the top income group in 1970 the chances were better than 85% that you would stay there. If you were in X_3 and were less than 40 in 1970 (age group 0-2), the chances were better than 95% that you would either stay where you were or move up to the top group. (84% of the 20-39 year group moved up).

**The only exception is males 40 to 59 moving from class X_3 to X_2.

What about those at the bottom of the distribution in 1970? They also shared in the favorable mobility patterns. If you had zero income as a male teenager in 1970, you had a 93% chance of moving up at least one class; and a 49% chance of moving up at least two classes--implying a move up to a job earning more than the minimum wage. If you were in the 20-39 group and earning less than the minimum wage in 1970 (in X_1), you had a 44% chance of moving up at least one income class. Note here that the minimum wage itself increased in real terms from 3600 to 4149 CR$ so part of this mobility is an expansion in the number of jobs covered by minimum wage legislation. Consistent with this interpretation, it appears that upward mobility was greater for those who started further up the labor pyramid. For the 20-39 age group, compare the very high probabilities that those who started in X_2 would move up to Y_3 or Y_4 or that those who started in X_3 would move to Y_4 with the much less favorable prospects for those starting in X_1. The growth process appears to have helped everyone. But it was most favorable for those placed high enough in the income distribution to take advantage of the rapid expansion in jobs with relatively high educational requirements and wages.

The reader may object that this differential pattern is not found in the X_0 column where there is also large upward mobility. But X_0 represents those earning zero income and is a somewhat special case. Most of this group are teenagers many of whom undboubtedly worked on farms or in family business while attending school and who entered the formal labor market during the decade. When they did, many were able to get good jobs. Note also that this group represents less than 3% of the non-teen age labor force.

The picture for female workers is a good deal less favorable than it is for males. There is far more downward mobility and less upward mobility. Whereas about 89% of males who started in X_2 or X_3 in the 20-39 age group moved up at least one income class, only about 62% of similarly placed women did. In the same age group 44% of X1 males moved up, compared to only 24% of the females. For males the zero income class appears to be transitory, with relatively few remaining over the decade. Except for women over 60 the situation is entirely different for females, where the probability is quite high that, if one started in a zero income job, one would still be there ten years later. Thus the data tell us that when measuring mobility one must distinguish by sex.

Consider next the age income profiles underlying the transition matrices. In Table four we show the average annual real income growth rates of different age cohorts of survivors. For this purpose we use a more disaggregated breakdown than the one used for our regression analysis. Table four makes clear the very steep income gradient during the early working years. Over the 1970's decade young workers gained relative to other survivors. Since those young workers tended to start at the bottom of the income pyramid[4], much of the upward mobility we have documented in Tables 2-3 must, in fact, have been young workers moving up and out of their low paying entry level jobs, which were taken by the next generation of new entrants.

<u>TABLE 4</u> about here

In trying to interpret and understand what this mobility evidence implies, it is important to go back to the social significance of inequality. We have argued that the snapshot approach gives a misleading picture because it does not allow us to track individuals over time. The mobility data confirm that the published aggregate data hides a substantial amount of upward mobility for those who were at the bottom of the 1970 income pyramid. That is comforting to the defenders of trickle down growth. However, if following Paglin [1975], one argues that it is cohort inequality that matters--that is, one's prospects relative to the other members of the same cohort--then Brazil's growth process does not look so appealing. For, as we pointed out, those who started out in the better positions in 1970, also tended to fare better over the ensuing decade. Thus growth seems to have widened the divergences within each age cohort, particularly among male workers, even while it made possible upward mobility for almost everyone.

Consider now the role of the growth variable itself in the regressions. Recall that we hypothesized that the transition probabilities should be related to the overall growth rate of regional income. Upward mobility should be higher and downward mobility lower in the faster growing regions. We should therefore expect the coefficients on the growth variable to be negative in the upper off diagonal of Table 2, where we are estimating downward mobility, and positive in the lower off diagonal. Table five shows the sign and significance of the growth coefficients in the Table two regressions. We included diagonal elements (P_{ii}) together with the downward mobility group and performed a one-tailed t-test at the 95% confidence level to evaluate the significance of the coefficients.[5]

TABLE 5 about here

As the reader can see, our hypothesis of a positive growth effect on mobility is confirmed by the data. The bulk of the downward mobility coefficients are negative and the upward positive, and all but one of the seven significant coefficients have the right sign. Another way to see the growth effect on mobility is to compare the P_{ij}'s in Table three with the coefficients on the X_j's in Table two. The former are calculated at the average growth rate of the relevant age class, while the latter would be the estimated transition probability if income growth were zero. Thus, for example, we see that for the 20-39 age males, if you were in X_3 in 1970, with zero growth you had a 19% chance of falling to Y_2, a 30% chance of staying where you started and a 45% chance of moving up to the top class. Rapid growth dramatically changes those probabilities. At the average growth of 9.9%, the probability of falling to X_2 falls to 5% while the probability of moving to the top class rises from 45% to 84%. One can see the same large growth impact in other cells of the matrix.

One should however enter a cautionary note here. While the growth effect appears to be large, it does not appear to be very significant in a statistical sense. As the reader can see from Table five only 6 of the 73 coefficients are significant and have the right sign. Since the overall fit of these sets of regressions is quite good, that has to mean that the initial position (the X's) must play a big explanatory role. Indeed, if one looks at Table two more closely, that appears to be true particularly along the diagonal. X_1 is a powerful predictor of Y_1, in both the male and female 20-59 age groups. The same is true for

X_3 and Y_3 or X_4 and Y_4 in the male regressions and for X_2 and X_3 in the 40-59 group of females and for the X_4 and Y_4 in the 20-39 females.

So much for the implications of our estimates. How good are the estimates? How well does our procedure work? There are several ways to address that question. Judging by overall goodness of fit, the estimates, particularly for non-teenagers look very good, explaining more than 90% of the variance in the regional 1980 distributions given the 1970 distribution. When one looks at the individual equations it is clear that one of the reasons is the very high significance level of X_1 in the Y_1 regression in most sex and age classes. The same can be said of X_4 in the Y_4 regression for 40-59 year old males and 20-39 aged females. Overall, 24 of the 144 coefficients in Table 2 for males and 24 of the 156 for females are significant at the 5% level.

As an alternative test of goodness of fit, we applied our estimated transition matrix to the all-Brazil data for each age-sex group.[6] To do this we first calculated weighted averages across states of (1) the per capita GNP growth rate, (2) shares of the population in the given income classes in 1980 (Y_j's) and (3) shares of the population in the income classes in 1970 (X_j's). For each age-sex group the weights used were the number of workers in each state corresponding to the age-sex group. We then calculated forecast values for each group by applying the growth-rate-dependent transition probabilities (using the weighted growth rate for the group) to the weighted X_j's to produce predicted values for the weighted Y_j's. These results are presented in Table six along with the actual weighted values.

TABLE 6 about here

In order to quantitatively evaluate the fit of our estimates to the actual values, we regressed the actual all-Brazil weighted population shares for each income class on the predicted all-Brazil population shares in a linear regression with no constant. Our observations are the different age-sex groups in each income class. We ran these regressions as an iterated SUR system.

Our system is:

$$Y_0^w = \beta_0 \hat{Y}_0^w + \epsilon_0$$

$$Y_1^w = \beta_1 \hat{Y}_1^w + \epsilon_1$$

$$Y_2^w = \beta_2 \hat{Y}_2^w + \epsilon_2$$

$$Y_3^w = \beta_3 \hat{Y}_3^w + \epsilon 3$$

$$Y_4^w = \beta_4 \hat{Y}_4^w + \epsilon_4$$

Where the Y_j^w's are the actual weighted population shares for income class j and the $\hat{Y}_j^w$'s are the predicted population shares calculated from our estimated transition probabilities.

If our estimates are "good" we should see unitary coefficients on the predicted population shares (unbiasedness) and high R^2 values. The results presented in Table seven confirm both expectations. Prediction errors are small and all coefficients are very close to one.

TABLE 7

RESULTS OF MODEL TEST

	β_0	β_1	β_2	β_3	β_4
Coefficients	.9930	.9144	1.0188	1.0553	1.0387
Standard Error	.0296	.0125	.0209	.0117	.0086

We tested the collective proximity of our estimated coefficients to unity through a likelihood ratio test based on comparing the value of the likelihood function for the unrestricted estimation with the value of the likelihood function when all the β's were restricted to equal one. This statistic is distributed as a chi-square with five degrees of freedom (five linear restrictions: $\beta_i = 1$, $i = 0,\ldots 4$). From our results, we cannot reject the null hypothesis that all the β's are equal to one. That is, our estimates of the Y_j's for all-Brazil are unbiased. Our estimates also have very high R^2 and very low mean square forecast error. We thus do remarkably well.

CONCLUSION

This paper develops a method for estimating Markov income-mobility matrices using census data. It allows for the first time to use non-panel data to make estimates of intracohort income mobility over time. The method uses regional observations as the basis for a non-linear programming approximation to a SUR regression with additional interval constraints on the estimated parameters.

The method was successfully applied to Brazil where it generates a very accurate and unbiased estimate to the predicted 1980 distribution of the labor force across income classes, given the observed 1970 distribution. Our results confirm previous assertions that the observed distribution, which became more unequal over the 1970s, also hides a great deal of upward mobility. We show that the probability of upward mobility for survivors was high in every age cohort, and was also positively related to regional income growth. All income classes benefitted from growth, but it also appears that those who started further up the distribution in 1970 tended to fare better over the ensuing decade. Thus while growth was beneficial to all groups, it also widened income divergences within each age cohort.

TABLE 1

DISTRIBUTION OF LABOR FORCE,
SURVIVORS AND NEW ENTRANTS
BY INCOME CLASS 1970-1980
(in percent)

Income Class	All over 15		Survivors		New Entrants
	1970	1980	1970	1980	1980
MALES					
0	6.7%	5.3%	7.6%	1.7%	10.6%
0-3599	37.6	18.4	37.2	17.0	20.5
3600-4999	23.1	13.6	22.8	12.2	15.6
5000-12000	18.9	34.9	18.6	36.8	32.0
>12000	13.7	27.8	13.8	32.3	21.1
FEMALES					
0	9.0%	7.9%	9.0%	6.8%	8.8%
0-3599	50.4	32.7	49.7	31.6	33.6
3600-4999	18.0	15.3	18.4	14.8	15.8
5000-12000	14.9	28.9	15.4	30.4	27.7
>12000	7.7	15.1	7.5	16.4	14.1

Source: Census Tapes

Table 2. Regression results, Brazil 1970-1980.

Males Ages 15-19 R-squared=.633

		Income Category In 1970									
		x 0		x 1		x 2		x 3		x 4	
		Base Coef	Growth Coef	Base Coef	Growth Coef	Base Coef	Growth Coef	Base Coef	Growth Coef	Base Coef	Growth Coef
Income Category In 1980	y 0	0.084	-0.00061	0.189	-0.008	0	0	0	0	0	0.000
		0.425	*0.017*	*0.245*	*0.010*	*0*	*0*	*0*	*0*	*0*	*0.000*
	y 1	3.137	-0.138	-0.886	0.058	0	0	0	0	0	0.000
		0.743	*0.033*	*0.471*	*0.023*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
	y 2	-0.193	0.013	0.250	0.000	0	0.000	0.060	-0.003	0.261	-0.012
		0.113	*0.007*	*0.375*	*0.020*	*0.000*	*0.000*	*0.557*	*0.030*	*2.876*	*0.126*
	y 3	-1.306	0.085	0.491	-0.008	1.051	-0.025	2.689	-0.118	0	0.000
		0.899	*0.046*	*0.521*	*0.027*	*1.691*	*0.093*	*2.929*	*0.129*	*0.000*	*0.000*
	y 4	-0.362	0.023	-0.166	0.016	0.761	-0.016	2.585	-0.102	1.130	-0.008
		0.657	*0.040*	*0.421*	*0.023*	*2.749*	*0.143*	*2.066*	*0.129*	*1.426*	*0.093*

Males Ages 20-39 R-squared=.918

		Income Category In 1970									
		x 0		x 1		x 2		x 3		x 4	
		Base Coef	Growth Coef	Base Coef	Growth Coef	Base Coef	Growth Coef	Base Coef	Growth Coef	Base Coef	Growth Coef
Income Category In 1980	y 0	0.587	-0.042	0.074	-0.004	0	0	0	0	0	0.000
		0.657	*0.053*	*0.048*	*0.004*	*0*	*0*	*0*	*0*	*0*	*0.000*
	y 1	0.082	-0.006	0.797	-0.027	0	0	0	0	0	0.000
		0.739	*0.480*	*0.139*	*0.014*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
	y 2	-0.111	0.021	0.221	-0.002	-0.095	0.015	0.193	-0.014	0.150	-0.011
		0.705	*0.062*	*0.110*	*0.010*	*0.259*	*0.029*	*0.326*	*0.029*	*0.350*	*0.034*
	y 3	0.185	0.012	0.042	0.018	0.845	0.009	[illegible]	-0.028	0.317	-0.023
		1.189	*0.115*	*0.192*	*0.019*	*0.566*	*0.053*	[illegible]	*0.058*	*0.518*	*0.038*
	y 4	-0.428	0.083	-0.007	0.003	-0.015	0.002	0.445	0.039	0.848	0.002
		0.734	*0.067*	*0.072*	*0.010*	*0.278*	*0.043*	*0.613*	*0.077*	*0.451*	*0.043*

Table 2 (continued)

Males Ages 40-59 R-squared=.942

Income Category in 1980		x 0 Base Coef	x 0 Growth Coef	x 1 Base Coef	x 1 Growth Coef	x 2 Base Coef	x 2 Growth Coef	x 3 Base Coef	x 3 Growth Coef	x 4 Base Coef	x 4 Growth Coef
	y 0	0.525	-0.062	0.059	-0.003	0	0	0	0	0	0.000
		1.012	*0.120*	*0.038*	*0.005*	*0.0014867*	*0.0001726*	*0.000*	*0.000*	*0.000*	*0.000*
	y 1	0.037	-0.004	0.831	-0.034	0	0	0	0	0	0.000
		2.240	*0.264*	*0.072*	*0.014*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
	y 2	0	0.000	0.155	0.002	0.265	-0.017	0.023	0.020	-0.036	0.011
		0.000	*0.000*	*0.077*	*0.014*	*0.355*	*0.062*	*0.456*	*0.086*	*0.091*	*0.029*
	y 3	0.012	-0.001	0.097	0.007	0.227	0.091	1.298	-0.135	0	0.000
		2.134	*0.249*	*0.086*	*0.014*	*0.259*	*0.029*	*0.317*	*0.048*	*0.000*	*0.000*
	y 4	0.425	0.068	-0.022	0.007	-0.080	0.025	0.112	0.041	1.035	-0.011
		2.379	*0.278*	*0.019*	*0.005*	*0.110*	*0.034*	*0.225*	*0.058*	*0.101*	*0.029*

Columns: Income Category in 1970

Males Ages 60 and over R-squared =.898

Income Category in 1980		x 0 Base Coef	x 0 Growth Coef	x 1 Base Coef	x 1 Growth Coef	x 2 Base Coef	x 2 Growth Coef	x 3 Base Coef	x 3 Growth Coef	x 4 Base Coef	x 4 Growth Coef
	y 0	0.000	0.000	0.046	-0.001	0.000	0	0	0.000	-0.010	0.004
		0.000	*0.000*	*0.033*	*0.003*	*0.017*	*0.0010945*	*0*	*0.000*	*0.050*	*0.007*
	y 1	-0.004	0.002	0.647	-0.016	0	0.000	0	0.000	0	0.000
		0.166	*0.073*	*0.080*	*0.010*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
	y 2	-0.183	0.083	0.197	0.005	-0.001	0.001	0.026	0.005	0.561	-0.039
		0.262	*0.106*	*0.076*	*0.013*	*0.056*	*0.027*	*0.232*	*0.027*	*0.206*	*0.013*
	y 3	1.161	-0.081	0.111	0.010	1.052	-0.024	0.716	-0.050	0	0.000
		1.416	*0.176*	*0.129*	*0.020*	*0.126*	*0.036*	*0.355*	*0.030*	*0.000*	*0.000*
	y 4	0	0.000	-0.008	0.004	-0.051	0.023	0.438	0.022	0.380	0.043
		0.000	*0.000*	*0.013*	*0.007*	*0.289*	*0.040*	*0.786*	*0.070*	*0.441*	*0.043*

Columns: Income Category in 1970

Table 2 (continued)

Females 15-19 R-squared=.630

Income Category In 1970

Income Category In 1980		x 0 Base Coef	x 0 Growth Coef	x 1 Base Coef	x 1 Growth Coef	x 2 Base Coef	x 2 Growth Coef	x 3 Base Coef	x 3 Growth Coef	x 4 Base Coef	x 4 Growth Coef
	y 0	-0.282	0.035	0.172	-0.008	0	0	0	0	0	0.000
		0.919	*0.050*	*0.116*	*0.007*	*0*	*0*	*0*	*0*	*0*	*0.000*
	y 1	2.386	-0.113	0.096	0.024	0	0	0	0	0	0.000
		1.552	*0.080*	*0.537*	*0.033*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
	y 2	-0.382	0.031	0.246	-0.006	0.571	-0.027	0.371	-0.018	2.434	-0.115
		0.415	*0.033*	*0.537*	*0.030*	*1.144*	*0.070*	*1.728*	*0.086*	*2.335*	*0.116*
	y 3	-0.190	0.015	0.061	0.010	1.025	-0.009	2.434	-0.115	0.000	0.000
		1.270	*0.070*	*0.242*	*0.013*	*1.008*	*0.056*	*0.471*	*0.036*	*2.186*	*0.106*
	y 4	0.000	0.000	0.024	0.004	0.000	0.000	2.041	-0.097	2.412	-0.114
		0.106	*0.007*	*0.129*	*0.007*	*0.000*	*0.000*	*1.386*	*0.076*	*3.774*	*0.222*

Females 20-39 R-squared=.906

Income Category In 1970

Income Category In 1980		x 0 Base Coef	x 0 Growth Coef	x 1 Base Coef	x 1 Growth Coef	x 2 Base Coef	x 2 Growth Coef	x 3 Base Coef	x 3 Growth Coef	x 4 Base Coef	x 4 Growth Coef
	y 0	0.498	0.001	0.126	-0.009	0	0	0	0	0	0.000
		0.321	*0.038*	*0.043*	*0.005*	*0*	*0*	*0*	*0*	*0*	*0.000*
	y 1	-0.038	0.012	0.851	-0.022	0	0	0	0	0	0.000
		0.086	*0.029*	*0.072*	*0.010*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
	y 2	0.221	-0.006	0.034	0.008	-0.146	0.047	0.722	-0.068	0.098	-0.009
		0.259	*0.029*	*0.043*	*0.005*	*0.168*	*0.014*	*0.283*	*0.034*	*0.364*	*0.034*
	y 3	0.729	-0.063	-0.010	0.015	0.309	0.055	0.163	0.025	1.108	-0.105
		0.523	*0.067*	*0.072*	*0.010*	*0.710*	*0.086*	*1.084*	*0.139*	*0.652*	*0.067*
	y 4	0	0.000	0.018	0.005	-0.090	0.029	0.288	0.018	1.084	-0.069
		0.001	*0.000*	*0.038*	*0.005*	*0.101*	*0.029*	*0.297*	*0.058*	*0.350*	*0.062*

Table 2 (continued)

Females 40-59

R-squared=.873

Income Category in 1980		Income Category in 1970: x 0 Base Coef	x 0 Growth Coef	x 1 Base Coef	x 1 Growth Coef	x 2 Base Coef	x 2 Growth Coef	x 3 Base Coef	x 3 Growth Coef	x 4 Base Coef	x 4 Growth Coef
	y 0	0.345	0.035	0.130	-0.008	0.000	0	0	0	0	0.000
		0.623	*0.086*	*0.062*	*0.010*	*0.014*	*0.010*	*0.005*	*0.039*	*0.000*	*0.000*
	y 1	0.400	-0.036	0.752	0.002	0	0	0	0	0	0
		0.317	*0.038*	*0.067*	*0.010*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
	y 2	0.018	0.021	0.021	0.011	0.547	-0.035	0.016	-0.001	0.284	-0.025
		0.058	*0.019*	*0.048*	*0.010*	*0.245*	*0.043*	*0.182*	*0.014*	*0.278*	*0.029*
	y 3	0.016	0.020	0.009	0.011	0.451	-0.002	0.948	-0.063	0.283	-0.025
		0.197	*0.043*	*0.034*	*0.005*	*0.129*	*0.034*	*0.168*	*0.024*	*0.360*	*0.077*
	y 4	0.004	0.005	0.016	-0.001	0.262	-0.017	0.131	0.044	0.433	0.051
		0.058	*0.010*	*0.038*	*0.010*	*0.312*	*0.067*	*0.341*	*0.086*	*0.388*	*0.053*

Females 60 and over

R-squared=.463

Income Category in 1980		Income Category in 1970: x 0 Base Coef	x 0 Growth Coef	x 1 Base Coef	x 1 Growth Coef	x 2 Base Coef	x 2 Growth Coef	x 3 Base Coef	x 3 Growth Coef	x 4 Base Coef	x 4 Growth Coef
	y 0	0	0.000	0.111	0.000	0	0.000	0	0.000	0	0.000
		0.000	*0.000*	*0.047*	*0.006*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
	y 1	0.285	-0.026	0.614	0.016	0.063	0.028	0.000	0.000	0.000	0.000
		0.914	*0.104*	*0.225*	*0.041*	*0.259*	*0.092*	*0.000*	*0.000*	*0.000*	*0.000*
	y 2	0	0.000	0.119	0.002	0.000	0.000	0.798	-0.072	0.840	-0.072
		0.000	*0.000*	*0.057*	*0.006*	*0.386*	*0.035*	*0.515*	*0.063*	*0.199*	*0.070*
	y 3	0.057	0.026	0.063	-0.002	0.937	-0.028	0.847	-0.069	0.834	-0.075
		2.593	*0.073*	*0.171*	*0.028*	*0.275*	*0.092*	*0.316*	*0.070*	*1.078*	*0.123*
	y 4	0.416	0.053	0.025	-0.002	0.000	0.000	0.000	0.000	0.000	0.000
		0.718	*0.092*	*0.038*	*0.006*	*0.060*	*0.013*	*0.136*	*0.002*	*0.095*	*0.041*

Note: Figures in italics are jackknife estimates of standard errors.

Table 3. Estimated Transition Matrices, Brazil 1970-1980.

Males 15-19 R-squared=.633

Income Category in 1980		Income Category in 1970: xo	x 1	x 2	x 3	x 4
	y 0	0.072	0.031	0	0	0
		0.066	*0.033*	*0*	*0*	*0*
	y 1	0.438	0.246	0	0	0
		0.153	*0.076*	*0.000*	*0.000*	*0.000*
	y 2	0.051	0.244	0	0.000	0.000
		0.033	*0.017*	*0.000*	*0.090*	*0.385*
	y 3	0.343	0.342	0.559	0.438	0
		0.176	*0.080*	*0.305*	*0.458*	*0.000*
	y 4	0.095	0.137	0.441	0.562	1.000
		0.149	*0.053*	*0.305*	*0.484*	*0.385*

Males 20-39 R-squared=.918

Income Category in 1980		Income Category in 1970: xo	x 1	x 2	x 3	x 4
	y 0	0.169	0.035	0	0	0
		0.149	*0.010*	*0*	*0*	*0*
	y 1	0.023	0.525	0	0	0
		0.206	*0.034*	*0.000*	*0.000*	*0.000*
	y 2	0.097	0.197	0.058	0.053	0.041
		0.144	*0.014*	*0.077*	*0.096*	*0.139*
	y 3	0.309	0.220	0.933	0.107	0.087
		0.153	*0.024*	*0.110*	*0.197*	*0.139*
	y 4	0.402	0.024	0.008	0.839	0.871
		0.182	*0.019*	*0.149*	*0.211*	*0.134*

Table 3 (continued)

Males 40-59 R-squared=.942

Income Category in 1980		Income Category in 1970 xo	x 1	x 2	x 3	x 4
	y 0	0.160	0.044	0	0	0
		0.307	*0.010*	*0.0004556*	*0.000*	*0.000*
	y 1	0.011	0.633	0	0	0
		0.681	*0.024*	*0.000*	*0.000*	*0.000*
	y 2	0	0.168	0.167	0.143	0.029
		0.000	*0.014*	*0.101*	*0.144*	*0.077*
	y 3	0.004	0.137	0.765	0.501	0
		0.652	*0.014*	*0.086*	*0.086*	*0.000*
	y 4	0.825	0.018	0.068	0.356	0.971
		0.724	*0.014*	*0.091*	*0.139*	*0.077*

Males 60 and over R-squared=.901

Income Category in 1980		Income Category in 1970 xo	x 1	x 2	x 3	x 4
	y 0	0	0.034	0.000	0	0.026
		0.000	*0.013*	*0.007*	*0*	*0.040*
	y 1	0.010	0.516	0	0	0
		0.411	*0.033*	*0.000*	*0.000*	*0.000*
	y 2	0.481	0.235	0.004	0.067	0.247
		0.607	*0.040*	*0.139*	*0.153*	*0.149*
	y 3	0.510	0.194	0.863	0.314	0
		0.574	*0.056*	*0.179*	*0.153*	*0.000*
	y 4	0	0.021	0.134	0.618	0.728
		0.000	*0.033*	*0.146*	*0.216*	*0.156*

Table 3 (continued)

Females 15-19 R-squared = .680

Income Category in 1980		Income Category in 1970: x 0	x 1	x 2	x 3	x 4
	y 0	**0.309**	0.037	0	0	0
		0.103	*0.020*	*0.000*	*0.000*	*0*
	y 1	**0.492**	0.496	0	0	0
		0.322	*0.050*	*0.000*	*0.000*	*0.000*
	y 2	**0.133**	0.153	0.118	0.077	0.502
		0.126	*0.036*	*0.172*	*0.289*	*0.468*
	y 3	**0.066**	0.230	0.882	0.502	0.000
		0.136	*0.023*	*0.172*	*0.129*	*0.050*
	y 4	**0.000**	0.084	0	0.421	0.498
		0.033	*0.010*	*0.000*	*0.222*	*0.239*

Females 20-39 R-squared = .906

Income Category in 1980		Income Category in 1970: x 0	x 1	x 2	x 3	x 4
	y 0	0.505	0.066	0	0	0
		0.072	*0.010*	*0*	*0*	*0*
	y 1	**0.048**	0.696	0	0	0
		0.106	*0.029*	*0.000*	*0.000*	*0.000*
	y 2	**0.181**	0.089	0.186	0.241	0.033
		0.072	*0.010*	*0.082*	*0.082*	*0.120*
	y 3	**0.266**	0.096	0.700	0.341	0.369
		0.091	*0.019*	*0.144*	*0.235*	*0.216*
	y 4	0	0.053	0.115	0.418	0.598
		0.0004311	*0.010*	*0.106*	*0.173*	*0.153*

Table 3 (continued)

Females 40-59 R-squared = .873

Income Category in 1980		Income Category in 1970				
		x 0	x 1	x 2	x 3	x 4
	y 0	0.514	0.090	0.000	0.000	0
		0.206	*0.019*	*0.034*	*0.019*	*0*
	y 1	0.230	0.764	0	0	0
		0.163	*0.029*	*0.000*	*0.000*	*0.000*
	y 2	0.118	0.072	0.379	0.009	0.163
		0.115	*0.014*	*0.149*	*0.106*	*0.158*
	y 3	0.110	0.061	0.441	0.650	0.163
		0.125	*0.010*	*0.168*	*0.134*	*0.120*
	y 4	0.029	0.013	0.180	0.340	0.674
		0.077	*0.010*	*0.120*	*0.177*	*0.201*

Females 60 and over R-squared = .463

Income Category in 1980	Inc. Cat. in 1980	Income Category in 1970				
		x 0	x 1	x 2	x 3	x 4
	y 0	0	0.114	0	0	0
		0.000	*0.027*	*0.000*	*0*	*0*
	y 1	0.168	0.687	0.193	0	0
		0.531	*0.093*	*0.507*	*0.000*	*0.000*
	y 2	0	0.130	0.000	0.470	0.509
		0.000	*0.036*	*0.229*	*0.335*	*0.531*
	y 3	0.176	0.055	0.807	0.530	0.491
		0.474	*0.083*	*0.494*	*0.352*	*0.564*
	y 4	0.657	0.015	0.000	0.000	0.000
		0.730	*0.033*	*0.066*	*0.153*	*0.295*

Note: Figures in italics are jackknife estimates of standard errors.

TABLE 4

ANNUAL GROWTH RATES OF REAL INCOME BY COHORT

AGE IN 1970	MALE	FEMALE
15-19	19.4%	15.9%
20-29	11.5	7.9
30-39	7.8	4.8
40-49	6.5	3.9
50-59	5.5	3.5
+ 60	7.9	3.2
Overall Brazil	10.6	7.9

Source: Computed from census tapes.

TABLE 5

SIGN AND SIGNIFICANCE OF GROWTH COEFFICIENTS

	Downward		Upward	
	Positive	Negative	Positive	Negative
Males				
15-19	0	3	4(2)	5(1)
20-39	0	4	8	2
40-59	2	1	7	2
Total	2	8	19(2)	9(1)
Females				
15-19	0	3	4	4
20-39	0	4(2)	6(1)	0
40-59	0	4	6(1)	4
Total	0	11(2)	16(2)	8

(number of significant coefficents in parentheses)

Table 6. Forecast Results

y0	y1	y2	y3	y4
	Weighted Population Shares for Males 15-19			
.0323	0.1812	0.1362	0.4172	0.2331
	Predicted Population Shares			
.0361	0.2471	0.1392	0.3804	0.2063
	Weighted Population Shares for Males 20-39			
.0136	0.1494	0.1111	0.3691	0.3568
	Predicted Population Shares			
.0196	0.1802	0.1039	0.3575	0.3382
	Weighted Population Shares for Males 40-59			
.0151	0.2106	0.1349	0.3314	0.3079
	Predicted Population Shares			
.0150	0.2208	0.1335	0.3267	0.3007
	Weighted Population Shares for Males 60 and Over			
.0192	0.2722	0.1693	0.3296	0.2098
	Predicted Population Shares			
.0231	0.2819	0.1707	0.3258	0.2018
	Weighted Population Shares for Females 15-19			
.0774	.3416	.1595	.3239	.0976
	Predicted Population Shares			
.0751	.4002	.1469	.3499	.0829
	Weighted Population Shares for Females 20-39			
.0623	.2835	.1485	.3181	.1876
	Predicted Population Shares			
.0663	.3138	.1398	.3019	.1764
	Weighted Population Shares for Females 40-59			
.0770	.3861	.1358	.2453	.1558
	Predicted Population Shares			
.0800	.4021	.1272	.2312	.1583
	Weighted Population Shares for Females 60 and Over			
.1004	.4751	.1509	.2507	.0230
	Predicted Population Shares			
.0862	.4874	.1626	.2223	.0390

REFERENCES

Adelman, I. (1989), "What is the Evidence on Income Inequality and Development?" Giannini Foundation Working Paper 502.

Adelman, I. and P. Whittle (1980), "Static and Dynamic Indices of Income Inequality," Canadian Journal of Development Studies, 1, 27-46.

Adelman, I., K. Subbarao, and P.S. Vashista (1985), "Some Dynamic Aspects of Rural Poverty in India," Economic and Political Weekly, 20 (39), 575-287.

Bonelli, R. and G.L. Sedlacek (1989), "Distribuicao de renda: evolucao no ultimo quarto de seculo," In G.L. Sedlacek and R. Paes de Barros, eds., Mercado de Trabalho e Distrribucao de Renda: uma Coletanea, (Rio de Janeiro: IPEA), 7-24.

Bacha, E.L. and L. Taylor (1978), "Brazilian Income Distribution in the 60's: Facts, Model Results and the Controversy," Journal of Development Studies 14, 271-297.

Crewe, I. and C. D. Payne (1966), "Another Game with Nature: An Ecological Regression Model of the British Two Party Vote Ratio in 1970," British Journal of Political Science 7, 43-81.

Denslow, D. and W.G. Tyler (1983), Perspectives on Poverty and Income Inequality in Brazil; An Analysis of the Changes During the 1970's, (Washington: The World Bank Staff Working Papers #61.

Fields, G. (1977), "Who Benefits from Economic Development? A Re-examination of Brazilian Growth in the 1960's," American Economic Review 67, 570-582.

Fishlow, A. (1972), "Brazilian Size Distribution of Income," American Economic Review 62, 391-402.

Goodman, L. (1953), "Ecological Regression and the Behavior of Individuals," American Sociological Review 18, 663-664.

Hoffman, R. and A.A. Kageyama (1986), "Distribuicao da renda no Brasil, entre Familias e entre Pessoas, em 1970 e 1980," Estudos Economicos 16, 25-51.

Langoni, C. G. (1973), Distribucao de Renda e Crescimento Economico do Brasil. (Rio de Janeiro: Expressao e Cultura).

Lee, T.C., G.G. Judge, and A. Zellner (1970), Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data (Amsterdam: North Holland).

Morley, S.A. (1981), "The Effect of Changes in the Population on Several Measures of Income Distribution," American Economic Review 71, 285-294.

__________ (1982), Labor Markets and Inequitable Growth: The Case of Authoritarian Capitalism in Brazil. (New York: Cambridge University Press).

Morley, S.A. and J.G. Williamson (1975), "Crescimento, Politica Salarial e Desigualdade: o Brasil Durante a Decada de 1960," Estudos Economicos 5, 107-139.

Pfeffermann, G. and R. Webb (1979), The Distribution of Income in Brazil. (Washington: The World Bank Working Paper #356).

Prais, J., (1955), "Measuring Social Mobility," Journal of the Royal Statistical Society, Series A-118, 55-66.

Shorrocks, A.F., (1976), "Income Mobility and the Markov Assumption," Economic Journal 86, 566-577.

Telser, L. (1963), "Least Squares Estimates of Transition Probabilities," in C. Christ et al., eds. Measurement in Economics (Stanford: Stanford University Press.

ENDNOTES

[1]The paper by Adelman 1989, in which the dependent variables are the rates of change of the income share between 1960 and 1970 and 1970 and 1980, is an exception.

[2]This view of income mobility was introduced by Prais (1955), and adopted by Shorrocks (1976), Adelman and Whittle (1980) and Adelman, Subbarao and Vashistra (1985).

[3]Bonelli and Sedlacek (1988), p. 14.

[4]The average 1979 income of workers aged 15-29 in 1970 was 4,289 CR$ compared to an average of 6,159 for the population as a whole.

[5]Degrees of freedom for the regressions are calculated according to:

df = # of equations X # of observations - # of coefficients +# of cross equation constraints +# of binding inequality constraints

Yielding:

	Regression	Degrees of Freedom
Males	15-19	60-50+5+17 = 32
	20-39	120-50+5+12 = 87
	40-59	120-50+5+16 = 91
	60 and over	60-50+5+16 = 31
Females	15-19	60-50+5+18 = 33
	20-39	120-50+5+14 = 89
	40-59	120-50+5+12 = 87
	60 and over	60-50+5+23 = 38

[6]The all-Brazil data was omitted from the regional census data on which the regression estimates of the P_{ij} are based.

DATA APPENDIX

The appendix gives a brief description of the census tapes we used and outlines the steps to calculate the income distribution of workers in 1970 who will stay in the labor force until at least 1980 and the income distribution of workers in 1980 who already worked in 1970. Hence we consider only workers who neither retire nor enter the work force during the period considered. If a worker migrated during the decade we determined her region of origin.

For 1970 we used public use data available in the one-percent census sample file. This public use data tape records information about 910,808 individuals living in 193,889 households of various types. We excluded the record of every individual younger than 14 years old or who is not in the labor force; 274,529 records of individuals remained. For 1980 the public use tape has 197,413 households and 893,278 individuals. We extracted the records of 202,453 workers over 25 years of age for our mobility estimations. Each of the data sets is a weighted sample of the entire Brazilian population.

We are interested in a particular set of characteristics to describe the individual. The characteristics are: region (12 outcomes), sex (2 outcomes), age group (6 outcomes), level of education (5 outcomes), income group (5 outcomes) and economic sector, which was only used in an intermediate step. The variable "region" describes where the worker was living in 1970. We divided Brazil into twelve regions whose boundaries are in many cases identical with the state boundaries. The reported income was multiplied by 24.75, the Rio de Janeiro cost of living index to inflate it to 1980 levels. Where respondents did not report income, we assigned the average income of workers with the same characteristics.

For 1980 we used two different versions of the public use tape because the official version did not contain information about the individual's previous residence. Fortunately we had a second version with this information as well as time of residence. This information was used to create the variable 1970 region for each 1980 respondent. We then constructed a 12 x 12 migration matrix for each possible age-education-sex combination. The elements of this matrix contain the number of workers who lived in the ith region in 1970 and in the jth region in 1980. We converted these entries to percentages of the 1980 population and applied these percentages to the official public use tape. We used the same percentages for each income class in the appropriate characteristics cell. This procedure redistributes the 1980 labor force into the regions they inhabited in 1970 and is thus compatible with the 1970 data.

If we now subtract the number of workers reported in the 1970 census for each particular set of characteristics from the number in 1980, we get a positive or negative number, indicating either net new entrants or retirements. Where new entrants are positive we calculate the number of surviving workers by subtracting the number of workers who have entered the labor force from the number of workers in 1980 with the same set of characteristics. Where new entrants are negative, we subtract our estimate of retirements from the 1970 total to get an estimate of those in the 1970 labor force who survived to 1980. Assuming that the proportion of new entrants or retirements are the same across income classes within each cell, we can then calculate the number of survivors in each income class in 1970 and 1980. This is the basic input to all the distribution data and regressions reported in the paper.