Intragenerational Social Mobility as a Markov Process: Including a Time- Stationary Markovian Model that Explains Observed Declines in Mobility Rates Over Time

Author(s): David D. McFarland

Source: *American Sociological Review*, Vol. 35, No. 3 (Jun., 1970), pp. 463-476
Published by: American Sociological Association
Stable URL: https://www.jstor.org/stable/2092989

Trow, Martin
1958 "Some implications of the social origins of engineers." Scientific Manpower. Washington, D.C.: National Science Foundation.
White, R. Clyde
1952 These Will Go To College. Cleveland: Western Reserve University Press.
Wilson, Alan B.
1959 "Residential segregation of social classes and aspirations of high school boys." American Sociological Review 24 (December):836–845.
Winer, Benjamin J.
1962 Statistical Principles in Experimental Design. New York: McGraw-Hill.

INTRAGENERATIONAL SOCIAL MOBILITY AS A MARKOV PROCESS: INCLUDING A TIME-STATIONARY MARKOVIAN MODEL THAT EXPLAINS OBSERVED DECLINES IN MOBILITY RATES OVER TIME *

DAVID D. MCFARLAND

University of Chicago

Previous Markovian models of intragenerational social mobility are examined, with emphasis on the substantive meaning of the mathematical assumptions employed, and on some qualitative consequences of particular combinations of assumptions. A new model is proposed which is a Markov chain at the level of any given individual but which involves a heterogeneous population. The corresponding population level process is not a Markov chain, differing at precisely those points where the failure of the Markov chain model to fit mobility data has been noted. This model is compared and contrasted with several others, notably those incorporating nonstationarity, and suggestions are given for future research.

INTRODUCTION: THE SIMPLE MARKOV CHAIN MODEL

THE Markov chain (Kemeny and Snell, 1960, 1962; Feller, 1968) has been proposed as a model for both intergenerational and intragenerational social mobility (Prais, 1955a, 1955b; Blumen, *et al.* 1955), as well as for numerous other social processes. A Markov chain is characterized by a number of "states," in which the process might be, and the matrix of probabilities of transitions between the various states in a single (fixed) unit of time, such as a year. The process is in exactly one state at any given time. In what will be called the *simple* Markov chain model of social mobility, the states correspond to occupational or other social status categories. The Markov chain assumptions are: (1) Stationarity. The transition probabilities remain constant over time. (2) Markovian. The probability of transition to a given state during the next time unit depends only on the current state of the process, and not on its history of previous moves from state to state.

Traditional Markov chain theory pertains to a single object moving from state to state; but in applications to social mobility, one considers an entire population, each person moving probabilistically from state to state. In this context a third assumption is made, though not always explicitly: (3) Homogeneity of population.[1] The various members of the population are subject to identical sets of transition probabilities. This assumption sometimes enters unsuspected. When one writes of *the* probability of some particular transition, rather than

* The research reported herein was carried out at the Population Studies Center, The University of Michigan, with the support of a fellowship from The Population Council. This paper has greatly benefited from the comments of the following: J. S. Coleman, D. L. Featherman, L. A. Goodman, N. Henry, K. C. Land, J. H. Levine, J. Matras, T. F. Mayer, R. McGinnis, T. W. Pullum, P. M. Siegel, L. O. Stone, and H. C. White. Even where their suggestions were not followed, they often led to a clearer presentation of the argument, and were appreciated.

[1] It should be noted that the term "homogeneous" is also used in the literature of Markov processes in reference to time homogeneity, i.e., what is called stationarity herein.

the *average* probabilty of that transition, he has implicitly assumed that the transition in question has the *same* probability for different persons, i.e., that the population is homogeneous. It is this assumption which permits one to use the *proportion* of persons making a particular transition as an estimate of the corresponding transition probability to which any *particular* person is subjected.

Limitations of Markov Chain Model

Duncan (1966) argues, and quite correctly, that the change in social structure over the "length of a generation" (actually, the average age of fathers at the births of their various sons, or about 30 years) cannot be faithfully represented by the transition matrix from respondent's father to respondent, from which it follows that intergenerational social mobility cannot be faithfully represented as a simple Markov chain. The reasoning is as follows: The father-to-son transition matrix does tranform *something* into the vector of men in the current working force; but that something that it transforms is *not* the vector of men in the working force about 30 years earlier; rather, it is the vector of fathers of men in the current working force. Some of these fathers were in the working force in the distant past, and are now dead; others are still in the current working force along with their sons. Furthermore, men who were in the working force a generation ago but who had no surviving sons are not represented at all; those with two sons are represented twice, etc. This argument shows that the notion of treating intergenerational social mobility as a simple Markov chain involves some serious difficulties. However, the logical difficulties Duncan points out do not carry over to intragenerational mobility.

Blumen, *et al.* (1955) and, more recently, Hodge (1966) argue against the simple Markov chain formulation of intragenerational social mobility on empirical grounds.[2] The Markov chain assumptions imply that the probability of remaining in a state for two successive time periods is equal to the square of the probability of remaining in the same state for a single time period, but empirical estimation—based on the homogeneity assumption, that the proportion of persons making a move is an appropriate estimate of the probability to which any one of them was subjected—yields larger values for the former than for the latter. More generally, the Markov chain assumptions imply that the k-step transition matrix —the matrix of transition probabilities during a period of k time units—is equal to the k-th power of the one-step transition matrix, which is empirically false. This argument against the Markov chain formulation is not conclusive, however, because of "lumpability" considerations (Kemeny & Snell, 1960: 123ff., 197ff.). If a stochastic process which *is* a Markov chain is modified by combining ("lumping") two or more states into a single state, the resulting "lumped" stochastic process will not generally be a Markov chain. So what these critics have really shown is that intragenerational social mobility is not a Markov chain *when states are defined the way they defined them;* the process still might be a Markov chain if the states were defined differently.

This alternative explanation of their results is nearly impossible to test, however. Occupations could be classified into states for a stochastic process in a large variety of ways, and searching for one which would make the process a Markov chain (even if such a classification into states existed) is somewhat like looking for the proverbial needle in a haystack. But in addition to this problem, there is a practical estimation problem: if very many states are used, the data will be insufficient to obtain reliable estimates of the transition probabilities.

Still, as recently as 1967, numerous authors were using the simple Markov chain model of social mobility (Bartholomew, 1967; Bartos, 1967; Beshers and Laumann, 1967; Lieberson and Fuguitt, 1967; Matras, 1967 [3]), although the model ap-

[2] Actually, Hodge's evaluation of the simple Markov chain model is somewhat equivocal. He does point out discrepancies between observed mobility and that predicted by the model, as discussed in the text. But he also concludes that the fit of the model, though not perfect, is sufficiently good to suggest that the concept of a "career," which is antithetical to the Markovian assumption, is of limited utility.

[3] In his latest paper, Matras (1967) carefully avoids the term "Markov chain," carrying out precisely the same matrix operations under the term

pears to have lost favor more recently. But those who use it as a model for intergenerational mobility must face the logical difficulties pointed out by Dnucan; and the burden of proof rests with those who wish to characterize intragenerational mobility as a simple Markov chain, since none of them has yet found the correct way of defining the states.

PREVIOUS MODIFICATIONS OF THE SIMPLE MARKOV CHAIN MODEL

It is possible to modify the simple Markov chain model by replacing one or more of its assumptions with alternative assumptions. Blumen *et al.* (1955) modified the homogeneity assumption. They postulated the population to consist of two types of persons, "movers" and "stayers," the former being subject to the usual type of transition probability matrix, and the latter having zero probability of movement to a different state. This modification of the model improved the fit to their data. (For further results on the Mover-Stayer Model, see Goodman, 1961; Bartholomew, 1967.)

Mayer (1967, 1968a) modifies the simple Markov chain model by relaxing the stationarity assumption.[4] Instead, he considers "uniform" nonstationary models, in which it is assumed that mobility rates decrease with age. The nonstationarity is uniform in the sense that the mobility rates for different transitions in social status decline according to the *same* function of age. Homogeneity is retained in a modified form: any two persons have identical sets of transition probabilities, provided they are in the same age group.

The data considered cover only a single cohort. Thus, as Mayer carefully acknowledges, he is unable to distinguish between time trends in mobility patterns on the one hand and changes in mobility rates with age on the other. Furthermore, for lack of appropriate data, the model was tested on a "synthetic cohort" rather than a real cohort. A synthetic cohort is constructed from cross-sectional data; this hypothetical cohort is assumed to be subject to the mobility rates of current 25-year-olds when age of cohort was 25, subject to the rates of current 35-years-olds when age of cohort was 35, etc. For these two reasons (plus the use of estimation procedures which may not be optimal), the adequacy of this type of model is difficult to appraise from Mayer's results.

A third modification of the simple Markov chain model, and one which has recently received considerable attention, is the so-called Cornell Mobility Model, a general purpose model intended not only for changes in occupation or social status, but also for changes in such other characteristics as industry of employment, region of residence, etc. (McGinnis, 1968; Henry, *et al.*, 1968; for applications to geographic mobility, see Myers, *et al.*, 1967; Morrison, 1967; Land, 1969).

The Cornell Mobility Model is a rather drastic modification of the simple Markov chain model, but a reformulation of what is meant by "state" brings it back into the Markov chain framework. In this model a phenomenon of "cumulative inertia" is postulated: the longer one has been in his current status (or industry, residence, etc.), the higher his probability of remaining there for yet another time unit.

This model gives rise to *apparent* violations of all three assumptions of the simple Markov chain, but the violations are only apparent, for a reformulation makes it into a Markov chain again. Whenever a person remains in a single status more than one time unit, his transition probabilities will change, an apparent violation of stationarity. (Note the use of the word "status" rather than "state." In the simple Markov chain the two are synonymous, but, in the present model, "state" has a different meaning, which will be specified below.) Second, the probability of a person's moving to a given status *does* depend on his history of previous moves (though only through the number of time units he has been in his present status), an apparent violation of the Markovian assumption. And different persons in the same status have different sets of transition probabilities because they

"The Linear Model." But for proofs of assertions he cites his earlier papers (1960, 1961), without actually retracting the Markov chain assumptions made therein. Both Matras and Bartholomew go on to consider more complex and more realistic models.

[4] Mayer's model also differs from the others discussed in treating time as continuous.

have been in their current status different lengths of time, an apparent violation of the assumption of a homogeneous population.

Both the Cornell Mobility Model and the Mayer model involve common-sense means of modifying the simple Markov chain to achieve a model which *will* explain the observed declines in mobility rates over time: if the proportion of people moving declines over time, then there is considerable appeal to the notion that, in one way or another, the *probability* of moving declines over time. Two ways in which this can occur are for the probability of each particular move to be a decreasing function of age (Mayer's approach), or for it to be a decreasing function of duration of stay in one's current occupational category (McGinnis' approach). However, we will show in the next section that this common-sense explanation—that the probability of movement declines—is by no means a necessary conclusion from empirical data: the observed declines can also be explained by a model in which each person's set of transition probabilities is constant over time.

McGinnis' justifications for modifying the simple Markov chain model in this particular manner differ somewhat from the common-sense justification given below: "It was begun with the observation that people are not necessarily homogeneous in their tendencies to be mobile even though they may be in a common location at a particular time. A number of sources made it seem equally plausible that movement out of a status position (or any other social location) is constrained chiefly by one's ties to that position. Moreover, the strength of these ties normally should be expected to grow with the passage of time" (McGinnis, 1968:716).

But one quickly runs into some serious estimation problems if he attempts to apply the Cornell Mobility Model directly to numerical data, as its authors apparently intend to do. A few details of the model will show why.

First of all, the process is reformulated from the way we described it above to get back into the Markov chain framework: in the reformulated process, a person's current "state" is defined to consist of a status-duration pair, telling *both* the person's current social status, and how long he has held that status. Then it is assumed that the reformulated process (though not the simpler process where states are identical to statuses) is a Markov chain, i.e., that it satisfies the assumptions with which this paper began.

The assumption that the reformulated process is stationary is equivalent to the assumption that the probability of a man's moving between two statuses (not states) varies with time *only* as a function of the duration of his stay in his current status. The assumption that the reformulated process is Markovian is equivalent to the assumption that one's probability of moving between two statuses depends on his history of previous moves *only* through the duration of stay in his current status, and hence that one's current "state" in the reformulated process contains all the history which is relevant. The assumption that the reformulated process is operating in a homogeneous population is equivalent to the assumption that different persons in the same status have different probabilities of making certain transitions between statuses *only* because they had been in their current status for differing durations.

Since this model neglects the basic demographic facts of birth and deaths, it provides no upper bound for duration in a particular status. Thus the model, in its original form, literally has an infinite number of unknown parameters to be estimated. To reduce the number of unknowns, the authors of the Cornell Mobility Model have taken two steps:

(1) In a modified version of the model (Henry *et al.*, 1968), only a finite number of durations are considered, so that the process has a finite number of states, rather than a denumerably infinite number of states. This makes it possible, at least in principle, to estimate the unknown parameters. If, to take some quite moderate figures for a numerical example, seven social statuses and ten durations are distinguished (and it is assumed that after one has remained in a given status for ten time units his transition probabilities remain constant), there would be $10 \times 7 \times (7-1) = 420$ un-unknown parameters to be estimated. This

is in marked contrast to the $7 \times (7-1) = 42$ unknown parameters to be estimated in the corresponding simple Markov chain model.[5]

(2) To further reduce the number of unknown parameters to be estimated, an additional assumption is made: that among those who do move, the status to which one moves is independent of the duration of his current status; i.e., the conditional probabilities of the various possible moves, given one's current state *and* that he will move, depend only on his current social status and not on the duration of his stay in that status. In the numerical example above, this reduces the number of independent quantities to be estimated to $10 \times 7 = 70$ probabilities of remaining in the same status, plus $7 \times (7-2) = 35$ conditional probabilities of various changes in status, given that one will change status, for a total of 105 independent estimates required.[6] This assumption appears to be strictly *ad hoc,* with no empirical basis, and there is no reason for us to take it any more seriously than it is taken by its authors (Henry *et al.,* 1968: Footnote 6). But some such assumption is required, in order to reduce the number of unknowns, if the model is actually to be estimated from real data. None of the published applications or tests of the model has produced anything even approaching a complete set of parameter estimates.

A TIME-STATIONARY MARKOVIAN MODEL

The purpose of this section is to point out that the observed decline in mobility *rates* over time does not require an explanation involving a corresponding reduction in mobility *probabilities,* as in models where mobility probabilities depend on age, on duration of current status, or on time directly. This reduction in mobility rates can also be explained by heterogeneity of a population in a model where each person's transition probabilities are constant over time.[7]

This should not be construed as a claim that this is "The Correct Model," but merely that this seems to be an alternative model which has not been adequately considered by the authors working in this area, and which is not contradicted by the data from which they have inferred that mobility probabilities decrease over time. And, as with the models considered above, its substantive content has a certain degree of plausibility.

Consider a population categorized into a finite number of social or occupational statuses, the latter denoted by positive integers, in an arbitrary order. During a particular time unit, any given person, say m, is characterized by his matrix of transition probabilities,

$$P(m) = \begin{bmatrix} P_{11}(m) & P_{12}(m) & \cdots \\ P_{21}(m) & P_{22}(m) & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix} \quad (1)$$

We employ $P_{ij}(m)$ to denote the conditional probability of person m's going from state i to state j during the time unit in question, given he is in state i at the beginning of the time unit. (A more general notation would be $P_{ij}(m, t)$, where t denotes a particular time unit; but we shall presently assume that the probability does not depend on time.) The stochastic process for a particular person would be completely specified by giving his initial state at time 0 and the transition matrices to which he will be subject in each of the subsequent time periods.

[5] The factor of 6, rather than 7, arises in both calculations because the 7 conditional probabilities in each set necessarily sum to unity, leaving only 6 independent quantities to be estimated in each set.

[6] The latter 35 parameters are not transition probabilities of the process, but the transition probabilities can be calculated from them together with the former 70 parameters, which are transition probabilities lying on the diagonal of the matrix.

[7] A note on priority: The circulation of preliminary drafts of this paper in 1968 brought several responses to the effect that the idea of a heterogeneous population was not original. McGinnis attributed the idea to a 1968 discussion with J. S. Coleman, which reminded me of an earlier discussion with Coleman which may well have been the stimulus to my own work on this topic (see McGinnis, 1968:721). T. F. Mayer responded by showing me an unpublished paper (1968b) which I had not seen previously, in which he considers some closely related matters. But L. A. Goodman and H. C. White both pointed out that the idea of a heterogeneous population is much older, having been mentioned briefly by Blumen *et al.* (1955: 153–154). Nevertheless, this article is apparently the first to systematically consider heterogeneity in some generality (as opposed to heterogeneity via a small number of linearly ordered and internally homogeneous classes), and in some detail.

We make the following assumptions: (1) Stationarity. For any given person, m, these probabilities are constant over time.[8] (2) Markovian. The process, for any given person, is Markovian; that person's probability of moving depends only on his current social status, and not on his history of previous moves. (3) Heterogeneity. Different persons will not necessarily have identical transition probability matrices.

The stationarity assumption may be a close approximation to reality over a period of several years, as in the study by Blumen *et al.* (1955); but we seriously doubt its accuracy if applied over an entire lifetime. Its use here will permit a contrast between two extreme types of models: stationarity without heterogeneity, and heterogeneity without stationarity.

We appreciate the sociological insights (quoted above) which led McGinnis and his colleagues to their model. In fact, the heterogeneity we assume was one of them; but the Cornell Mobility Model incorporates only one of many possible types of heterogeneity —that due to different durations of stay in one's current status. The second sociological insight behind the Cornell Mobility Model—that attachments grow over time—undoubtedly has some truth to it, but is probably much more relevant to *geographic* mobility than to occupational mobility. (In fact, it was geographic mobility which was considered in the few attempts cited above to apply the ideas behind the model.) With occupational mobility there is a countervailing tendency for the propensity to move to *increase* with the duration of one's current status.

During his tenure in an occupation, a person often accumulates additional education, experience, and maturity which he did not have on entering this occupation, and which both make him dissatisfied with his current occupation and make him attractive to those who control entrance to other occupations with higher status. In fact, the norms of our society favor continuing occupational achievement throughout one's lifetime. Thus, while spending some time as a factory worker, one may be developing both the interest and the qualifications to move to a position in inspection, supervision, etc. Similar considerations apply to the engineer or accountant moving into sales or management, the college professor moving into administration, the employee of a small business setting up his own firm, and the person in practically any occupation becoming a politician. McGinnis (1968: Footnote 8) admits a similar point.

On the basis of these sociological considerations, we do not find McGinnis' basis for the "cumulative inertia" assumption especially compelling. We do, however, feel compelled by the biological, psychological, and sociological consideration that people do differ in their qualifications for different roles; that society must at least make some approximation to placing people in roles for which they are qualified; and that, therefore, people differ in their probabilities of moving between roles (and between the corresponding statuses). Thus the heterogeneity of the population is a fundamental part of this model—and it includes heterogeneity arising from sources other than differing durations of stay in current statuses, such as intelligence and other psychological traits, type of family background, ethnicity, and educational attainment.

In the other models discussed, the proportion of a group of people making a particular move is taken as an estimate of the probability that any particular member of that group will do so. It should be empha-

[8] The assumption that transition probabilities are constant should not be confused with an assumption to the effect that the social structure, as indicated by the proportional distribution of the population into the various categories, remains constant. In fact, except in very special situations, the stationary model implies that the social structure will change over time.

To each particular transition matrix there correspond zero, one, or several stationary distributions, i.e., distributions which are unchanged by the continued operation of the process. In the case of regularity, discussed below, a stationary distribution exists and is unique. One special case where stationarity of transition probabilities does *not* imply changing social structure is the case where both the following conditions are met: (1) each of the transition matrices applying to the various subsets of the population possesses at least one stationary distribution; and (2) the initial distribution of each population subset happens to be identical to one of the stationary distributions for that subset's matrix. The other special case is where the nonstationarity of one subset exactly cancels the nonstationarity of other subsets during each time interval, so that the social structure of the entire population remains constant despite micro-level changes.

sized that such is not the case in this model: in fact, in this model there is no such thing as *the* probability of making a particular move—each person may have a *different* probability of making it.

One attribute of Markov chain models not mentioned previously is the "ergodic" property of the particular type of chains commonly used as models, namely "regular" Markov chains. These Markov chains have the property that, after a sufficient passage of time, the probability of being in any particular state depends only on the various transition probabilities, and not on the initial state. Furthermore, this long-run distribution of the probabilities of being in the various states is stable—it is unchanged by further operation of the process. So those who use a regular Markov chain as a model can determine the (expected) distribution of persons in the various states after the process had stabilized, possibly comparing that with the current distribution to obtain an indication of whether the process had already operated a long time under present conditions.

We shall now show that the process defined in our model also has such a long-run stable distribution, but with important differences from that of Markov chains which operate on homogeneous populations. As previously, let $P(m)$ denote the matrix of transition probabilities for person m. We want to combine these transition matrices for the various persons in such a manner as to yield the expected value of the one-step transition matrix for the entire population, i.e., a matrix Q_1 whose $i-j$ element is the expected proportion of the people in category i at time 0 who are in category j at time 1. To do this, we define $N_0(m)$ as a diagonal matrix with an entry of unity in the diagonal position corresponding to person m's initial state and zeros elsewhere; and let $N_0 = \sum_m N_0(m)$ be the diagonal matrix whose $j-j$ element is the number of people initially in category j. Then N_0^{-1}, the inverse of N_0, is a diagonal matrix whose $j-j$ element is the reciprocal of the corresponding element of the matrix N_0. Using this notation, the expected one-step population transition matrix is

$$Q_1 = N_0^{-1} \sum_m N_0(m)P(m). \qquad (2)$$

Similarly, the expected k-step population transition matrix is

$$Q_k = N_0^{-1} \sum_m N_0(m)[P(m)]^k. \qquad (3)$$

Next we make one additional assumption: (4) Regularity. No person has any transition probability precisely equal to zero. There exists some small positive number, ϵ, such that for each person m and for each pair of states x and y, $P_{xy}(m) \geqq \epsilon$.

Actually, regularity could be achieved with a much weaker assumption,[9] but that does not appear to have any particular advantages in this case. Even in the strong form stated here, the regularity assumption does no injustice to the empirical facts: if we make epsilon sufficiently small, there is no means of reliably inferring within the duration of any one mobility study (or the lifetime of a given man) whether one of a man's probabilities is actually zero, rather than epsilon or greater.

Successively higher powers of the matrix $P(m)$, the transition matrix for the regular Markov chain to which person m is subject, converge to a matrix $P^*(m)$, say, whose rows are identical. Thus the probability that person m is, after a sufficient length of time, in any particular state depends only on person m's transition probabilities, and not on his initial state. Furthermore, these long-run probabilities are stable, remaining constant under continued operation of the process.

Here is apparent one major advantage of a model which allows for a heterogeneous population. In a model where everyone has the same transition probabilities, an immediate corollary of the argument in the preceding paragraph is any statement of the form, "Everyone, regardless of origin, has the same probability of eventually being .... (Insert *President,* or some other desirable status.)." Such statements, which can be derived from the models with homogeneous populations, sound like myths serv-

[9] The usual form of the regularity assumption is as follows: There exists a finite number, k, such that no k-step transition probability is zero. This is implied by our strong form of regularity, in which $k=1$.

ing as "opium of the people," and not at all like realistic appraisals of the situation. Any realistic model must certainly allow for *unequal* probabilities of eventually reaching highly valued statuses—and with the inequalities depending on many more variables than age or duration of one's current status.

Continuing the derivation of long-run properties of our model: If we substitute $P^*(m)$, the matrix of long-run transition probabilities for person m, into Equation (3) in the limit as k becomes sufficiently large, we find the long-run value of the k-step population transition matrix as k increases:

$$Q^* = N_0^{-1} \sum_m N_0(m) P^*(m). \qquad (4)$$

Since the long-run individual matrices $P^*(m)$ are constant, so is the long-run population transition matrix, Q^*. Hence, if this process continues sufficiently long, at the population level there will result a stable distribution of persons among the various states, as in the simple Markov chain model.

But the set of Markov chains, like many other sets, is not closed under the operation of taking averages. Although the population level process is a weighted average of individual level Markov chains, the population level process is not itself a Markov chain. There are two important differences:

(1) The long-run population level transition matrix, Q^*, unlike the long-run individual level transition matrices, $P^*(m)$, does not necessarily have identical rows, since the premultiplication of $P^*(m)$ by $N_0(m)$ in Equation (4) incorporates only a single row of $P^*(m)$—the row corresponding to person m's initial state—into the matrix Q^*. Hence the proportion of persons originating in a given initial state who are in a given state after stability is reached may differ for different initial states. This fact can be explained heuristically: An individual's probability of being in a particular state after stability is reached does not depend directly on his initial state, but *does* depend on his matrix of transition probabilities, and the latter may be distributed unequally among the various origin states. In our heterogeneous model it would be only by coincidence that more than a few persons have identical probabilities of eventually reaching some highly desirable status. In this respect the heterogeneous model, unlike the others discussed, corresponds to what we take to be a realistic view, and not to the wishful thinking that "My Johnny has as good a chance of becoming President as any other boy does."

(2) The matrix Q_k, unlike the k-step transition matrix of a Markov chain, is *not* generally equal to the k-th power of the one-step transition matrix, Q_1. So this process differs from the simple Markov chain at precisely the point where the latter's disagreement with the data inspired other models discussed previously. This discrepancy between Q_k and $(Q_1)^k$ can easily be seen from Equations (2) and (3): Q_k is a weighted average of k-th powers of the matrices $P(m)$, while $(Q_1)^k$ is the k-th power of the weighted average of the $P(m)$ matrices themselves.[10] These discrepancies permit the heterogeneous model to fit the data where the simple Markov chain failed.

The process at work in this model can be described heuristically, at least to the point of explaining why the proportion of people staying in a given status from time 0 to time 2 is *not* equal to the square of the proportion staying from time 0 to time 1; or, equivalently, why the proportion of those still in the given status at time 1, who remain from time 1 to time 2, is not equal to the proportion of those in the given status at time 0 who remain from time 0 to time 1. The expected proportion of persons in a given status who make any particular transition is equal, by Equation (2), to the average of their various probabilities of doing so. Now those who leave the given status during the first time interval differ from those who remain, in that the latter tend to be persons with higher probabilities of staying than the former. Thus the group remaining after one time period will have a higher average probability of staying than did the initial group; and hence the expected proportion of the former group remaining throughout the second time period is larger than the expected proportion of the initial group remaining throughout the first time period.

[10] An analogous distinction arises in elementary statistics: the average of the square of a variable and the square of the average of the same variable are not identical; in fact, their difference is the variance of the variable.

Thus this model gives rise to empirical consequences which appear, at first glance, as if the probability of movement declines over time. But in fact the probabilities do not decline; it is just that those with high probabilities of movement tend to move early, and those still remaining after several time periods are predominantly persons with low probabilities of movement. This model and the Cornell Mobility Model have causation going in opposite directions. In the Cornell model one's probability of movement becomes low *because* of his long duration in his current status; in this model one has a long duration in his current status *because* his probability of movement is (and always was) low.

In a certain sense the model in this section is a generalization of the Mover-Stayer Model of Blumen *et al.* (1955): their model permits a heterogeneous population only to the extent of having two classes of men, each class being homogeneous; the model given herein has an arbitrarily heterogeneous population.[11] Mayer (1968b) has done some preliminary work toward an intermediate model—one with a moderate number of different classes of men, each class being homogeneous; his paper provides a partial answer to the question of how many classes would be necessary to explain observed mobility data.

Our model provides the possibility that one can be a high probability mover until he gets into the "right" job for him, and thereupon have a negligible probability of moving. This career pattern, which appears to be quite common, is not possible in the Mover-Stayer Model, in Mayer's (1968b) generalization or in the Cornell Mobility Model. Another way of allowing this type of career pattern, in a different type of model, was proposed by Mayer (1968a). The latter model has two states—one regular and one absorbing—corresponding to each social status, and each time a person is in a given social status he has a positive probability of being absorbed permanently therein. In this model, the cumulative probability of absorption in a given status increases with the duration of stay therein, producing empirical consequences similar to those of the Cornell Mobility Model but by a somewhat different mechanism. The stationarity assumption of our model, together with its lack of absorbing states, means that we attribute the apparent obsorption of some persons to a different cause—that their probability of movement, though constant, is very low.

We conclude this section with a *caveat* concerning the long-run and stable properties of this model, at both individual and population levels. (The same observations, of course, will apply to other models as well.) Nothing we have done above gives any indication of the speed with which stability is approached.[12] But the empirical process of social mobility, unlike the mathematical model, can not continue indefinitely until stability is reached. If stabilization requires, say, 100 years, all the men will die before it is accomplished, and the mathematical theorems about stability will be irrelevant to the empirical phenomenon. And even if stabilization requires, say, only 20 years, the stationarity assumption will become highly suspect. However, we will see in the next section that this is not so serious—the main result we derived is still valid without the stationarity assumption.

FURTHER DISCUSSION OF THE MODEL

This paper contains no computations purporting to "apply" the model of the preceding section to actual mobility data. In part, this is because the model has not yet been sufficiently specified to make parameter estimation possible—it, like the Cornell Mobil-

[11] In a stricter sense, our model is not a generalization of the Mover-Stayer Model: one of our assumptions—that no person has any transition probability precisely equal to zero—directly contradicts one of theirs, where a Stayer has zero probability of changing statuses. However, see the remark in the text immediately following the regularity assumption.

[12] The Blumen *et al.* (1955:59, Table 4.3) transition matrix, using ten categories of employment by industry plus an eleventh category for unemployment, and based on a three month time interval, is stable to several decimal places by the thirty-second power, or after only eight years. In the remainder of the paper we tentatively assume that other transition matrices would have similar convergence rates. However, a different classification of states (e.g., by prestige of occupation, rather than by industry) could conceivably result in a quite different rate of stabilization. This question needs further empirical study.

ity Model, currently involves too many unknown parameters. But the lack of a section treating actual mobility data in this paper is primarily in resistance to a common tendency to turn to numerical data *prematurely*. A great amount of energy and paper has been wasted attempting to "apply" various inadequate models to data, when the models' inadequacy could more easily have been discovered—and perhaps remedied—by a careful theoretical analysis of the models' assumptions and/or their logical consequences.

This model seems inadequate for several reasons. One apparent inadequacy is the stationarity assumption. There is no way of testing a single proposition of this nature which does not refer directly to observables (a test of one assumption depends on the validity of the model's other assumptions); yet it runs counter to our intuitive impression of the process in question. In particular, it seems much more difficult for men over 40, except in politics and administration, to obtain new jobs. And (by the same logic we applied previously to someone else's assumption) we can hardly expect the reader to take the stationarity assumption any more seriously than we do. In the previous section, it served the function of showing that even when there is no nonstationarity whatever, heterogeneity will give the apparence of nonstationarity. A second reason for using it there was its simplicity: if one does not assume stationarity, then he must consider the infinitely many different ways the probabilities could conceivably change over time.

In the preceding section the *stationary* regular Markov model, operating on a homogeneous population, was shown to imply the implausible conclusion that everyone, regardless of origin, has the same long-run probability of reaching a given desired status. Would a lack of stationarity nullify this argument, making the resulting model more plausible? The answer is No. Nonstationarity, per se, is not sufficient. Hajnal (1956, 1958) [13] has shown that the same conclusions follow for a *nonstationary* regular Markov process operating on a homogeneous population: the long-run probability of being in any particular state is (asymptotically) independent of the initial state, depending only on the sequence of transition probabilities to which the process has been subjected. And since the latter are the same for each person if the population is homogeneous, everyone has the same probability of eventually being in some particular desired status.[14, 15]

The argument which does not assume stationarity is complicated, and will not be reproduced here, but the implication concerning the plausibility of various models is the same: one must modify either the homogeneity assumption, the regularity assumption, or the Markovian assumption.

Some readers may be more inclined to discard the regularity assumption than to discard either the homogeneity assumption or the Markovian assumption. In fact, this is the second approach taken by Mayer (1968a). The regularity assumption (both strong and weak forms) rules out the possibility of either a set of states, called an "ergodic set," or a single state, called an "absorbing state," which can never be left once it is entered.

Now there is certainly one class of events with zero probability: the class of logically impossible events. Some examples of mobility events which are logically impossible are provided by the Cornell Mobility Model:

[13] See Footnote 1 for a difference in terminology between Hajnal's articles and the present paper.

[14] As in the stationary case (cf. Footnote 12 and the related paragraph in the text), a nonstationary model could avoid this implausible conclusion if the sequence of transition matrices is chosen in such a manner as to prolong the convergence time until after most of the members of the population disappear from mortality. This approach, postulating high mobility at the younger ages and almost no mobility at the older ages, would seem somewhat more plausible than the approach of Footnote 12, which postulates a set of *stationary* transition probabilities concentrated so heavily on the diagonal as to prolong convergence (cf. Mayer, 1968a).

[15] Another conclusion reached in the preceding section—convergence to a stable distribution which thereafter remains constant—does not hold in the case of nonstationary transition probabilities. The long-run probability of being in any particular state continues to change even after it becomes (asymptotically) independent of the initial state. This can be explained heuristically as follows. Each transition matrix determines a "target" toward which it moves the distribution during the interval of time when it is applied. But the actual distribution never reaches a "target" distribution, since the "target" keeps moving as the transition matrices continue changing over time.

e.g., moving from a state with a duration of three time units to a state with a duration of five time units in a single unit of time. Such logically impossible events certainly have zero probability. But it seems extremely implausible to assert that any logically possible move has a probability of precisely zero.

If this criterion is accepted, then no ergodic set or absorbing state can be directly observable. For example, a move is always *logically* possible by a man with any combination of status and duration, no matter how long the duration in his current status. Hence, if our criterion is accepted, one would not modify the finite version of the Cornell Mobility Model by making the highest duration considered into absorbing states, although this possibility has been explored by the authors of the model. Rather, ergodic sets or absorbing states would have to be unobservable, and a logical impossibility of moving from them would be imposed *by definition.* An example of this kind, mentioned earlier, is Mayer's (1968a) second model, which is patterned after Cohen's (1963) model for the Asch experiment. In this model, each observable social status corresponds to two nonobservable states, one absorbing and the other not, and as long as a man remains in a given social status, there is no way of determining for certain whether or not he has been absorbed.

At the beginning of the preceding section we discussed some strong *a priori* reasons for believing that the men in a mobility process are not homogeneous. No comparable argument appears to exist for nonobservable ergodic sets or absorbing states. The choice between these two hypotheses appears to be partially a matter of taste, not subject to empirical test. One widely accepted criterion in such matters is parsimony. The model in the preceding section, at least in its current form, can not account for observed numerical mobility data in terms of a small number of parameters; but the success of the Mover-Stayer Model suggests that perhaps models with nonobservable absorbing states can.

However, in one important sense, heterogeneous models such as the present model and the Mover-Stayer Model are potentially much superior to the Mayer (1968a) absorbing state model. In Mayer's model the question of *who* moves and who does not move never receives a sociological answer; it is primarily a matter of chance, determined by the random coincidence of who happens to be absorbed. In contrast, in models with heterogeneous populations the transition probabilities could be taken as *dependent* variables, and the matter of who moves and who does not move would have a sociological explanation involving the matching of attributes of particular persons and requirements of particular statuses. For this reason we find it preferable to relax the homogeneity assumption.

So far in this paper we have seriously questioned the stationarity assumption, the assumption of a homogeneous population, and the regularity assumption; but throughout all this, the Markovian assumption has remained untouched (although sometimes, as in the Cornell Mobility Model, this required a reformulation). But no paper entitled "Intragenerational Social Mobility as a Markov Process" could really be complete without an explicit discussion of the Markovian assumption.

In part, the use of this assumption is a matter of mathematical convenience: it plays a crucial role in the derivation of the long-run properties of the process; such a derivation would be at least much more difficult, and perhaps impossible, without the Markovian assumption (depending on what alternative was used to replace it). But this assumption, like any other mathematical assumption in a model of an empirical phenomenon, also has substantive meaning, and to this we now turn.

The Markovian assumption is sometimes characterized as a probabilistic analog of the principle of scientific determinism (Parzen, 1962:187). The latter principle states that the complete description of the state of a physical system at any single point in time (without any *direct* reference to its history of prior states) is sufficient to permit the derivation of its state at any future point in time. In a probabilistic system, as opposed to a completely deterministic system, it is the set of probabilities of various possible states, rather than the precise state, which is derived for some future point in time. And the principle of scientific determinism applied in this context becomes the Markovian assumption.

It is important to note that, in the deterministic case, the complete description of the current state of a physical system necessarily contains *indirect* information about the history of the system: if a ball is currently rolling up hill, we can be certain that sometime in the past it was the recipient of a force that pushed it in that direction. The same type of considerations arise in social phenomena. Whether or not a man ever attended college is a matter of history, at least in the way we usually think about it. Nevertheless, this information is part of his *vita* which is currently being examined by his prospective employer; i.e., this information is contained in a *complete* description of his *current* situation. In fact, the *vita*, the job application form, the credit record, the record of law enforcement agencies—all these are devices specifically designed by society to ensure that a man's current situation *does* include certain parts of his history.

For this reason, the Markovian assumption is entirely compatible with empirical evidence that one's future mobility does depend to a large extent on such historical matters as his educational attainment and certain features of his family of origin (Blau & Duncan, 1967). But this compatibility can be attained only by incorporating all the currently relevant historical matters into the description of one's current state in the Markovian process or, equivalently, using a heterogeneous population whose probabilities depend on such historical matters. The failure to do so is another inadequacy of the model in the preceding section. In this respect the Cornell Mobility Model is a step in the right direction, but doesn't go far enough.

This substantive meaning of the Markovian assumption has not, unfortunately, been made very clear by a number of authors, giving one the impression that mathematical convenience was their only reason for making the assumption. The mathematical technique for incorporating parts of a man's history into the description of his current state is not new: it is well-known in the context of higher-order Markov chains (e.g., Goodman, 1962); and it was used in the Cornell Mobility Model discussed above. The reason for discussing it here is to point out, substantively, one mechanism by which it takes place in the real world, namely, the current scanning of pertinent historical records in preparation for the decision on whether to hire a man.

TOWARD A MORE ADEQUATE MODEL OF INTRAGENERATIONAL SOCIAL MOBILITY

We have already made several suggestions regarding a more adequate model, but let us consolidate and reiterate them here before making some concluding remarks.

First, an adequate model must involve either ergodic sets (which includes absorbing states as a special case), a heterogeneous population, or both, in order to avoid the implausible conclusion that long-run achievement is statistically independent of origin. Models with absorbing states may require fewer parameters than models with heterogeneous populations, and hence appeal to parsimony. But, on the other hand, in models with heterogeneous populations the transition probabilities could conceivably be taken as dependent variables which are a function of individual characteristics of the persons involved, such as intelligence and personality. This would make the question of *who* has a low (or zero) probability of movement more than a matter of chance, more than the coincidence of who happens to get absorbed, as it is in a model with absorbing states.

Second, if the personal characteristics of continued relevance to mobility are not incorporated into the model by making the transition probabilities a function of them, they must be incorporated in some equivalent manner. One equivalent procedure is to make the states more complex than mere occupational statuses or status-duration pairs, incorporating in the specification of "states" the historical matters, such as family background, intelligence, and education, which have a continuing influence on mobility.

Third, the stationarity assumption is rather implausible. We did not suggest any specific alternative, although the discussion of other models raised the possibility that mobility probabilities might decline with age and/or duration of current status. An additional reason for suggesting nonstationarity, not mentioned earlier, is that technological developments and economic conditions seem to determine, to a large extent, the relative numbers of jobs available in the various occupational categories, and hence it would seem that changes in these societal characteristics would produce changes in the relative sizes of the various transition probabilities.[16, 17]

In summary, a fully adequate Markovian model of intragenerational social mobility would seem to require a considerably larger number of parameters than any of the models discussed. Hence, in trying to fit it to numerical mobility data, one would run into very serious estimation problems, worse than those faced by the Cornell Mobility Model (finite version), since the amount of data required increases rapidly with the number of parameters to be estimated. Hence we may be faced with the unhappy conclusion that any really adequate model would be too cumbersome to be fitted to numerical data. If this conclusion is correct, then we must be content to fit simplified and inadequate models to numerical data. But at least we can, and should, be aware of some of their inadequacies.

REFERENCES

Bartholomew, D. J.
1967 Stochastic Models for Social Processes. New York: Wiley.

Bartos, O. J.
1967 Simple Models of Group Behavior. New York: Columbia University Press.

Beshers, J. M. and E. O. Laumann
1967 "Social distance: A network approach." American Sociological Review 32:225–236.

[16] Some steps in this direction, though outside the Markovian framework considered in this paper, have been taken by White (1963, 1968) and Coleman (1968). Also related are Goodman's (1965) and Duncan's (1966) discussions of previous work on "structural" mobility; Matras (1961) has considered some effects of differential fertility.

[17] In personal correspondence, R. McGinnis suggests that the stationarity assumption has a logical status similar to that of the Markovian assumption, as discussed in the preceding section: if the states are defined in such a manner as to include all relevant independent variables, then the transition probabilities must necessarily be stationary. But incorporation of societal variables, as well as individual variables, into the specification of states may pose some special problems; this matter deserves further investigation.

Blau, P. M. and O. D. Duncan
1967 The American Occupational Structure. New York: Wiley.

Blumen, I., M. Kogan and P. J. McCarthy
1955 The Industrial Mobility of Labor as a Probability Process. Cornell Studies of Industrial and Labor Relations, Vol. 6. Ithaca, N. Y.: Cornell University.

Cohen, B. P.
1963 Conflict and Conformity: A Probability Model and Its Application. Cambridge: The M.I.T. Press.

Coleman, J. S.
1968 "Demand and supply considerations in mobility." Presented to the Cornell Conference on Human Mobility, Ithaca, October 31–November 2, 1968.

Duncan, O. D.
1966 "Methodological issues in the analysis of social mobility." Pp. 51–97 in N. J. Smelser and S. M. Lipset, eds., Social Structure and Mobility in Economic Development. Chicago: Aldine.

Feller, W.
1968 An Introduction to Probability Theory and Its Applications. Vol. 1, Third Edition. New York: Wiley.

Goodman, L. A.
1961 "Statistical methods for the Mover-Stayer Model." Journal of the American Statistical Association 56:841–868.
1962 "Statistical methods for analyzing processes of change." American Journal of Sociology 68:57–78.
1965 "On the statistical analysis of mobility tables." American Journal of Sociology 70:564–585.

Hajnal, J.
1956 "The ergodic properties of non-homogeneous finite Markov chains." Proceedings of the Cambridge Philosophical Society 52:67–77.
1958 "Weak ergodicity in non-homogeneous Markov chains." Proceedings of the Cambridge Philosophical Society 54:233–246.

Henry, N. W., R. McGinnis and H. W. Tegtmeyer.
1968 "A finite model of mobility." Presented to the annual meeting of the Population Association of America, Boston, April 1968.

Hodge, R. W.
1966 "Occupational mobility as a probability process." Demography 3:19–34.

Kemeny, J. G. and J. L. Snell.
1960 Finite Markov Chains. Princeton: Van Nostrand.
1962 Mathematical Models in the Social Sciences. New York: Blaisdell. Appendix C: "Finite Markov chains."

Land, K. C.
1969 "Duration of residence and prospective migration: Further evidence." Demography 6:133–140.

Lieberson, S. and G. V. Fuguitt
1967 "Negro-white occupational differences in the absence of discrimination." American Journal of Sociology 73:188–200.

Matras, J.
1960 "Comparison of intergenerational occupational mobility patterns." Population Studies 14:163–169.
1961 "Differential fertility, intergenerational mobility, and change in the occupational structure." Population Studies 15:187–197.
1967 "Social mobility and social structure: Some insights from the linear model." American Sociological Review 32:608–614.

Mayer, T. F.
1967 "Birth and death process models of social mobility." Unpublished ms. Michigan Studies in Mathematical Sociology, paper number 2.
1968a "Age and mobility: Two approaches to the problem of nonstationarity." Presented to the annual meeting of the American Sociological Association, Boston, August 1968.
1968b "Multi-class Markov models and the rank criterion." Unpublished ms.

McGinnis, R.
1968 "A stochastic model of social mobility." American Sociological Review 33:712–722.

Morrison, P. A.
1967 "Duration of residence and prospective migration: The evaluation of a stochastic model." Demography 4:553–561.

Myers, G. C., R. McGinnis and G. Masnick
1967 "The duration of residence approach to a dynamic stochastic model of internal migration: A test of the axiom of cumulative inertia." Eugenics Quarterly 14:121–126.

Parzen, E.
1962 Stochastic Processes. San Francisco: Holden-Day.

Prais, S. J.
1955a "Measuring social mobility." Journal of the Royal Statistical Society 118 (Series A): 55–66.
1955b "The formal theory of social mobility." Population Studies 9:72–81.

White, H. C.
1963 "Cause and effect in social mobility tables." Behavioral Science 8:14–27.
1968 "Opportunity chains." Sociological Abstracts 16(4): suppl. ASA p. 49. Abstract of paper presented to the annual meeting of the American Sociological Association, Boston, August 1968.

RACE, SOCIAL STATUS, AND CRIMINAL ARREST *

EDWARD GREEN

Eastern Michigan University

The Negro-white arrest rate differential for selected years between 1942 and 1965 in a northern industrial community is analyzed with respect to age, sex, and the socioeconomic variables of employment status, occupation, and migration. Although the incidence of recorded Negro crime has greatly increased since 1942 owing to the increase in the Negro population, the rate of Negro crime has decreased. The magnitude of the excess of the Negro over the white arrest rate reflects the wider distribution among Negroes of the lower social class characteristics of unemployment, employment in unskilled and semiskilled occupations, and migration from the rural South. The findings do not support color-caste theories which interpret Negro criminal behavior as a response to racial proscriptions or which construe Negro criminality as a function of racially suppressive law enforcement tactics.

INTRODUCTION

The recurrently higher official arrest rate of Negroes over whites poses a persistent issue in the study of deviance relative to ethnicity. Although it is well established that criminogenic conditions such as poverty, family instability, slum residence, and migration are much more concentrated among Negroes than whites, the extent to which the differential accounts for the racial variance in crime rates remains problematical (Sellin, 1928:64). One point of view holds out the prospect that under comparable circumstances the white and Negro crime rates would not differ substantially (Wolfgang, 1964:61), a presumption which finds some support in the historical experience of lower class white migrant groups who as recent arrivals on the American urban scene also incurred high arrest rates which later declined in relation to their upward social movement and cultural assimilation. A less sanguine view holds that the circumstances of whites and Negroes are not fully comparable, that the experience of the Negro in America differs not only in degree but in kind from that

* This research is part of a larger study supported by the Walter E. Meyer Research Institute of Law.