11

A MARKOVIAN MODEL OF INCOME DYNAMICS

by

John J. McCall

Discussion Paper Number 11
May, 1971

A MARKOVIAN MODEL OF INCOME DYNAMICS

John J. McCall*

I. INTRODUCTION

This paper presents an analysis of the Continuous Work History Sample of the Social Security Administration. The data cover the ten years 1957-1966 and include information on race, age, sex, and estimated annual earnings for 984,500 individuals. The earnings mobility of prime working age males and females is analyzed. The "prime working age" category includes all those between the ages of 25 and 55.[1] The unique features of the data are their longitudinal character and the large sample size. The first feature permits the generation of a ten year earnings profile for each individual in the sample. Because of the large sample size, considerable confidence can be attached to the empirical findings.

As with all empirical data, these are also deficient in several respects. An extensive discussion of these limitations is contained in Section III. For now, it should be noted that the time series is short (10 periods), the data are for individuals and not for families, and many people are not covered by Social Security. The second and third limitations are especially important to a study whose focus is on poverty dynamics. It is imperative that the conclusions of this study be considered in conjunction with these data limitations.

The analysis of the Social Security data follows the methodology outlined in [14] and exercised in [15, 18]. At some points it may be useful to refer to these papers.

In the recent literature on poverty, several authors have investigated

the influence of the rate of change of gross national product (economic growth) on the poverty population. All agree that growth does indeed have a beneficial effect on the poverty population. These authors disagree, however, concerning the extent of this positive influence. Some claim that almost all groups participate in growth with, in some cases, the gains being greater for the non-white poor than for the white poor.[2] Others are less sanguine about the overall effects of growth. They have advanced the "backwash thesis," which claims that certain subgroups in poverty are so isolated from our society that their economic welfare is immune to aggregate growth.[3] Finally, there are some who claim that sustained economic growth will, for all practical purposes, eventually eliminate poverty.[4] Clearly, these hypotheses have different policy implications. If the "backwash thesis" is true, then elimination of poverty requires special programs in addition to sustained economic growth. The willingness of the general populace to engage in anti-poverty programs is undoubtedly affected by the rate of growth of GNP, that is, the alleviation of poverty has a higher priority among the general populace when real incomes are rising. However, the "backwash thesis" does not consider these secondary effects of growth but maintains that the direct effects of growth are not sufficient for the elimination of poverty. If the direct effects of growth reached the entire population and tended to eliminate all poverty, less emphasis should be placed on specially designed anti-poverty programs such as Neighborhood Youth Corps and Job Corps.

Given the limitations of the Social Security data, the use of the word poverty in the context of this study refers only to covered earnings (earnings that are reported to Social Security) of employees and is

irrespective of family size. Consequently, there is no one-to-one correspondence between those who would be designated poor here and those who are poor according to the "official" definition of Poverty.[5] To avoid confusion this study distinguishes between two sets of individuals, those who have a low earning status (L) for the year under investigation and those who do not have a low earning status ($\bar{L}$). These two sets are obviously mutually exclusive and, when combined with the set of individuals who were not covered by Social Security for the particular year, the uncovered (U), compose the entire continuous work history sample for that year.

Accordingly, in order to test these hypotheses using Social Security data, individuals are classified into four groups, stayers in a low earning status (L), stayers in non-low earning status ($\bar{L}$), stayers in the uncovered category, and movers. Uncovered indicates that the individual's earnings have for one reason or another not been reported to Social Security for that particular year.[6] Stayers in L ($\bar{L}$) were those who remained in L ($\bar{L}$) for the entire ten year period. Stayers in uncovered were those whose earnings were not covered by Social Security for nine of the ten years. Movers constitute the remaining individuals.

Since the focus of this study is on movements into and out of the low earnings category, the income distribution was dichotomized according to three different definitions of L: $1,500, $3,000, and $4,500, mainly to measure the sensitivity of the empirical findings to changes in earning levels.[7] A subsidiary reason is the presence of several poverty standards in use--for example, the Orshansky definition and that of the family assistance plan. Although there is no one-to-one correspondence between the definition of low earnings and these poverty definitions, they are

irrespective of family size. Consequently, there is no one-to-one correspondence between those who would be designated poor here and those who are poor according to the "official" definition of Poverty.[5] To avoid confusion this study distinguishes between two sets of individuals, those who have a low earning status (L) for the year under investigation and those who do not have a low earning status ($\bar{L}$). These two sets are obviously mutually exclusive and, when combined with the set of individuals who were not covered by social security for the particular year, the uncovered (U), compose the entire continuous work history sample for that year.

Accordingly, in order to test these hypotheses using Social Security data, individuals are classified into four groups: stayers in a low earning status (L), stayers in non-low earning status ($\bar{L}$), stayers in the uncovered category, and movers. Uncovered indicates that the individual's earnings have for one reason or another not been reported to Social Security for that particular year.[6] Stayers in L ($\bar{L}$) were those who remained in L ($\bar{L}$) for the entire ten year period. Stayers in uncovered were those whose earnings were not covered by Social Security for nine of the ten years. Movers constitute the remaining individuals.

Since the focus of this study is on movements into and out of the low earnings category, the income distribution was dichotomized according to three different definitions of L: $1,500, $3,000, and $4,500, mainly to measure the sensitivity of the empirical findings to changes in earning levels.[7] A subsidiary reason is the presence of several poverty standards in use--for example, the Orshansky definition and that of the family assistance plan. Although there is no one-to-one correspondence between the definition of low earnings and these poverty definitions, they are

certainly positively related.

Section II of this study summarizes the Markovian methods used to formulate the model. The hypotheses that emerge from this model are also presented together with the appropriate statistical tests.

Section III describes the Social Security data, mentioning some of their limitations for the analysis of income dynamics. The results of the various statistical tests are then reported for males in the following order: (1) the validity of the stayer-mover model is established; (2) the equality of stayer probabilities is then tested for various race and age subgroups; (3) tests for the stationarity of the mover matrixes are then performed; (4) the stationarity distributions associated with several of the mover matrixes are reported; (5) the simple least squares regressions of the transition probabilities on percentage change in GNP are calculated. The results of a logit analysis of these same data are also indicated.

Section IV reports the results of the various statistical tests for females. The format of this section is identical to that of Section III.

Section V presents a summary of the main findings.

II. A MARKOVIAN MODEL OF EARNINGS MOBILITY

It is assumed that movements into and out of the low earnings category can be described by a modified Markov model. The usual first order Markov model assumes that movements into and out of L depend only on one's present income position (L or $\bar{L}$) and are independent of all prior income history. Letting 1 and 2 denote L and $\bar{L}$, respectively, the p_{ij} entry in the matrix,

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

denotes the conditional probability of going to j in the next period (year) given that you are currently in i. The mathematical analysis of such stochastic processes is fairly elementary and accounts for their popularity as first approximations to the dynamics of real world phenomena.[8]

On purely theoretical grounds, a model as simple as this would yield an inadequate description of income dynamics. In theory the length of time in L or $\bar{L}$ should have an important influence on the probability of moving to L or $\bar{L}$ during the subsequent period. For this reason a modified Markov process is used to describe movements into and out of L. In this modified process, four different classes are distinguished: stayers in L, stayers in $\bar{L}$, stayers in uncovered (U), and movers; uncovered indicates that the individual has for one reason or another not been covered by Social Security for the period in question. The behavior of the movers is assumed to follow a simple first order Markov process, while the stayers are assumed to remain in L, $\bar{L}$, or U with probability one. This Markovian formulation proved to be a very convenient device for organizing the massive Social Security longitudinal file.

In the analysis here an individual can occupy three states: L, $\bar{L}$, and U. Denoting L, $\bar{L}$, and U by 1, 2, 3, respectively, the law of motion for this process is given by:

$$P_{ij} = \begin{cases} s_i + (1-s_i)\, m_{ii}, & i = j \quad (i,j = 1,2,3) \\ (1-s_i)\, m_{ij}, & i \neq j, \end{cases}$$

In this process the proportion of stayers in L, $\bar{L}$, and U are denoted by s_1, s_2, and s_3, respectively.[9] The mover matrix, m_{ij}, is assumed to be a simple first order Markov process. For long time series, Goodman suggests

the following approximations to maximum likelihood estimators of the parameters

$$m_{ij} = h_{ij} \quad (i, j=1, 2, 3)$$

and

$$s_i = f_i \quad (i=1, 2),$$

where h_{ij} is the proportion of individuals in the i^{th} income class in the initial period who were in the j^{th} income class in the following period (considering only individuals in the i^{th} income class in the initial period who were not continuously in the class for all n periods), and f_i is the proportion of individuals in the i^{th} income class in the initial period who remained in that class for all n periods. Even though this time series is quite short, these approximations will be made.[10]

The first economic hypothesis that emerges from this model is the "backwash thesis," which maintains that there are hard core poor who are immune to overall economic growth. This hypothesis is easily tested by determining if s_i is significantly different from zero. [9] This test can also be used to test whether the stayer proportions in $\bar{L}$ and U are significantly different from zero. In this way the validity of the stayer-model can be ascertained.

The second hypothesis to be tested is the equality of both s_1 and s_2 for whites (W) and non-whites (N). Because of racial discrimination one would anticipate that $s_1^N - s_1^W > 0$ and $s_2^W - s_2^N > 0$.[11]

The third economic hypothesis is that because of economic growth the mover transition matrix will be nonstationary over the 1957-66 time period. The stationarity of these matrixes can be tested using χ^2 methods developed by Goodman [9]. The hypothesis not only predicts nonstationarity but also

the specific behavior of four of the transition probabilities over the period of growth. In particular, the L to L and $\bar{L}$ to L transition probabilities should be negatively related to g, the percentage change in GNP, whereas the L to $\bar{L}$ and $\bar{L}$ to $\bar{L}$ transition probabilities should bear a positive relation to g. These hypotheses are easily tested by running simple linear regressions of each transition probability on g. Since the dependent variable is a probability and must be in the interval (0, 1), a logit regression is more appropriate than the simple linear regressions. The logit regression proceeds in two stages in estimating the regression

$$\ln \left(\frac{p}{1-p}\right) = \alpha + \beta g$$

where p is a transition probability.[12] For a discussion of this procedure see [13, 25].

The following argument suggests that non-white movers should benefit more from the tight labor markets that accompany sustained economic growth than their white counterparts. In periods of tight labor markets the number of qualified whites who are currently searching for employment diminishes as the white unemployed labor pool becomes dominated by those who have tried and failed. Under these circumstances, it will no longer be economical to use such simple screening devices as race, and employers will begin sampling from the non-white distribution.[13] This hypothesis can be tested by measuring the relative changes in white and non-white transition probabilities over this ten year period.

The estimated relations between transition probabilities and percentage change in growth can be used to generate a steady state distribution that is also a function of g, by replacing each of the nine transition probabilities of the mover matrix by its estimate.

$$\hat{m}_{ij} = \hat{\alpha}_{ij} + \hat{\beta}_{ij}\, g$$

The associated stated steady state vector (π_1 (g), π_2 (g), π_3 (g)) can then be calculated for various values of g.

III. EMPIRICAL FINDINGS FOR PRIME WORKING AGE MALES

This section describes the Social Security data, emphasizing their limitations for poverty analysis, and presents empirical results for the 1957-1966 Continuous Work History Sample. The focus will be on males in the prime working age group (25-54). Within this group, three subgroups are distinguished: 25-34, 35-44, and 45-54; and three earning levels are discussed, $1,500, $3,000, and $4,500.

Social Security data were available for the years 1957-1966. These data include the one percent Continuous Work History Sample maintained by the Social Security Administration as an aid in administering the Old Age Survivors Disability and Health Insurance system (OASDHI). Earnings are reported to the Social Security Administration for about 80 million people each year under OASDHI. Government employees are the largest group not covered by this program. However, for our analysis here the most important group not covered by Social Security comprises domestics, migrant laborers, other casual workers, and the unemployed. It is likely that most members of this group are low earners.

Basic demographic data are obtained when the individual applies for a social security number. These include date of birth, race, sex, and place of birth. When employment commences in a covered activity, the employer reports quarterly wages together with geographic location (by county and state) and industrial affiliation (by 4-digit industry codes of the Standard Industrial Classification Manual). These quarterly data and the

basic demographic information are the main ingredients of the Continuous Work History Sample.

Although such data are unique in their coverage and content, they do have certain limitations for the analysis of low earners, the most important of which are:

(1) The data are collected for individuals rather than for households. Consequently, an individual may be earning low wages and be a member of a prosperous household, or he may be earning high wages (when considered by himself) while he and his large family are clearly destitute.

(2) There is no information on income from non-employment sources. An individual may appear to be in poverty on the basis of his wage income but be relatively affluent when his non-wage (dividends, interest, and so on) income is considered.[14]

(3) The problem of uncovered employment is perhaps the most important deficiency in the Social Security data. No information is available concerning an individual's income behavior while he is in the uncovered state. Furthermore, the movement from covered to uncovered can occur for a variety of indistinguishable reasons. Among the most important are: (a) if the maximum taxable income is earned in a prior quarter, the individual will appear as uncovered in the current quarter; (b) if the individual is unemployed for whatever reason, death, disability, retirement, etc.,[15] or earns less than $50 he will appear as uncovered in the quarter; (c) if his employer fails to file a quarterly report (casual labor and domestics), he

will appear uncovered; (d) if an individual in fact does move from covered to uncovered employment, he will appear as uncovered; (e) coverage changed over the ten year period, 1957-66.[16]

In spite of these limitations these data seem to possess great potential for studying the dynamics of poverty and income formation.

For each earning level, an individual may occupy one of three states, L, $\bar{L}$, and U. Movements among these states are assumed to be governed by the stayer-mover probability process described in the previous section. With three income classes, three age categories, two sex categories, and two race categories, there are 36 states. The single 36 x 36 transition matrix naturally decomposes into twelve 3 x 3 transition matrixes. Transition matrixes like these are calculated for each of the successive periods, 1957-1958, 1958-1959, . . . 1965-1966. Similar transition matrixes are calculated for each of the three low earnings definitions. Those individuals who remained in L or $\bar{L}$ for the ten periods and those who were uncovered nine of the ten are withdrawn from these transition matrixes giving rise to revised transition matrixes, the mover matrixes. Comparisons between male non-whites and male whites will be our primary concern here. Similar comparisons between sexes and between female non-whites and female whites are presented in the next section. Because of the ambiguity of the uncovered category, special assumptions are made to condense the 3 x 3 matrixes to 2 x 2 matrixes, i.e., the uncovered state is eliminated.

Two kinds of stayer proportions or probabilities will be calculated. The first is the conditional probability of remaining in L ($\bar{L}$) over the whole time period, given that a person was in L ($\bar{L}$) in 1957. This stayer

proportion is estimated by dividing the total number in L ($\bar{L}$) for all ten periods by the number in L ($\bar{L}$) in the first period. This is done for each of the six male non-income categories. These proportions are displayed in Table 1 for each of the three earning levels. The proportion of male white stayers in L is significantly less (at .001) than the proportion of male non-white stayers in L for each of the three age categories.

Table 1 reports the male stayer proportions by race, age, and earnings level. The stayer proportions are all significantly different from zero.[17] This validates the use of the stayer-mover model. The fact that the proportion of stayers in L exceeded zero tends to substantiate the backwash thesis. There are groups in our society who were relatively immune to the economic growth that occurred in the 1957-66 period. Furthermore, the probability of staying in L for the remaining nine years given poverty in 1957 is an increasing function of age. This probability is, of course, also an increasing function of the earning level. The difference in these probabilities across race agree with the hypotheses in Section II: the probability of staying in L was significantly greater for non-whites than for whites[18] and the probability of staying in $\bar{L}$ was significantly greater for whites than for non-whites.[19] These results were true for all age groups and all earnings lines.

The proportion of stayers in L is positively related to age for both whites and non-whites and for all three definitions of low earnings. This is not surprising because younger workers may be investing in job skills in the first few years and then transiting to non-poverty in one of the later years. Perhaps more importantly, the younger worker is more mobile and in general more adaptable to improving economic conditions. The

proportion of stayers in $\bar{L}$ was greatest for the 35-44 age group. This was true for both whites and non-whites and for all three definitions of L.

The second stayer probability is the unconditional probability of remaining in L or $\bar{L}$ for the ten periods. This probability is obtained by dividing the total number of stayers in L or $\bar{L}$ by the total number in the appropriate category; for example, male, white, 25-34.[20] These stayer probabilities are reported in Table 2 for male whites and non-whites, for each age group and for each low earnings definition. Within each category the proportion of movers is obtained by subtracting the proportion of stayers in L plus the proportion of stayers in $\bar{L}$ from one. The proportion of movers in each non-income category is also recorded in Table 2. It is important to keep these mover proportions in mind when discussing differences in various mover matrixes. The relations between the male white and male non-white unconditional proportions are the same as those for the conditional stayer proportions.

The proportion of movers is greater for non-whites than for whites. This is true for all age groups and for all poverty lines.

ANALYSIS OF THE MALE MOVER MATRIXES FOR 1960 AGE GROUP[21]

The male mover matrixes are first analyzed for the 1960 age groups; that is, an individual is placed in one of the three age groups based on his age as of 1960. The behavior of each individual is then analyzed for the ten years.

Deleting the stayers from the transition matrixes gives the mover matrixes discussed earlier. Mover matrixes are shown in Table 3 for white and non-white males between the ages of 35 and 44 (age as of 1960) for each of the transition periods, 1957-1958, 1960-1961, and 1965-1966, when a

$3000 earnings line is used. The steady state distributions associated with the 1957-1958 and 1965-1966 mover matrixes are also presented in Table 3 -B. Table 3 -C shows the corresponding population (stayers and movers) steady state distribution. These distributions show the proportion of individuals in each category at any point in time assuming that the particular transition matrix persists indefinitely. More specifically, let π_i be the steady state proportion in state i. Then the vector $\pi = (\pi_1, \pi_2, \pi_3)$ denotes the steady state distribution. Letting M denote the mover matrix, π can be calculated from the following equations:

$$\pi = \pi M \quad \text{and}$$

$$\Sigma\pi_i = 1.$$ [22]

Casual observation of the transition matrixes reveals their non-stationarity, that is, the transition probabilities are not constant over time.[23] For example, the probability of a non-white transiting from L to L in 1957-58 was .74, in 1960-61 .68, and in 1964-65 .59. As previously hypothesized, one would anticipate that the persistence of high growth rates would affect the mover transition matrix. In particular, it would be expected that for the movers, as GNP increased the probabilities of transits from L to L and $\bar{L}$ to L would decline, while the probabilities of L to $\bar{L}$ and from $\bar{L}$ to $\bar{L}$ would increase.[24]

These regressions tend to verify the growth hypotheses. Note first that for both whites (W) and non-whites (N), the coefficients of g have the right sign. Furthermore, the values of these coefficients are highly significant. Finally in all four cases the absolute size of the g coefficient is larger for non-whites than for whites. This is consistent with the hypothesis that non-white movers are more sensitive to growth than

their white counterparts. Results like these held for all age groups and all earning levels.[25] The results of the simple linear regressions are virtually identical to both the weighted and unweighted logit regressions. For this reason the logit results are not presented.

ELIMINATION OF THE UNCOVERED CATEGORY

The uncovered category can be removed by the following method. Let $\lambda = \frac{P_{31}}{P_{31}+P_{32}}$, represent the fraction of individuals moving (between years) from uncovered to L who were previously in L; it also represents the fraction of individuals moving (between years) from L to uncovered who stay in L; finally, it represents the fraction of those who stayed in uncovered (for two consecutive years) who also remained in L. The fraction, $(1 - \lambda)$ has a similar interpretation for the $\bar{L}$ category. A λ was calculated for each of the nine mover matrixes. Several of the condensed mover matrixes are presented in Table 5.[26]

The associated steady state proportions are displayed in Table 5-B, for the movers, and in 5-C for both movers and stayers.

For the 35-44 group and a $3000 earning level, the simple regressions of the revised transition probabilities on g were (all statistically significant):

	Durbin-Watson Statistic
$p_{11}^{W} = .78 - .61g,\ r^2 = .52$	.89
$p_{12}^{W} = .32 - 1.3g,\ r^2 = .54$	1.7
$p_{11}^{N} = .91 - 1.0g,\ r^2 = .73$	2.0
$p_{21}^{N} = .41 - 2.6g,\ r^2 = .64$	1.8

Comparing both these regressions with those previously obtained and Table 3 with Table 5 demonstrates invariant conclusions. Both white and non-white movers benefit from increases in g with non-whites benefiting more.

IV. EMPIRICAL FINDINGS FOR FEMALES

The presentation of results follows the same order as in Section III. Because of these similarities, the tabular presentations are accompanied by very brief descriptions.

The empirical findings presented here emphasize comparisons of the economic performance between white and non-white females. Differences among age categories (25-34, 35-44, 45-54) are also noted for both.

A cursory examination of the differences between males and females clearly illustrates that the economic performance of males is, in general, superior to that of females. There are several explanations of this, the most obvious being job market discrimination and the fact that because of the females' role in our society many utility producing services they perform are not or cannot be converted into their income equivalents. By the same fact, females' participation in the labor force lacks the stability of the male. For this reason employers are less willing to invest in female human capital. This lowers their productivity and wages.

The effect of improved economic conditions on female labor participation rates could in theory be either positive or negative.[27] The rise in family income accompanying growth could cause the wife to reduce her labor force participation. On the other hand, occupations previously closed to women would in periods of sustained growth demand their services. These improved job opportunities could cause female labor participation to

increase. Unfortunately, the results of the analysis here are inconclusive on this point since the unit of measure is the individual and not the family. Nevertheless, they do suggest that, for both whites and non-whites, the female labor participation rate is positively related to economic growth.

With respect to females whose low earning (L) or non-low earning ($\bar{L}$) status is affected by sustained growth (the so-called movers), both whites and non-whites exhibited substantial improvements, with non-whites benefiting slightly more than whites. These differences in white and non-white performance are not nearly as great, however, as those discovered in the case of non-white and white male movers. These relative improvements were again measured by changes in the transition probabilities of the mover matrix. For both white and non-white females, the probabilities of moving from L to $\bar{L}$ and from $\bar{L}$ to $\bar{L}$ increased, while the probabilities of moving from $\bar{L}$ to L and from L to L declined.

Contrary to the performance of male movers, the non-white females did not always benefit more from growth than their white counterparts. The benefits accruing to non-whites relative to whites as a consequence of growth depended on age, earning level, and the transition probability used to measure improvement.

Table 6 presents the female stayer proportions by race, age and earning level. The stayer proportions are all significantly different from zero at the .01 level. Thus the stayer-mover Markov model gives a better description of female earnings mobility than the simple Markov model. The proportion of females who remained in L through 1966 given occupancy in L during 1957 is significantly higher for non-whites than for whites at the .001 level. This result is independent of age and earning level and corresponds to the

analysis of non-white male stayers relative to whites (see Table 1). However, this finding does require more interpretation than in the case of males. Perhaps those who were stayers in L were mainly working wives who chose to work part-time. This topic requires further study. It should be noted though that approximately one-quarter of non-white families have a female head. With the exception of the 25-34 age group, the non-low earning stayer proportions were significantly (at .005) greater for whites than for non-whites. This was true for all earning levels. For the 25-34 age group the differences were significant for the $1500 (at .005) and $3000 (at .10) earning levels, but insignificant for the $4500 earning level.

The proportion of stayers in L is positively related to age for both whites and non-whites at all three earning levels.

The second stayer probability is the unconditional probability of remaining in L or $\bar{L}$ for the ten periods. This probability is obtained by dividing the total number of stayers in L or $\bar{L}$ by the total number in the specific non-income category, for example, female, white, 25-34. These stayer probabilities are reported in Table 7 for female whites and non-whites, for each age group and for each earning level. Note that the probabilities of staying in L are somewhat similar for non-white males and white females (compare Table 2). Within each non-income category the proportion of movers is obtained by subtracting the proportion of stayers in $\bar{L}$ from one. The proportion of movers in each non-income category is also recorded in Table 7. It is important to keep these mover proportions in mind when discussing differences in various mover matrixes. In Table 7 note that the proportion of movers is roughly the same for whites and non-whites for each of the age categories and earning levels. This was not

the case for males.

ANALYSIS OF THE MOVER MATRIXES FOR 1960 AGE GROUP

The mover matrixes are first analyzed for the 1960 age groups; that is, an individual is placed in one of the three age groups based on his age as of 1960. The behavior of each individual is then analyzed for the ten years.

Deleting the stayers from the transition matrixes gives the mover matrixes discussed earlier. Mover matrixes are shown in Table 8 for white and non-white females between the ages of 35 and 44 (age as of 1960) for each of the transition periods, 1957-1958, 1960-1961, and 1965-1966, when a $1500 earning level is used. The steady state distributions associated with the 1957-1958 and 1965-1966 mover matrixes are also presented in Table 8 - B. These distributions show the proportion of individuals in each category at any point in time assuming that the particular transition matrix persists indefinitely. The corresponding population (movers and stayers) steady state distributions are shown in Table 8 - C.

Casual observation of the transition matrixes reveals their non-stationarity, that is, the transition probabilities are not constant over time.[28] For example, the probability of a non-white transiting from $\bar{L}$ to $\bar{L}$ in 1957-1958 is .65, in 1960-1961 is .80, and 1964-1965 is .83. As previously hypothesized, one would anticipate that changes in growth rates would affect the mover transition matrix. To measure this relation each of the transition probabilities was regressed on percentage change in GNP. For the 35-44 age group and a $1500 earning level the simple regressions shown in Table 9 were obtained.

These regressions yielded mixed results. In some cases, there was no significant relationship between g and a particular transition probability

while in others the relation was significant. Also in some cases non-whites were more strongly influenced by g than whites while in other cases the reverse was true. Contrary to the findings for males, movements from uncovered to $\bar{L}$ are strongly influenced by g for both whites and non-whites. Changes in g also influence movements of non-whites from U to L. Female labor participation rates are sensitive to changes in economic growth.

V. CONCLUSIONS

This study has presented an analysis of the Continuous Work History Sample of the Social Security Administration. This longitudinal sample covered the years 1957-1966 and included 984,500 individuals. The analysis focused on males and females in the 25-55 age group. The earnings mobility of both groups were investigated using a simple Markov model. Movements across three different earning levels ($1500, $3000, $4500) were measured. The following results were obtained for males.

(1) The probability of remaining in a low earnings category all ten years given low earnings in 1957 was significantly greater than zero. This was true for both white and non-white males for all age groups, and for each of the three measures of low earnings.

(2) The probability of remaining in a low earnings category all ten years given low earnings in 1957 was significantly larger for male non-whites than for male whites. This was true for all age groups and all earnings levels.

(3) The probability of remaining in a non-low earnings category all ten years given high earnings in 1957 was significantly greater than zero. This was also true for both white and non-white males for all age groups and for each of the three measures of low earnings. This result in

conjunction with (1) validates the use of the stayer-mover Markov model.

(4) The probability of remaining in a non-low earnings category all ten years given high earnings in 1957 was significantly greater for male whites than for male non-whites. This was true for all age groups and all earning levels.

(5) The low earning and non-low earning stayers were deleted from the sample and the behavior of the male movers was analyzed. For all male movers, a strong systematic relationship was observed between each of the low earning and non-low earning transition probabilities and percentage change in GNP.

(6) Economic growth had a stronger influence on non-white male movers than on white male movers, that is, the relative changes in the transition probabilities across each of these earnings levels were greater for non-whites than for whites.

Analysis of the female population produced the following results:

(1) The probability of remaining in a low earnings category all ten years given low earnings in 1957 was significantly greater than zero. This was true for white and non-white females, for all age groups, and for each of the three measures of low earnings.

(2) Year-by-year transition matrixes were also calculated for the female movers, that is, all those who were not stayers in L, $\bar{L}$, or U. Both whites and non-white females benefited from growth. However, the relationships between the four L-$\bar{L}$ transition probabilities and percentage change in GNP were not nearly as strong as those for males. Differences between white and non-white female movers were also much smaller than for males.

(3) The transition probabilities from U to L and from U to $\bar{L}$ were more closely associated with percentage change in GNP than the corresponding male probabilities. These results suggest that the labor participation rates for females identified as movers have a strong positive relation to growth.

Table 1. MALE STAYER PROPORTIONS, 1957-1966[a]

Low earnings			Non-low earnings	
Non-white	White	Age group	None-white	White
$1500 earning level		(as of 1960)	$1500 earning level	
.03	.02	25-34	.53	.69
.05	.04	35-44	.57	.73
.06	.05	45-54	.55	.70
$3000 earning level			$3000 earning level	
.14	.05	25-34	.48	.64
.17	.08	35-44	.53	.69
.19	.10	45-54	.52	.65
$4500 earning level			$4500 earning level	
.28	.10	25-34	.34	.57
.31	.14	35-44	.40	.62
.32	.17	45-54	.39	.58

[a]These proportions can be interpreted as the conditional probability of staying in L ($\bar{L}$) all ten years given presence in L ($\bar{L}$) in 1957.

Table 2. MALE STAYERS AND MOVERS AS PROPORTIONS OF EACH AGE-RACE CATEGORY

Non-white stayers in					White stayers in			
L	$\bar{L}$	U	Non-white movers	Age group[a]	L	$\bar{L}$	U	White movers
				(1500)				
.008	.25	.05	.69	25-34	.003	.49	.03	.48
.011	.32	.05	.62	35-44	.004	.54	.04	.41
.015	.32	.06	.61	45-54	.005	.50	.05	.44
				(3000)				
.07	.11	.05	.77	25-34	.01	.36	.03	.59
.08	.18	.05	.69	35-44	.02	.45	.04	.50
.09	.18	.06	.68	45-54	.02	.41	.05	.52
				(4500)				
.20	.03	.05	.73	25-34	.05	.18	.03	.74
.20	.06	.05	.69	35-44	.05	.29	.04	.62
.21	.06	.06	.68	45-54	.07	.26	.05	.63

[a]In the 25-34 age group there were 12,760 non-whites, 84,447 whites.

In the 35-44 age group there were 12,757 non-whites, 94,802 whites.

In the 45-54 age group there were 10,316 non-whites, 80,370 whites.

Table 3. MATRIX ANALYSIS FOR MALES (35-44) WHEN A $3000 EARNING LEVEL IS USED[a] (1960 age group)

	White			Non-white		
	L	$\bar{L}$	U	L	$\bar{L}$	U
A. Mover matrixes						
1957-58						
L	.62	.23	.15	.74	.13	.14
$\bar{L}$	.28	.69	.03	.40	.58	.02
U	.22	.06	.72	.28	.02	.70
1961-62						
L	.56	.26	.19	.68	.17	.14
$\bar{L}$	.13	.83	.03	.17	.81	.01
U	.21	.05	.74	.28	.02	.70
1965-66						
L	.52	.27	.21	.59	.26	.16
$\bar{L}$	.10	.87	.03	.13	.86	.01
U	.13	.04	.83	.19	.02	.78
B. Steady state distributions						
1957-58	.40	.35	.24	.56	.18	.26
1965-66	.19	.49	.32	.26	.51	.23
C. Total steady state distributions[b]						
1957-58	.22	.63	.16	.47	.30	.23
1965-66	.12	.70	.20	.26	.53	.21

[a] This group includes 94,802 whites and 12,757 non-whites. Almost 50 percent of the whites are movers, whereas 69 percent of non-whites are movers. Approximately 45 percent of whites were stayers in $\bar{L}$, only 18 percent of non-whites were in this category. Almost 8 percent of non-whites stayed in L while 1.6 percent of whites were similarly affected. The percent who stayed in uncovered were about the same for the two groups: 4 percent for whites and 5 percent for non-whites.

[b] This distribution includes both movers and stayers.

Table 4. REGRESSION OF TRANSITION PROBABILITIES ON GROWTH IN GNP

Probability	Constant	Coefficient of g	r^2	Durbin-Watson Statistic[a]
1. p_{11}^W	.63	1.1	.94	2.1
2. p_{11}^N	.76	1.5	.72	1.4
3. p_{12}^W	.20	.74	.53	1.0
4. p_{12}^N	.09	1.4	.80	1.3
5. p_{21}^W	.25	1.7	.63	1.1
6. p_{21}^N	.36	2.7	.70	1.2
7. p_{22}^W	.72	1.7	.63	1.2
8. p_{22}^N	.62	2.8	.71	1.2

[a]The Durbin-Watson statistic was calculated for each of these regressions. However, the sample size is so small that it is difficult to detect correlation of errors. See [1], p. 424.

Table 5. MATRIX ANALYSIS FOR MALES (35-44) WHEN A $3000 EARNING LEVEL IS USED (Uncovered category removed)

	White		Non-white	
	L	$\bar{L}$	L	$\bar{L}$
A. Mover matrixes				
1957-58				
L	.75	.25	.88	.12
$\bar{L}$	.36	.64	.45	.55
1961-62				
L	.74	.26	.85	.15
$\bar{L}$	.22	.78	.23	.77
1965-66				
L	.72	.28	.79	.21
$\bar{L}$	.20	.80	.18	.82
B. Steady state distributions				
1957-58	.59	.41	.79	.21
1965-66	.43	.57	.46	.53
C. Total steady state distributions				
1957-58	.34	.66	.69	.31
1965-66	.26	.74	.44	.56

Table 6. FEMALE STAYER PROPORTIONS, 1957-1966

Low earnings			Non-low earnings	
Non-white	White	Age group	None-white	White
$1500 earning level		(as of 1960)	$1500 earning level	
.05	.02	25-34	.34	.30
.11	.04	35-44	.45	.52
.18	.07	45-54	.48	.56
$3000 earning level			$3000 earning level	
.17	.08	25-34	.30	.27
.28	.16	35-44	.43	.49
.32	.20	45-54	.47	.54
$4500 earning level			$4500 earning level	
.29	.18	25-34	.28	.26
.41	.31	35-44	.33	.43
.44	.36	45-54	.32	.47

Table 7. FEMALE STAYERS AND MOVERS AS PROPORTIONS OF EACH AGE-RACE CATEGORY

Non-white stayers in					White stayers in			
L	$\bar{L}$	U	Non-white movers	Age group (1960)	L	$\bar{L}$	U	White movers
				(1500)				
.015	.054	.125	.806	25-34[a]	.004	.088	.137	.770
.037	.100	.109	.754	35-44[b]	.008	.163	.113	.716
.067	.114	.104	.715	45-54[c]	.015	.232	.099	.654
				(3000)				
.079	.010	.125	.786	25-34	.030	.036	.137	.796
.139	.025	.109	.727	35-44	.060	.072	.113	.754
.180	.025	.104	.691	45-54	.090	.106	.099	.705
				(4500)				
.140	.001	.125	.734	25-34	.089	.004	.137	.770
.222	.003	.109	.665	35-44	.156	.014	.113	.718
.264	.003	.104	.629	45-54	.215	.022	.099	.663

[a] This age group contained 12,161 non-whites, 103,049 whites.

[b] This age group contained 10,818 non-whites, 74,603 whites.

[c] This age group contained 7,509 non-whites, 59,253 whites.

Table 8. MATRIX ANALYSIS FOR FEMALES (35-44) WHEN A $1500 EARNING LEVEL IS USED (1960 age group)

	White			Non-white		
	L	$\bar{L}$	U	L	$\bar{L}$	U
A. Mover matrixes						
1957-58						
L	.58	.21	.21	.65	.14	.21
$\bar{L}$	.26	.69	.05	.30	.65	.05
U	.17	.03	.80	.20	.02	.77
1961-62						
L	.53	.25	.22	.63	.16	.21
$\bar{L}$	.13	.83	.04	.17	.80	.03
U	.22	.04	.73	.26	.02	.72
1965-66						
L	.51	.28	.20	.60	.21	.18
$\bar{L}$	.12	.85	.03	.14	.83	.02
U	.16	.04	.79	.21	.04	.75
B. Steady state distributions						
1957-58	.33	.27	.40	.40	.19	.41
1965-66	.22	.49	.29	.30	.44	.26
C. Total steady state distributions						
1957-58	.26	.30	.45	.34	.21	.46
1965-66	.17	.47	.36	.26	.41	.33

Table 9. REGRESSION OF TRANSITION PROBABILITIES ON GROWTH IN GNP

Probability	Constant	Coefficient of g	r^2	Durbin-Watson Statistic
1. p^W_{11}	.58	.63	.79	1.3
2. p^N_{11}	.64	.23	.16	1.4
3. p^W_{12}	.20	.75	.86	1.8
4. p^N_{12}	.13	.59	.51	1.8
5. p^W_{21}	.23	1.32	.55	1
6. p^N_{21}	.26	1.35	.43	1.1
7. p^W_{22}	.71	1.48	.54	1
8. p^N_{22}	.70	1.47	.40	1.1
9. p^W_{u1}	.19	.19	.04	1.5
10. p^N_{u1}	.22	.38	.19	1.7
11. p^W_{u2}	.03	.19	.57	1.9
12. p^N_{u2}	.02	.21	.57	1.2

FOOTNOTES

*Professor of Economics, University of California, Los Angeles.

The research herein was supported by the office of Economic Opportunity and the National Science Foundation (GS-2697). The opinions expressed herein are those of the author and should not be construed as representing opinions or policy of any agency of the U. S. Government.

The author is indebted to K. J. Arrow, B. Fox, A. Carlin, S. J. Carroll, J. Hirshleifer, T. Ozenne, A. Pascal, F. Sloan and the referee for their valuable comments and suggestions. The calculations reported herein are based on programs designed by K. Hall and T. Ozenne.

The author is a consultant to the RAND Corporation.

[1]These were subdivided into three age groups: 25-34, 35-44, and 45-54.

[2]See [8] and [23].

[3]See [1] and [2]. [3] is also germane. An individual is immune to growth if he remains in poverty during periods of sustained growth.

[4]In the period considered, the annual percentage change in GNP ranged from one to nine percent. Overall, the period was marked by considerable growth with relatively small changes in the price level.

[5]The distinguishing characteristics of "poverty" are discussed in [7] and [20].

[6]It should be noted that for all individuals in the ten year file, earnings must have been covered for at least one year.

[7]In measuring income dynamics in general it would have been preferable to have partitioned the income distribution into four categories: 0-1500, 1501-3000, 3001-4500, > 4500. However, once the dichotomous procedure was adopted, it was too costly to repartition. This will be done in a

subsequent study.

[8]See the pioneering work of Solow [21] for a thorough discussion of Markov models and their application to income distributions.

[9]For a discussion of the stayer-mover model see [4] and [9].

[10]The exact method used is described in [9], pp. 854-855. Because of the small size of n, there is an upward bias in the estimates of the stayer proportions.

[11]The standard statistical procedure was used to test these inequalities. See [5].

[12]Throughout this study it is assumed that constant values of g do not cause changes in the transition probabilities; that is, the process is stationary for fixed growth rates. Obviously this is a very crude approximation. Persistently high values of g will undoubtedly influence these transition probabilities. Appropriate modifications will be made in a subsequent study.

[13]A more complete discussion of this hypothesis is contained in [17].

[14]For a complete description of these data, see [18], [22], [26], and [27].

[15]Individuals who are disabled and become eligible for Social Security benefits cannot be identified by these Social Security data, i.e., they are treated as uncovered.

[16]See Appendix A of [18] for the scope and incidence of these changes.

[17]The Goodman test described in the previous section was applied to these data. The null hypothesis of equality with zero was rejected in every case at significance levels above .01.

[18]The null hypothesis of equality was rejected in every case at

significance levels beyond .025.

[19]For these differences the significance level was always greater than .001.

[20]It is important for policy to distinguish between these two probabilities. The conditional probabilities of staying in L or $\bar{L}$ could be quite high, while the unconditional probabilities of staying in L for all ten years could be quite small. It is also necessary to calculate the unconditional probabilities in order to obtain the mover proportions. It should also be remembered that the estimates of the conditional probabilities are biased upward.

[21]This analysis was also conducted for a current age group, i.e., the age intervals--25-34, 35-44, 45-54--were held fixed over time. The 1960 age grouping confounds the effects of aging and of transitions as of a fixed age. The current age analysis is better designed to hold the effects of aging constant. However, differences between these analyses were negligible. Hence the current age results are not presented.

[22]For a more complete description of the properties of Markov chains see [14] and references listed there.

[23]Statistical tests verified these casual observations. Using the χ^2 tests developed by Goodman [10], the hypothesis of stationarity could be rejected at very high levels of significance.

[24]In making the distinction between stayers and movers only those who remained in L ($\bar{L}$) for the entire ten year period were movers. Hence, for any two successive years the mover matrix can have transitions from L ($\bar{L}$) in year t to L ($\bar{L}$) in year t + 1.

[25]Three alternative explanations of these differences are (1) the South to North and rural to urban migration rates were higher for non-whites than whites, (2) improvements in the quality of education were greater for non-whites than whites, and (3) civil rights legislation and the subsequent reduction in job discrimination caused non-whites to improve faster than whites. None of these alternative explanations is investigated here.

[26]The method assumes that emergence from uncovered gives information concerning the income distribution of the uncovered category. The procedure may be biased, but given the paucity of information regarding the uncovered category, the bias cannot be measured. All those who were ever uncovered, are covered for at least one of the ten periods. Their income during this covered period is our measure of uncovered behavior.

[27]See [6] and [19].

[28]Using the χ^2 tests developed by Goodman [10], the hypothesis of stationarity could be rejected at very high levels of significance.

REFERENCES

[1] Aaron, Henry, "Foundations of the 'War on Poverty' Reexamined," American Economics Review, December 1967, pp. 1229-1240.

[2] Anderson, W. H. Locke, "Trickling Down: The Relationship Between Economic Growth and the Extent of Poverty Among American Families," Quarterly Journal of Economics, November 1964, pp. 511-524.

[3] Batchelder, Alan B., "Decline in the Relative Income of Negro Man," Quarterly Journal of Economics, November 1963, pp. 525-548.

[4] Blumen, Isadore, Marvin Kogan, and P. J. McCarthy, The Industrial Mobility of Labor as a Probability Process, Cornell University Press, Ithaca, 1955.

[5] Brownlee, K. A., Statistical Theory and Methodology, Wiley, New York, 1965.

[6] Cain, Glen G., Married Women in the Labor Force, The University of Chicago Press, Chicago, 1966.

[7] David, Martin, "Welfare, Income and Budget Needs," Rev. of Econ. and Stat., Vol. 40, Nov. 1959, pp. 393-399.

[8] Gallaway, Lowell, "The Foundations of the War on Poverty," American Economic Review, March 1965, pp. 122-131.

[9] Goodman, L. A., "Statistical Methods for the Mover-Stayer," Journal of American Statistical Association, December 1961, pp. 841-868.

[10] _____, "Statistical Methods for Analyzing Processes of Change," American Journal of Sociology, July 1962, pp. 57-58.

[11] Kosters, Marvin and Finis Welch, "The Distributional Incidence of Cyclical Fluctuations and the Minimum Wage," The RAND Corporation, RM-6273-OEO (forthcoming).

[12] Malinvaud, E., Statistical Methods of Econometrics, Rand-McNally and Company, Chicago, Illinois, 1966.

[13] McCall, John J. and Neil Wallace, "A Supply Function of First-Term Re-enlistees to the Air Force," The Journal of Human Resources, Summer 1969, pp. 293-310.

[14] McCall, John J., "An Analysis of Poverty: A Suggested Methodology," Journal of Business, January 1970, pp. 31-43.

[15] _____, "An Analysis of Poverty: Some Preliminary Findings," Journal of Business, April 1971.

[16] _____, "Economics of Information and Job Search," Quarterly Journal of Economics, February 1970, pp. 113-126.

[17] _____, "Information, Search, and Racial Discrimination," in The American Economy in Black and White: Essays on Race Discrimination in Economic Life, A. Pascal, editor (forthcoming).

[18] _____, "Earnings Mobility and Economic Growth," The RAND Corporation, R-576-OEO, October 1970.

[19] Mincer, Jacob, "Labor Force Participation of Married Women," in Aspects of Labor Economics, NBER, Princeton University Press, 1962.

[20] Orshansky, Mollie, "Counting the Poor: Another Look at the Poverty Profile," Social Security Bulletin, January 1965, pp. 3-29.

[21] Solow, Robert M., On the Dynamics of the Income Distribution, Unpublished Doctoral Dissertation, Harvard, 1951.

[22] Steinberg, Joseph and C. H. Cooper, "Social Security Statistical Data, Social Science Research and Confidentiality," Social Security Bulletin, October 1967, pp. 3-15.

[23] Tobin, James, "On Improving the Economic Status of the Negro,"

Daedalus, Fall, 1965, pp. 878-898.

[24] Wohlstetter, Albert and Sinclair Coleman, "Race Differences in Income," The RAND Corporation, R-578-OEO, November 1970.

[25] Zellner, Arnold and Tong Hun Lee, "Joint Estimation of Relationships Involving Discrete Random Variables," Econometrica, April 1963, pp. 382-393.

[26] Workers Under Social Security 1960, U. S. Department of Health, Education and Welfare, Social Security Administration.

[27] Longitudinal Employer-Employee Data, U. S. Department of Health, Education and Welfare, Social Security Administration, April 1970.