

Extensions of the Mover-Stayer Model¹

Seymour Spilerman

University of Wisconsin

A major drawback to the traditional Markov formulation of mobility processes is that it assumes population homogeneity with respect to transition behavior. This assumption is clearly violated in most instances of social mobility. In an attempt to relax the homogeneity requirement and still retain the essential character of a Markov process, Blumen, Kogan, and McCarthy (1955) developed the "mover-stayer" model in which heterogeneity is attributed to the presence of two types of persons who differ in their rates of movement. In the present paper, the mover-stayer model is generalized to permit a continuous distribution of persons by rate of mobility. The model is illustrated with simulated data and then applied to an analysis of interregional migration.

INTRODUCTION

Applications of Markov processes to the study of social mobility have commonly concluded with the observation that individuals differ in their transition behavior. Although the Markov model requires population homogeneity, transitions from an origin state rarely conform to this assumption. Some persons simply move more often than, or differently from, others. This has been found with industrial mobility data (Blumen, Kogan, and McCarthy 1955), with intergenerational and intragenerational occupational mobility (Hodge 1966; Lieberman and Fuguitt 1967), and with geographic migration (Rogers 1966; Tarver and Gurley 1965).

The main difficulty derives from the Markov model having been constructed with repeated state changes by a single object in mind. In the analysis of social mobility, however, the movements of an entire population are at issue. If this population is heterogeneous in its transition behavior, then even if each individual were to satisfy the central assumption of a first-order Markov process—namely, that his probabilities of making particular transitions are determined solely by his present state and are independent of past history—the population-level process would not be Markovian.

¹ The research reported here was supported by funds granted to the Institute for Research on Poverty at the University of Wisconsin by the Office of Economic Opportunity, pursuant to provisions of the Economic Opportunity Act of 1964. I wish to thank David Dickens for his assistance with the statistical computations, and I. Richard Savage, and Neil Henry for their comments. An earlier version of this paper was presented at the Third Conference on the Mathematics of Population, University of Chicago, July 20–24, 1970.

Attempts to relax the homogeneity requirement while retaining the essential Markov framework have led to research in two directions. In one approach, interest has focused on the construction of subpopulation matrices and on ways to categorize individuals that would permit the "within-category" variation in transition behavior to be reduced. Operationally, this has usually meant disaggregating the population on attributes which are expected, either from theoretical considerations or empirical investigation, to relate to mobility and to construct a separate transition matrix for each subpopulation. For example, Rogers (1966) and Tarver and Gurley (1965), analyzing geographic migration, disaggregate the population to produce transition arrays by age categories and race. In this spirit, McFarland (1970) has reported an analytic method for combining subpopulation or individual-level transition matrices and projecting from these to the k -step population matrix, and I have (Spilerman 1972) presented a regression procedure for disaggregating the population matrix in order to obtain the individual-level transition arrays.

The above strategy casts the problem of heterogeneity into a framework in which each person is viewed as making a single transition during a unit time interval, but following a matrix relevant to the subpopulation that shares his particular attributes. A conceptually different approach to heterogeneity is embodied in an alternative assumption, namely, that all individuals move according to an identical transition matrix when they move but differ in their *rates* of mobility (see Spilerman [1972] for a discussion on the convergence of these two perspectives). Work in this direction has resulted in the development of the "mover-stayer" model (Blumen, Kogan, and McCarthy 1955). Under the specifications of this process, heterogeneity is handled by postulating two types of individuals—stayers, who remain permanently in their states of origin, and movers, who are homogeneous in their transition behavior and therefore follow a Markov process with a common transition matrix. Several estimation methods for the parameters of the mover-stayer model have been developed by Goodman (1961).

Aside from the novel conceptual perspective provided by this model, which seems appropriate to the analysis of geographic migration or intra-generational occupational mobility where repeated moves can be made by a person, it has the advantage of not requiring individual-level attribute data (although parameter estimation can be improved if such information on the waiting time to transition is available [Goodman 1961]). Since much of our mobility data lacks significant detail at the individual level, the mover-stayer model can be applied where the construction of subpopulation transition matrices is not possible.

Although the mover-stayer model postulates two types of persons, this is done out of necessity for keeping the process mathematically tractable,

not because the authors genuinely believed that instances of population heterogeneity can generally be attributed to two types of persons. In fact, in their concluding chapter, Blumen, Kogan, and McCarthy (BKM) (1955) discuss strategies for extending the mover-stayer model to incorporate a wider range of heterogeneity in the rate of transition, although they do not develop such a generalization. An extension of the mover-stayer model in which the rate of individual mobility is specified by a continuous distribution is constructed in this paper. Following the mathematical presentation, the extension is applied to regional migration data.

THE MOVER-STAYER MODEL AND BKM'S COMMENTS
ON GENERALIZATION

The Mover-Stayer Model

In their study of industrial mobility, BKM (1955, p. 62) report that calculations of k -step transition matrices from a Markov chain consistently underpredict the main diagonal elements of the observed k -step matrix. That is, if

$$P(1) = \begin{bmatrix} \hat{p}_{11} & \cdot & \cdot & \hat{p}_{1m} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \hat{p}_{m1} & \cdot & \cdot & \hat{p}_{mm} \end{bmatrix}$$

is the observed one-step transition matrix, and

$$P(k) = \begin{bmatrix} \hat{p}_{11}^{(k)} & \cdot & \cdot & \hat{p}_{1m}^{(k)} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \hat{p}_{m1}^{(k)} & \cdot & \cdot & \hat{p}_{mm}^{(k)} \end{bmatrix}$$

is the observed k -step transition matrix, then the k -step matrix predicted from a stationary Markov process, $\hat{P}(k) = P(1)^k$, will have main diagonal elements $(\hat{\hat{p}}_{11}^{(k)} \dots \hat{\hat{p}}_{mm}^{(k)})$, which commonly have the property that $\hat{\hat{p}}_{ii}^{(k)} < \hat{p}_{ii}^{(k)}$ for $i = 1, \dots, m$.

Although one might suspect that overtime change in the p_{ij} elements of $P(1)$ is responsible for this result, this is generally not the case. For example, Hodge (1966) reports similar findings with occupational mobility data, even though the $P(1)$ matrices he uses are time dependent. More formally, Hodge's analysis shows that if $P_t(1)$ for $t = 1, \dots, k$ are *observed* one-step transition matrices for successive time intervals, and if $\hat{\hat{P}}(k)$ is the predicted k -step array,

$$\hat{P}(k) = \prod_{t=1}^k P_t(1),$$

then the relationship between the main diagonal elements of $\hat{P}(k)$ and $P(k)$ may have the same structure as that described between $\hat{P}(k)$ and $P(k)$. The problem is not one of the p_{ij} elements of $P_t(1)$ changing over time for the population but, rather, that some persons are less apt to move than others in each time interval.

To contend with this situation, BKM suggest decomposing the population into movers and stayers,

$$P(1) = S + (I - S)M, \quad (1)$$

where S is a diagonal matrix containing as entries the proportion of persons in each origin state who remain there permanently, $I - S$ is a diagonal matrix with entries which indicate the proportion in a state who are potentially mobile, and M is the transition matrix for mobile individuals. The assumptions of the mover-stayer model, then, are (a) a proportion of the population in each state that never moves, (b) the population which is mobile is homogeneous in its pattern of movement and follows a Markov process, and (c) the process is stationary. We therefore have for the predicted k -step matrix,

$$\hat{P}(k) = S + (I - S)M^k.$$

Follow-up work on this model (Goodman 1961) has been concerned primarily with deriving consistent estimators for the matrices S and M and testing hypotheses relating to the mover-stayer process. Conceptually, however, there is a need to develop models which incorporate a greater range of heterogeneity. Instead of postulating two types of persons, we should like a process which handles several types and, ideally, a continuous range of individual differences in the rate of movement. Blumen, Kogan, and McCarthy address this problem in their concluding chapter (1955, pp. 138-46). Since the extension developed here proceeds from their suggestion, I first present their remarks.

BKM's Comments on Extending the Mover-Stayer Model

Instead of requiring every person to make a fixed number of transitions in each time interval, we assume that transitions are random occurrences²

² A transition from state i to state i would be considered movement in this terminology. For example, if the states of the system were geographic regions, an $i \rightarrow i$ transition would represent *intraregional* movement. Alternatively, we might speak of the expected number of *exposures* to movement as do BKM (1955, p. 139) but allow an individual to not move (an $i \rightarrow i$ transition) at an exposure.

Extensions of the Mover-Stayer Model

and that the rate of movement by an individual refers to his *expected* number of transitions, not to the actual number. This is conceptually reasonable, since an individual with rate equal to, say, three moves per unit time interval will not necessarily make this exact number of transitions in every time unit. He may make zero or one moves in some intervals, four or five in others. Over a long time period he will nevertheless average three moves per time unit. A formal way of stating this is to assume that individuals move in accordance with a Poisson process³ with parameter value (expected number of moves per unit of time) $\lambda = 3$.

Consider for the moment only individuals with a common rate of movement equal to λ . The $P(1)$ matrix would then be given by

$$P(1) = \sum_{v=0}^{\infty} r_v(1)M^v, \quad (2)$$

where $r_v(t)$ is a Poisson probability,

$$r_v(t) = \frac{(\lambda t)^v e^{-\lambda t}}{v!}$$

(with $t = 1$ in equation 2), and indicates the proportion of individuals who are expected to make v transitions during the time interval $(0, t)$ from among those having a rate equal to λ , and M is the transition matrix followed at each move. We also assume that the $r_v(t)$ values are the same for all states of the process.

With the above specification, it is easy to show⁴ that $P(1) = e^{-\lambda[I-M]}$ for any matrix M . This result is important because the k -step transition matrix $P(k)$ is now given by

$$P(k) = e^{-\lambda k(I-M)} = [e^{-\lambda(I-M)}]^k = [P(1)]^k.$$

³ A precise specification of the (stationary) Poisson process is given by the following four assumptions: (1) λ is constant over time. (2) In an infinitesimal time interval Δt , at most one event can occur. (3) The probability of an event in Δt equals $\lambda \Delta t$; the probability of no event in Δt equals $1 - \lambda \Delta t$. (4) The occurrence of an event during $(t, t + \Delta t)$ is independent of the past behavior of the process. The derivation of the Poisson distribution from these assumptions is a straightforward procedure (see Feller 1957, p. 400).

⁴

$$\begin{aligned} P(t) &= \sum_{v=0}^{\infty} r_v(t)M^v = \sum_0^{\infty} \frac{(\lambda t)^v e^{-\lambda t}}{v!} M^v \\ &= e^{-\lambda t} \sum_0^{\infty} \frac{(\lambda t M)^v}{v!} = e^{-\lambda t} e^{\lambda t M} = e^{-\lambda t[I-M]}. \end{aligned}$$

The convergence of the infinite sum to $e^{\lambda t M}$ will hold for an arbitrary matrix M (Gantmacher 1959, p. 113).

Thus, relaxation of the fixed-number-of-moves assumption does not necessarily prevent the population-level process from being Markovian. In the particular case where transitions are Poisson events⁵ and the population is homogeneous in its transition rate, the Markov requirement will, in fact, be satisfied.

Now assume that we have g types of persons who differ in their rates of mobility. Each individual, however, follows the same M matrix when making a transition. If a proportion q_1 of the population moves with rate λ_1 , a proportion q_2 moves with rate λ_2 , etc., we could write separate equations identical to equation (2) for each subpopulation. Alternatively, let $r_v(t)$ equal the expected proportion of the *total* population which makes v transitions during the interval $(0,t)$, irrespective of the individual mobility rates. Then, combining the coefficients of M^v from the separate processes, we have

$$r_v(t) = \sum_{i=1}^g q_i \frac{(\lambda_i t)^v e^{-\lambda_i t}}{v!}. \quad (3)$$

If we generalize this result from g types of persons to sampling from a continuous distribution $f(\lambda)$, we obtain

$$r_v(t) = \int_0^{\infty} \frac{(\lambda t)^v e^{-\lambda t}}{v!} f(\lambda) d\lambda. \quad (4)$$

Equation (4) says that the expected proportion of individuals who make v transitions in the time interval $(0,t)$ equals the sum of the products of two quantities: (a) the proportion of individuals with rate equal to λ , and (b) the probability that an individual with rate λ will make v transitions. The summation is taken over all possible values of λ , which is assumed to have a continuous density function.

Blumen, Kogan, and McCarthy develop the generalization of the mover-stayer model to this point (as does Bartholomew [1967, pp. 27-37] in a recent review of mobility models). In the next section, we present a solution to the proposed extension, provide an estimation procedure for the parameters of the model, and discuss strategies for testing whether the assumptions of the model are met by data.

AN EXTENSION OF THE BASIC MODEL

Assumptions and the Derivation

In order to extend the mover-stayer model, it is necessary to specify a form for $f(\lambda)$ in equation (4). Since we have little a priori knowledge

⁵ In fact, the assumption that the occurrence of transitions follows a Poisson process leads to a *continuous-time* Markov formulation.

Extensions of the Mover-Stayer Model

about this distribution, we assume a very general family of curves and use the observed data to estimate parameters for the specific distribution. We do, however, restrict $f(\lambda)$ to be unimodal or decline exponentially. This seems reasonable, since several studies of mobility (Goldstein 1964; Lipset and Bendix 1959, p. 158; Palmer 1954, p. 50; and Taeuber, Chiazze, and Haenszel 1968, p. 46) report distributions of persons by number of moves which have these forms.

Specifically, we assume that $f(\lambda)$ can be approximated by a gamma density,

$$f(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \quad \lambda > 0, \alpha > 0, \beta > 0, \quad (5)$$

where

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy.$$

The gamma distribution is a very general family of unimodal functions and is often assumed where the shape of the actual curve is unknown.

With this assumption regarding $f(\lambda)$, we obtain from equations (4) and (5) (see Chiang 1968, p. 49, for details on the integration)

$$\begin{aligned} r_v(t) &= \int_0^\infty \frac{(\lambda t)^v e^{-\lambda t}}{v!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda \\ &= \frac{\Gamma(v + \alpha)}{v! \Gamma(\alpha)} \beta^\alpha t^v (\beta + t)^{-(v+\alpha)}. \end{aligned} \quad (6)$$

Using the relation $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$, this result may be written as

$$r_v(t) = \binom{\alpha + v - 1}{v} \left(\frac{t}{\beta + t} \right)^v \left(\frac{\beta}{\beta + t} \right)^\alpha, \quad (7)$$

which is a negative binomial distribution. Thus, under the assumption that each individual's transitions follow a Poisson process, with the individual rates of mobility specified by a gamma density, the proportion of the population making v moves in $(0, t)$ will satisfy a negative binomial distribution.

Substituting this result into equation (2) yields for the one-step transition matrix,

$$P(1) = \sum_{v=0}^{\infty} r_v(1) M^v$$

$$= [p(t)]^a \sum_{v=0}^{\infty} \binom{v + \alpha - 1}{v} [q(t)M]^v, \tag{8}$$

where $p(t) = \beta/(\beta + t)$, $q(t) = t/(\beta + t)$, and $t = 1$.

Conditions for a Closed Form Solution to $P(1)$

Recall that for X , a scalar, and any real number α , $(1 - X)^{-\alpha}$ has the binomial expansion

$$\begin{aligned} (1 - X)^{-\alpha} &= 1 + \alpha X + \frac{\alpha(\alpha + 1)}{2!} X^2 \\ &\quad + \frac{\alpha(\alpha + 1)(\alpha + 2)}{3!} X^3 + \dots \\ &= \sum_0^{\infty} \binom{v + \alpha - 1}{v} X^v, \tag{9} \end{aligned}$$

where the condition for convergence of the infinite sum is $|X| < 1$. By extension, we write for the infinite sum of matrices in equation (8),

$$\sum_{v=0}^{\infty} \binom{v + \alpha - 1}{v} [q(t)M]^v = [I - q(t)M]^{-\alpha}, \tag{10}$$

where I is the identity matrix. We now discuss the condition for convergence of equation (10) and the computation of the right side when α is an arbitrary real number, not necessarily integer valued.

Analogous to the condition on X in the scalar case (eq. 9), the requirement for convergence of the infinite matrix sum is that all eigenvalues of $q(t)M$ are less than one in absolute value (Gantmacher 1959, p. 113). Since M is a transition matrix, it is stochastic and its largest eigenvalue equals one. However, $q(t) = t/(\beta + t) < 1$ for finite t (since $\beta > 0$), and this ensures that the eigenvalues of $q(1)M$, in particular, will be strictly less than one.

Substituting the result of equation (10) into equation (8), we obtain for $P(1)$,

$$P(1) = [p(1)]^a [I - q(1)M]^{-\alpha}. \tag{11}$$

A second consideration concerns the computation of $[I - q(t)M]^{-\alpha}$ when α is an arbitrary real number.⁶ For an integer k and a nonsingular

⁶ Since $[I - q(t)M]^{-\alpha} = ([I - q(t)M]^{-1})^{\alpha}$ and an inverse is defined only for a nonsingular array, this matrix power is defined only if the term in brackets is nonsingular. This condition, however, will be satisfied for $q(t) < 1$, i.e., for finite t .

Extensions of the Mover-Stayer Model

matrix A , the matrix power A^k always exists. For an arbitrary real number α , we define A^α to be

$$A^\alpha = e^{\alpha \log A},$$

which will hold for nonsingular A (Gantmacher 1959, p. 240).

If the eigenvectors of A are linearly independent (which will be the case if A has distinct eigenvalues), then we can diagonalize A ,

$$A = HDH^{-1}, \tag{12}$$

where D is a diagonal matrix with the eigenvalues of A as entries, and H is a matrix containing the eigenvectors of A as columns.⁷ In this circumstance, $A^{-\alpha}$ can be written⁸ as

$$A^{-\alpha} = e^{-\alpha \log A} = He^{-\alpha \log D} H^{-1} = HD^{-\alpha}H^{-1}, \tag{13}$$

where $D^{-\alpha}$ is a diagonal matrix with elements of the form $\mu^{-\alpha}$. Consequently, letting $[I - q(1)M] = A$ in equation (11), $P(1)$ can be computed from M by the diagonalization procedure of equation (13).

Obtaining the M Matrix from $P(1)$

Equation (11) shows how we can obtain the population transition matrix if we know M , the individual level transition matrix, and α and β , the parameters of $f(\lambda)$. Unfortunately, our problem is generally the reverse of this situation. Given an observed population transition matrix $P(1)$, and estimates of α and β from the empirical distribution of the number of moves by an individual, we wish to obtain M so that equation (11), rewritten in general form, may be used to project to $P(t)$ for some $t > 1$.

No difficulty is encountered in solving equation (11) for M as long as the main diagonal elements of $P(1)$ are greater than one-half. We obtain

$$M = \frac{1}{q(1)} \{I - p(1)[P(1)]^{-1/\alpha}\}. \tag{14}$$

The requirement on the p_{ii} entries constitutes a sufficient (though not necessary) condition for ensuring the uniqueness of M .⁹ When this re-

⁷ If A does not have linearly independent eigenvectors equal in number to its order (which may be the case if the eigenvalues are not distinct), then A cannot be diagonalized. It can, however, be put in Jordan form (Bellman 1960, p. 191) which creates computational difficulties but, frequently, not theoretical ones. With real data, it is rare that the eigenvalues are not distinct, so only the case where $[I - q(t)M]$ can be diagonalized is considered in this paper.

⁸ The eigenvalues of D may be complex numbers, in which case $\log \mu = \log r + i\theta$, $-\pi < \theta \leq \pi$, where r and θ are the polar form components of the eigenvalue μ . To guarantee uniqueness the principal branch of the logarithm is used.

⁹ An analogous restriction is necessary to obtain the matrix of transition intensities,

quirement is not satisfied multiple solutions may exist to equation (14). In this situation, if individual histories are available, the researcher has an option of either reducing the real-time duration of the unit time interval so that the condition on $P(1)$ will be met, or calculating M directly from the observed transitions.

Because most mobility processes operate "slowly" relative to frequency of sampling, the restriction on the p_{ii} entries is not a particularly severe one, even in the absence of individual-level data. Where it is satisfied, M can be obtained by diagonalizing $P(1)$ in the manner discussed previously:

$$[P(1)]^{-1/a} = Ke^{-(1/a) \log Q} K^{-1} = KQ^{-1/a}K^{-1},$$

where Q is the eigenvalue matrix of $P(1)$, and the columns of K are the corresponding eigenvectors. Equation (14) therefore provides a method for estimating M from the population-transition matrix $P(1)$, under the assumption that population heterogeneity in the rate of movement can be specified by a gamma density.

The remaining parameters of the model, α and β of the gamma distribution, can be estimated directly from observed data on the number of moves by an individual. If \bar{v} and S_v^2 are the sample mean and variance of this variable, then estimates of α and β can be obtained in terms of these values from the mean and variance formulas for a negative binomial variate (Chiang 1968, p. 50). This yields

$$\left. \begin{aligned} \hat{\beta} &= \frac{\bar{v}}{S_v^2 - v} \\ \hat{\alpha} &= \hat{\beta}\bar{v} \end{aligned} \right\} \quad (15)$$

Projection

Having computed M , we can now project forward in time to find $P(t)$, the t -step transition matrix

$$P(t) = \left(\frac{\beta}{\beta + t} \right)^a \left[I - \left(\frac{t}{\beta + t} \right) M \right]^{-a}. \quad (16)$$

This result will hold for finite t . For very large t , $q(t) \approx 1$ and $[I - q(t)M] \approx [I - M]$, which is singular since M is stochastic. Equation (16) is not defined in this circumstance, but other considerations (see Appendix A) suggest that $\lim P(t) = M(\infty)$, the equilibrium matrix for M , which may be found by the usual Markov methods.¹⁰

corresponding to a continuous-time Markov process, from a probability transition matrix (see Coleman 1964, p. 181). I am indebted to Burton Singer for an enlightening conversation on this subject.

¹⁰ This assumes an absence of $\lambda = 0$ individuals. If stayers are present, then $\lim P(t)$ is given by the mover-stayer formulation (BKM 1955, pp. 111-14), $\lim P(t) = S + (I - S)M(\infty)$.

If we view the process as embedded in continuous time so that noninteger values of t are meaningful, these can also be used with equation (16). We are not restricted to multiples of the initial time period, since it is the distribution of the population with respect to the expected number of transitions that is changing and this change is continuous. By any time t , of course, each person will have made an integer number of transitions. Likewise, no mathematical difficulty is presented with projecting backward in time. Starting with $P(1)$, for example, we can find $P(1/2)$.¹¹ This flexibility is useful because we can often obtain better estimates for α and β after a sizable number of moves have been made. The projections of the model could then be compared with transition matrices for shorter time intervals.

Testing the Model

At the outset, we assumed that the matrix $P(t)$ could be written in the form presented in equation (2), and consequently that $r_v(t)$, the proportion of the population expected to make v transitions by time t , is the same for all system states. Blumen, Kogan, and McCarthy (1955, p. 139) also make this assumption in their comments on extending the basic mover-stayer model. A necessary condition for this requirement to be satisfied is that the individual rates of transition not be a function of the state a person is in. Where data are available this assumption can be tested by computing the distribution of waiting times to a transition separately for individuals in each of the system states, and comparing these distributions.

There are two other assumptions of the model which can be tested—that individuals move in accordance with a Poisson process, and that the gamma density provides a reasonable approximation to the distribution of mobility rates in the population. To test one of these, we must, however, assume the validity of the other. If we assume that transitions are Poisson events, we can test the gamma specification by comparing the fit of the negative binomial estimates from equation (7) with the actual distribution of the number of moves. If the fit is poor, then the gamma assumption for $f(\lambda)$ in equation (5) should be altered and the Poisson compounded with a

¹¹ One referee suggested the following interpretation for this procedure. Consider an individual who is in state i at time 0, state j at time $1/2$, and state k at time 1. In arriving at equation (14), the information that his intermediate state is j was discarded; only the information that he made two moves from time 0 to time 1, beginning in state i and terminating in state k , is used. How, then, can $P(1/2)$ be reconstructed? The argument is that because of the assumption of an identical transition matrix for every person and a continuous distribution of mobility rates, there are other individuals who move just like him, only more slowly, who will be in state i at time 0, state j at time 1 , and the information they contribute to $P(1)$ serves as a proxy for the information he would have contributed to $P(1/2)$.

different curve.¹² Alternatively, repair work can be done on the gamma distribution (see discussion on the spiked gamma in the following section).

There are direct tests for whether the occurrence of events conforms to a Poisson process, such as on the interarrival times, but they assume an absence of heterogeneity. If a large number of interarrival times were available for each person, we could test the Poisson assumption separately for each individual without concern for the distribution of λ in the population. Most social data are not so rich in detail; consequently, it is necessary to assume the correctness of $f(\lambda)$ in order to test the Poisson assumption. Therefore, if we believe $f(\lambda)$ to be gamma, a comparison between the actual distribution of moves and the negative binomial estimates would provide a test of the Poisson specification. In practice, however, the form of $f(\lambda)$ would seem to be more problematic and the more interesting question.

AN EXAMPLE USING SIMULATED DATA

The advantage of illustrating the model with constructed data is that we have full knowledge of the actual mobility characteristics of the hypothetical population. We will assume an individual-level transition matrix and a population distribution by rate of movement as presented in table 1. In practice, this information usually would not be available.

TABLE 1
STRUCTURE OF THE SIMULATED DATA

A. INDIVIDUAL-LEVEL TRANSITION MATRIX				B. DISTRIBUTION OF THE POPULATION BY RATE OF MOBILITY	
				λ	Proportion of the Population with This λ Value
$M = \begin{bmatrix} .600 & .200 & .100 & .100 \\ .150 & .700 & .100 & .050 \\ .100 & .100 & .750 & .050 \\ .050 & .050 & .100 & .800 \end{bmatrix}$				0.1	0.25
				1.0	0.35
				2.0	0.20
				3.0	0.10
				4.0	0.06
				5.0	0.04
					1.00

We further assume that the six types of persons in the population (table 1, part B) move in accordance with a Poisson process which is specified by the indicated λ value for each subpopulation. Consequently, the Poisson

¹² If the fit is highly deficient, or if the assumption in the preceding paragraph is untenable, an individual-level transition matrix M which satisfies equation (8) may not exist. In this case, a nonstochastic M will be obtained from equation (14).

Extensions of the Mover-Stayer Model

distribution was used to generate an expected proportion of each subpopulation who make $v = 0, 1, 2, \dots$ moves during the time interval (0,1). These values, multiplied by the respective subpopulation proportions in the total population, were aggregated to produce a distribution of the total population by number of moves. This distribution is presented in column 1 of table 2. Each of the $r_v(1)$ values has been multiplied by 1,000

TABLE 2
DISTRIBUTION OF NUMBER OF MOVES FROM OBSERVED
(SIMULATED) DATA AND FROM NEGATIVE
BINOMIAL ESTIMATES

v	[1000 $r_v(1)$]	[1000 $\hat{r}_v(1)$]
Number of Moves	Number of Persons with v Moves (Observed Data) (1)	Number of Persons with v Moves (Calculated from Negative Binomial $\alpha = 1.371, \beta = .915$) (2)
0	388	363
1	226	260
2	153	161
3	97	94
4	59	54
5	34	30
6	19	17
7	11	9
8	6	5
9	2	3
10	1	1
1,000 $\sum r_v(1)$	996*	997*

NOTE.— $\bar{v} = 1.498$; $S_v^2 = 3.133$.

* Value is less than 1,000 because of rounding error.

so we can refer to the number of persons making a specified number of moves. These “observed” data were then used with the equation

$$P(1) = \sum_{v=0}^{10} r_v(1)M^v$$

to generate an “observed” transition matrix $P(1)$. This process was repeated¹³ for the intervals (0,3) and (0,6), so we have three observed transition matrices: $P(1)$, $P(3)$, and $P(6)$. The matrices produced by this construction are presented in table 3. Normally, these transition arrays and the distribution of the population by number of moves (table 2, col. 1)

¹³ The upper limit on the summation was set to 30 for $t = 3$ and to 60 for $t = 6$.

TABLE 3
OBSERVED TRANSITION MATRICES FOR THE
TIME INTERVALS (0, 1), (0, 3), AND (0, 6)

$P(1) =$	<table style="border-collapse: collapse; margin: auto;"> <tr><td style="padding: 2px 10px;">.650</td><td style="padding: 2px 10px;">.156</td><td style="padding: 2px 10px;">.101</td><td style="padding: 2px 10px;">.093</td></tr> <tr><td style="padding: 2px 10px;">.118</td><td style="padding: 2px 10px;">.719</td><td style="padding: 2px 10px;">.101</td><td style="padding: 2px 10px;">.063</td></tr> <tr><td style="padding: 2px 10px;">.090</td><td style="padding: 2px 10px;">.102</td><td style="padding: 2px 10px;">.747</td><td style="padding: 2px 10px;">.061</td></tr> <tr><td style="padding: 2px 10px;">.056</td><td style="padding: 2px 10px;">.064</td><td style="padding: 2px 10px;">.101</td><td style="padding: 2px 10px;">.779</td></tr> </table>	.650	.156	.101	.093	.118	.719	.101	.063	.090	.102	.747	.061	.056	.064	.101	.779
.650	.156	.101	.093														
.118	.719	.101	.063														
.090	.102	.747	.061														
.056	.064	.101	.779														
$P(3) =$	<table style="border-collapse: collapse; margin: auto;"> <tr><td style="padding: 2px 10px;">.448</td><td style="padding: 2px 10px;">.218</td><td style="padding: 2px 10px;">.178</td><td style="padding: 2px 10px;">.155</td></tr> <tr><td style="padding: 2px 10px;">.165</td><td style="padding: 2px 10px;">.529</td><td style="padding: 2px 10px;">.178</td><td style="padding: 2px 10px;">.128</td></tr> <tr><td style="padding: 2px 10px;">.143</td><td style="padding: 2px 10px;">.177</td><td style="padding: 2px 10px;">.555</td><td style="padding: 2px 10px;">.125</td></tr> <tr><td style="padding: 2px 10px;">.110</td><td style="padding: 2px 10px;">.135</td><td style="padding: 2px 10px;">.178</td><td style="padding: 2px 10px;">.577</td></tr> </table>	.448	.218	.178	.155	.165	.529	.178	.128	.143	.177	.555	.125	.110	.135	.178	.577
.448	.218	.178	.155														
.165	.529	.178	.128														
.143	.177	.555	.125														
.110	.135	.178	.577														
$P(6) =$	<table style="border-collapse: collapse; margin: auto;"> <tr><td style="padding: 2px 10px;">.365</td><td style="padding: 2px 10px;">.235</td><td style="padding: 2px 10px;">.215</td><td style="padding: 2px 10px;">.184</td></tr> <tr><td style="padding: 2px 10px;">.177</td><td style="padding: 2px 10px;">.441</td><td style="padding: 2px 10px;">.215</td><td style="padding: 2px 10px;">.166</td></tr> <tr><td style="padding: 2px 10px;">.164</td><td style="padding: 2px 10px;">.210</td><td style="padding: 2px 10px;">.462</td><td style="padding: 2px 10px;">.164</td></tr> <tr><td style="padding: 2px 10px;">.141</td><td style="padding: 2px 10px;">.181</td><td style="padding: 2px 10px;">.215</td><td style="padding: 2px 10px;">.463</td></tr> </table>	.365	.235	.215	.184	.177	.441	.215	.166	.164	.210	.462	.164	.141	.181	.215	.463
.365	.235	.215	.184														
.177	.441	.215	.166														
.164	.210	.462	.164														
.141	.181	.215	.463														

TABLE 4
POPULATION TRANSITION MATRICES PREDICTED FROM A
STATIONARY MARKOV PROCESS

$\hat{P}(1) = P(1) =$	<table style="border-collapse: collapse; margin: auto;"> <tr><td style="padding: 2px 10px;">.650</td><td style="padding: 2px 10px;">.155</td><td style="padding: 2px 10px;">.101</td><td style="padding: 2px 10px;">.093</td></tr> <tr><td style="padding: 2px 10px;">.118</td><td style="padding: 2px 10px;">.719</td><td style="padding: 2px 10px;">.101</td><td style="padding: 2px 10px;">.063</td></tr> <tr><td style="padding: 2px 10px;">.090</td><td style="padding: 2px 10px;">.102</td><td style="padding: 2px 10px;">.747</td><td style="padding: 2px 10px;">.061</td></tr> <tr><td style="padding: 2px 10px;">.056</td><td style="padding: 2px 10px;">.064</td><td style="padding: 2px 10px;">.101</td><td style="padding: 2px 10px;">.779</td></tr> </table>	.650	.155	.101	.093	.118	.719	.101	.063	.090	.102	.747	.061	.056	.064	.101	.779
.650	.155	.101	.093														
.118	.719	.101	.063														
.090	.102	.747	.061														
.056	.064	.101	.779														
$\hat{P}(3) = [P(1)]^3 =$	<table style="border-collapse: collapse; margin: auto;"> <tr><td style="padding: 2px 10px;">.346</td><td style="padding: 2px 10px;">.261</td><td style="padding: 2px 10px;">.209</td><td style="padding: 2px 10px;">.183</td></tr> <tr><td style="padding: 2px 10px;">.199</td><td style="padding: 2px 10px;">.446</td><td style="padding: 2px 10px;">.209</td><td style="padding: 2px 10px;">.147</td></tr> <tr><td style="padding: 2px 10px;">.170</td><td style="padding: 2px 10px;">.209</td><td style="padding: 2px 10px;">.478</td><td style="padding: 2px 10px;">.143</td></tr> <tr><td style="padding: 2px 10px;">.126</td><td style="padding: 2px 10px;">.154</td><td style="padding: 2px 10px;">.209</td><td style="padding: 2px 10px;">.511</td></tr> </table>	.346	.261	.209	.183	.199	.446	.209	.147	.170	.209	.478	.143	.126	.154	.209	.511
.346	.261	.209	.183														
.199	.446	.209	.147														
.170	.209	.478	.143														
.126	.154	.209	.511														
$\hat{P}(6) = [P(1)]^6 =$	<table style="border-collapse: collapse; margin: auto;"> <tr><td style="padding: 2px 10px;">.231</td><td style="padding: 2px 10px;">.280</td><td style="padding: 2px 10px;">.264</td><td style="padding: 2px 10px;">.224</td></tr> <tr><td style="padding: 2px 10px;">.211</td><td style="padding: 2px 10px;">.317</td><td style="padding: 2px 10px;">.264</td><td style="padding: 2px 10px;">.207</td></tr> <tr><td style="padding: 2px 10px;">.200</td><td style="padding: 2px 10px;">.260</td><td style="padding: 2px 10px;">.337</td><td style="padding: 2px 10px;">.203</td></tr> <tr><td style="padding: 2px 10px;">.174</td><td style="padding: 2px 10px;">.224</td><td style="padding: 2px 10px;">.264</td><td style="padding: 2px 10px;">.336</td></tr> </table>	.231	.280	.264	.224	.211	.317	.264	.207	.200	.260	.337	.203	.174	.224	.264	.336
.231	.280	.264	.224														
.211	.317	.264	.207														
.200	.260	.337	.203														
.174	.224	.264	.336														

are the kinds of data which can be collected.¹⁴ We now assume that only this information is available.

Projections from a stationary Markov chain require that the observed $P(1)$ matrix be raised to the requisite power. Markov estimates of $P(3)$, $\hat{P}(3) = [P(1)]^3$, and of $P(6)$, $\hat{P}(6) = [P(1)]^6$, are presented in table 4. As a result of the heterogeneity in rate of movement which was built into the data the main diagonal elements from this projection are, as expected,

¹⁴ We also have available the distribution of $r_v(t)$ for $t = 3, 6$, analogous to col. 1 of table 2. This information will not be used in the illustration, but a test of the Poisson assumptions could also be based upon the change over time in the distribution.

consistently smaller than the corresponding observed values reported in table 3. Moreover, the discrepancy increases over time.

Turning to the model proposed in this paper, \bar{v} and S_v^2 (from table 2, col. 1) were first used with equation (15) to estimate α and β : $\hat{\alpha} = 1.371$, $\hat{\beta} = 0.915$. Having estimated these parameters, the negative binomial formula (eq. 7) can be used to generate a predicted distribution of moves. This distribution is presented in column 2 of table 2, alongside the observed values. While there are some sizable deviations between expected and observed figures, this method will usually produce a superior fit than simply dichotomizing the population into stayers and movers,¹⁵ especially when the heterogeneity is considerable.

The estimates of α and β , together with the observed $P(1)$ matrix from table 3, allow the M matrix to be derived using equation (14),

$$\hat{M} = \begin{bmatrix} .613 & .191 & .099 & .098 \\ .144 & .707 & .099 & .050 \\ .097 & .099 & .754 & .050 \\ .050 & .050 & .099 & .801 \end{bmatrix}$$

This array is an estimate of the individual-level transition matrix M which, by the assumptions of the model, is the same for all persons. Equation (16) can now be used with \hat{M} , $\hat{\alpha}$, and $\hat{\beta}$ to project to $P(t)$ for any value of t . Estimates of $P(1)$, $P(3)$, and $P(6)$ are presented in table 5.

A comparison of these predicted arrays with the observed transition matrices (table 3) and with the Markov projections (table 4) reveals the superiority of the present model. The main diagonal entries, in particular, decline less rapidly than in the Markov projections.¹⁶ The deviations from observed values, incidentally, cannot be attributed to any inadequacy with the Poisson assumptions, since these were used to generate the data. Rather, the deviations result from an inability of the gamma density to fit perfectly the constructed distribution of λ values in the population, although the discrepancy is not severe. (Compare the negative binomial estimates with the observed distribution of moves in table 2). In fact, using the computed values of α and β , the gamma density $f(\lambda)$ for this population can be drawn directly from equation (5). This graph is presented in figure 1. Superimposed on the curve are vertical lines which indicate the points of

¹⁵ Nevertheless, the inadequacy of the mover-stayer dichotomization for data analysis can be less severe than would at first appear. As BKM (1955, p. 142) point out, it is not the case that all movers need make a single transition during a time unit, only that they follow a Poisson process with a common λ value. A generalization of the mover-stayer model in which permanent stayers are permitted, as in the original formulation, is discussed in conjunction with the "spiked gamma" (with vodka, please).

¹⁶ Estimates of the main diagonal entries from the present model converge to the equilibrium values predicted by the Markov model. At any finite t , however, estimates from the model herein are greater than the corresponding Markov values.

TABLE 5

POPULATION-TRANSITION MATRICES PREDICTED
FROM THE GENERALIZED MOVER-STAYER MODEL*

$$\hat{P}(1) = \begin{bmatrix} .650 & .156 & .101 & .093 \\ .118 & .719 & .101 & .063 \\ .090 & .102 & .747 & .061 \\ .056 & .064 & .101 & .779 \end{bmatrix}$$

$$\hat{P}(3) = \begin{bmatrix} .423 & .232 & .184 & .162 \\ .175 & .510 & .184 & .131 \\ .150 & .184 & .539 & .127 \\ .113 & .138 & .184 & .565 \end{bmatrix}$$

$$\hat{P}(6) = \begin{bmatrix} .319 & .256 & .229 & .197 \\ .193 & .404 & .229 & .174 \\ .176 & .224 & .428 & .171 \\ .148 & .187 & .229 & .434 \end{bmatrix}$$

* Estimates are from equation (16).

concentration of the simulated data (from table 1, part B). It is apparent that the heterogeneity in the population is reasonably well represented by this gamma density, although many real social processes will actually permit a better fit than the arbitrary distribution constructed here (e.g. Spilerman 1970, table 4).

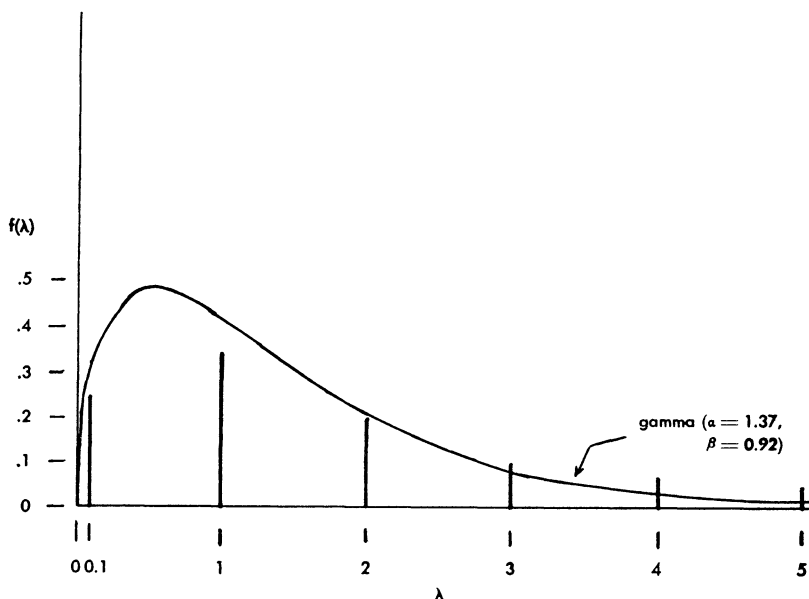


FIG. 1.—Distribution of the population by expected rate of movement, from simulated data.

Spiked Gamma

If the comparison between the negative binomial estimates and the actual distribution of the population by number of moves is less than satisfactory (as judged, for example, by a χ^2 test), the best recourse would be to compound the Poisson with a different family of curves which might allow a more adequate fit to $f(\lambda)$. However, if the gamma assumption fails in the way it is most likely to, by a very heavy concentration of nonmovers, this defect can be repaired by use of the "spiked gamma."

The spiked gamma is a direct generalization of the mover-stayer model, since, unlike the previous extension, the presence of permanent stayers ($\lambda = 0$ persons) is permitted. It is an extension of the mover-stayer model in that heterogeneity is allowed among the movers, who are assumed to be distributed by rate of movement according to a gamma density.¹⁷ One procedure for estimating the parameters of this process would be to first apply a mover-stayer method to separate out stayers from the remainder of the population, then treat movers according to the present model. In fact, having removed stayers, a quick test of the need to even use the model of this paper, rather than the simpler Markov methods, can be obtained by comparing the mean and variance of the distribution of moves. For a Poisson distribution, $\text{var}(v) = E(v)$, while, if heterogeneity is present in the distribution of λ , $\text{var}(v) > E(v)$. Thus, by comparing \bar{v} and S_v^2 , the likely adequacy of a Poisson (mover-stayer) assumption for the movers can be ascertained. (See Spilerman [1970, p. 633] for a lengthier discussion and application of this point.)

An alternate procedure for estimating the size of the spike would be to assume that a gamma density provides the correct distribution of movers and choose that division of the population failing to move which minimizes the deviations of the observed values from the expected distribution of moves. Thus, we would use the gamma to estimate the number of *movers* who happen to make zero transitions during the interval (0,1). The advantage of this approach is that it will allow a "best fit" of the gamma to the distribution of movers who make $v \geq 1$ moves to be obtained, since the term for $v = 0$ does not influence the parameter estimates.

The procedure here is to fit a negative binomial to the observed distribution of moves (table 2, col. 1) except that information about the $v = 0$ term is not used. The "truncated" negative binomial¹⁸ must be employed to

¹⁷ In the BKM mover-stayer model, this gamma specification for movers is replaced by the more restrictive assumption that they are concentrated at a single λ point.

¹⁸ An analogous procedure, using the truncated Poisson distribution (Coleman 1964, p. 366), can be used to estimate the parameters of the BKM mover-stayer model.

estimate α and β when the zero term of the observed distribution is missing. This probability distribution is defined by the equation¹⁹

$$R_v = \frac{1}{1 - p^\alpha} \left(\alpha + \frac{v - 1}{v} \right) q^v p^\alpha \quad (17)$$

for $v \geq 1$, where the two independent parameters α and q are estimated by (see Appendix B)

$$\hat{q} = 1 - \frac{\bar{v}(1 - R_1)}{S_v^2},$$

$$\hat{\alpha} = \frac{1}{\hat{q}} [\bar{v}(1 - \hat{q}) - R_1].$$

In these formulas, \bar{v} and S_v^2 are the sample mean and variance (with the $v = 0$ observation deleted from the computations), and R_1 is the proportion of the observed population with $v \geq 1$ moves who make a single transition. The remaining parameter of the gamma distribution, β , may be calculated from \hat{q} : $\hat{\beta} = (1/\hat{q}) - 1$.

These estimates of α and β are now used with the regular negative binomial formula (eq. 7) to estimate the number of *movers* who failed to move during the interval (0,1). The difference between these observed and calculated numbers, appropriately standardized,²⁰ provides an estimate for the size of the spike at $\lambda = 0$. Calculations using the alternate method were carried out with the data in column 1 of table 2. The results are presented in table 6, column 2, alongside the observed distribution. Because of the added degree of freedom in estimating the spiked distribution, the estimates are clearly superior to those obtained from the regular negative binomial.

With the size of the spike estimated by either method, we have a division of the population into stayers and movers and may project to $P(t)$:

$$P(t) = S + (I - S) \left(\frac{\beta}{\beta + t} \right)^\alpha \left[I - \left(\frac{t}{\beta + t} \right) M \right]^{-\alpha}. \quad (18)$$

In this equation, S is a diagonal matrix containing as entries the proportion

¹⁹ Equation (17) is not written as a function of time since only the values for $t = 1$ are considered here.

²⁰ The negative binomial estimates are standardized by forcing

$$\sum_1^\infty \hat{n}_v = \sum_1^\infty n_v,$$

where n_v is the observed number of persons making v transitions during the time interval (0, 1).

Extensions of the Mover-Stayer Model

TABLE 6

DISTRIBUTION OF NUMBER OF MOVES FROM OBSERVED DATA AND
FROM NEGATIVE BINOMIAL, WITH PARAMETERS ESTIMATED
FROM TRUNCATED NEGATIVE BINOMIAL

v	[1000 $r_v(1)$]	[1000 $\hat{r}_v(1)$]
Number of Moves	Number of Persons with v Moves (Observed Data) (1)	Number of Persons with v Moves ($\alpha=2.231, \beta=1.251$) Calculated from Truncated Negative Binomial (2)
0	388	388*
1	226	224
2	153	160
3	97	101
4	59	58
5	34	32
6	19	17
7	11	9
8	6	5
9	2	2
10	1	1
1,000 $\Sigma r_v(1)$	996†	997†

NOTE.— $\bar{v} = 2.448$; $S_v^2 = 2.776$; the $v = 0$ observation is excluded from these calculations.

* This value includes 162 persons in the spike.

† Value is less than 1,000 because of rounding error.

of the initial population in a state who are stayers;²¹ $I - S$ is a corresponding matrix for movers; and the remaining parameters are estimated as before, but now using only information on movers. The improved predictions for $\hat{P}(t)$, using equation (18), are presented in table 7.

GEOGRAPHIC MIGRATION

Using data made available by Karl Taeuber from his analysis of residential mobility in the United States (Taeuber, Chiazzè, and Haenszel 1968), the model of this paper was applied to interregional transitions by males.²² The Taeuber data were collected in 1958 from retrospective reports about prior residences and are described in detail elsewhere (Taeuber, Chiazzè, and Haenszel 1968). For the purpose of this study, four geographic regions were defined as states of the process: (1) Northeast, (2) North Central,

²¹ Estimates of the proportion of stayers in each state are obtained directly when the mover-stayer estimation procedures are used. With the alternate method, one assumes that stayers constitute an identical proportion of the nonmovers in each state.

²² A computer program which performs the calculations associated with this model has been written by David Dickens. Copies of the program may be obtained from University Microfilms, Ann Arbor, Michigan (request item S-404).

TABLE 7
 TRANSITION MATRICES PREDICTED FROM
 THE SPIKED GAMMA MODEL*

$\hat{P}(1) =$	[.650	.156	.102	.093]
		.118	.718	.102	.063	
		.090	.103	.746	.061	
		.056	.064	.102	.778]
$\hat{P}(3) =$	[.446	.219	.179	.156]
		.166	.529	.178	.128	
		.144	.177	.555	.124	
		.111	.135	.177	.577]
$\hat{P}(6) =$	[.371	.232	.215	.183]
		.175	.448	.213	.165	
		.163	.207	.468	.162	
		.142	.180	.211	.468]

* Estimates are from equation (18).

(3) South, and (4) West. The time points that were used are $t_0 = 1937$, $t_1 = 1944$, $t_2 = 1951$, and $t_3 = 1958$. These were selected to provide residence histories for the adult years of this cohort.

The data proved to be less than ideal for illustrating the versatility of this model to incorporate a wide range of heterogeneity. One difficulty is that the histories were collected only for the four most recent residences of an individual and for his residence at birth. Persons who have had more than five addresses, therefore, have gaps in their residence histories and had to be excluded from the analysis. Unfortunately, this meant that persons with high rates of mobility were deleted and consequently the heterogeneity in proneness to move was being artificially reduced.

A second difficulty with these data stems from the little interregional migration which appears to take place (see table 8, part B). In part, this derives from the truncation of the distribution at $v = 4$ moves, since individuals with many residence changes are most likely to have had some regional migration experience. However, it is also a consequence of the phenomenon we are examining. Apparently, persons do not change geographic region very frequently during a seven-year interval; indeed, only one out of five residence changes resulted in a move to a different region, using this four-category definition of region.

Table 8 presents the observed transition matrix for the population during 1937-44 (part A) and the observed and predicted distributions of the population by number of moves for this period (part B). In light of the above comments, these data pertain only to persons who made four or fewer residence changes during 1937-58. Comparing the observed distribution of moves (part B, col. 1) with the distribution predicted from the negative

Extensions of the Mover-Stayer Model

TABLE 8

OBSERVED POPULATION TRANSITION MATRIX AND DISTRIBUTION OF MOVES FROM GEOGRAPHIC MIGRATION DATA, TOGETHER WITH DISTRIBUTION PREDICTED FROM NEGATIVE BINOMIAL

A. OBSERVED POPULATION TRANSITION MATRIX (1937-44)

$P(1) =$	$\frac{n}{z}$
	.970 .010 .012 .008] 3437
	.007 .947 .015 .030] 4160
	.011 .028 .938 .023] 4110
	.003 .015 .017 .966] 1341

B. DISTRIBUTION OF THE POPULATION BY NUMBER OF MOVES DURING 1937-44

(v)	(n _v)	(n̂ _v)
Number of Moves	Number of Persons with v Moves (1)	Number of Persons with v Moves (Calculated from Negative Binomial, α=1.771, β=6.382) (2)
0	10,120	10,082
1	2,328	2,419
2	507	454
3	93	77
4	12
5	*	1
Σn_v	13,048	13,045

NOTE.— $\bar{v} = 0.278$; $S_v^2 = 0.321$.

* Number not available.

binomial (col. 2), it is evident that the fit is reasonably good, except at the tail end of the distribution. The difficulty at the tail probably results from a tendency to underreport moves when many were made. Remember that we are dealing here with recollections in 1958 of residences during 1937-44.

Using the $P(1)$ matrix from 1937-44 together with $\hat{\alpha}$ and $\hat{\beta}$ from column 2 of table 8, \hat{M} , the estimate of the individual-level transition matrix, was constructed using equation (14):

$$\hat{M} = \begin{bmatrix} .891 & .035 & .046 & .028 \\ .026 & .799 & .059 & .116 \\ .042 & .109 & .763 & .085 \\ .009 & .057 & .065 & .869 \end{bmatrix}$$

Matrix \hat{M} , therefore, indicates how individuals transfer each time they move. It shows, in particular, that the Northeast is most successful in retaining its residents *when they move*; the South, least successful. Using

the estimates of α , β , and M , we can now project to $P(t)$ for any time t . Ordinarily, equation (16) would be used for this purpose. However, to compensate for the effect of truncating the distribution, projection was accomplished using equation (8) with the upper limit of the sum set to four, the maximum number of moves by an individual retained for the analysis. As a result, negative binomial predictions of five or more moves have been deleted from the estimation of $P(t)$. The observed data, and projections obtained by using a Markov chain, as well as those by the present method, are presented in table 9 for the periods 1937-51 and 1937-58.

TABLE 9
OBSERVED AND PREDICTED TRANSITION MATRICES FOR 1937-51 AND 1937-58,
FROM GEOGRAPHIC MIGRATION DATA

	$P(2)$ —(1937-51)	$P(3)$ —(1937-58)
Observed matrices	$\begin{bmatrix} .947 & .015 & .019 & .019 \\ .008 & .923 & .018 & .050 \\ .018 & .045 & .907 & .030 \\ .004 & .023 & .019 & .954 \end{bmatrix}$	$\begin{bmatrix} .934 & .017 & .025 & .022 \\ .009 & .908 & .020 & .063 \\ .020 & .052 & .890 & .038 \\ .004 & .024 & .018 & .954 \end{bmatrix}$
Projection from Markov process $[\hat{P}(t) = P(1)^t]$	$\begin{bmatrix} .942 & .019 & .023 & .016 \\ .014 & .898 & .030 & .059 \\ .021 & .053 & .881 & .044 \\ .005 & .029 & .032 & .933 \end{bmatrix}$	$\begin{bmatrix} .915 & .028 & .034 & .024 \\ .020 & .853 & .043 & .085 \\ .031 & .076 & .829 & .064 \\ .008 & .043 & .047 & .904 \end{bmatrix}$
Projection from present model* ($\alpha = 1.771$, $\beta = 6.382$)	$\begin{bmatrix} .946 & .017 & .021 & .016 \\ .009 & .926 & .023 & .042 \\ .018 & .033 & .912 & .037 \\ .008 & .036 & .023 & .932 \end{bmatrix}$	$\begin{bmatrix} .924 & .024 & .028 & .023 \\ .012 & .897 & .032 & .059 \\ .024 & .046 & .878 & .052 \\ .001 & .050 & .032 & .906 \end{bmatrix}$

* $\hat{P}(t) = \sum_{v=0}^4 \hat{r}(t)M^v$.

By comparing the main diagonal entries, especially for the 1937-58 matrices, it is evident that the model of this paper produces a superior fit to the data than is obtained from the Markov projection, although the latter estimates are themselves not poor. One reason, incidentally, why the predictions from both models are not better is because the stationarity requirement is violated. We are dealing here with a cohort through a 21-year period and extrapolating to a terminal year in which the population is 14 years older than at termination of the period used for parameter estimation. Thus, if age has an effect on migration behavior, as it surely does (e.g., Morrison 1967, pp. 558-59), we have a transition matrix which is changing over time. By comparison, BKM did not have this concern, since their data covered only a three-year period.

While the migration data preclude our demonstrating the suitability of

Extensions of the Mover-Stayer Model

this model in situations of considerable population heterogeneity, we can investigate the reason why the Markov estimates are reasonably good. We are assuming here that transitions are Poisson events in which λ , the parameter of the distribution, varies over individuals. In this circumstance, an estimate of the variance of λ is given by $\hat{\sigma}_\lambda^2 = S_v^2 - \bar{v}$. Substituting the values of S_v^2 and \bar{v} from table 8, we obtain $\hat{\sigma}_\lambda^2 = .043$, a value which is not very large. (By comparison, $\hat{\sigma}_\lambda^2 = 1.635$ for the simulated data in table 2.) Thus, primarily because of the deletion of individuals with more than four moves, little heterogeneity remains in the population, and the Markov chain model which formally requires all persons to have an identical parameter value now provides reasonably good projections. In fact, having estimated α and β , we can graph $f(\lambda)$ to ascertain the appearance of the heterogeneity in λ . This graph is presented as figure 2. Note that the scale

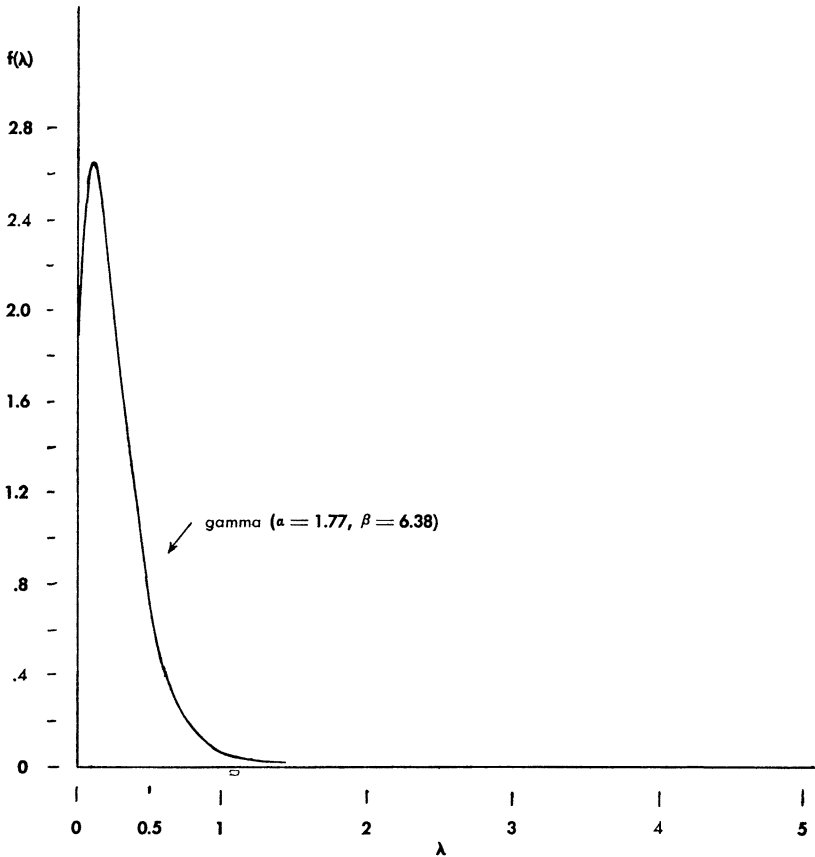


FIG. 2.—Distribution of the population by expected rate of movement, from geographic migration data.

unit on the Y -axis is one-half the size of the corresponding unit in figure 1. It is evident that the population is indeed highly concentrated over a narrow range of λ values.

CONCLUSIONS

The extensions developed in this paper, like the original mover-stayer model, cast the burden of explaining heterogeneity onto variations in the rate of mobility, since, by assumption, all persons follow an identical transition matrix at each move. If individual-level data on social characteristics are available, they can be used to determine the components of heterogeneity. Elsewhere (Spilerman 1970, pp. 646–48), I have argued that a regression methodology, in which the number of moves made by an individual is the dependent variable, is both consistent with this formulation and provides an approximation to analyzing the λ values themselves in terms of the independent variables.

While the emphasis here has been on the analysis of a stationary process, these methods will also shed light on the structure of time-varying processes. Data at two consecutive time points are required for parameter estimation in this model. Therefore, if the parameters are recalculated for adjacent time intervals of a time-varying process, we can ascertain whether the nonstationarity is primarily attributable to changes in the M matrix, which would suggest an alteration in the manner of selecting destination states at a transition, or to change in the gamma distribution, which would be indicative of a shift in the rate at which individuals are making transitions.

It is well known that the negative binomial distribution can be derived from an assumption of positive reinforcement (Coleman 1964, p. 300) as well as from this heterogeneity model. In the context of geographic migration, reinforcement would mean that, with each move, an individual's probability of making a subsequent transition is increased. Although this conceptualization seems forced, it becomes more plausible in an alternate formulation. Making a statement about the process by which moves occur is equivalent to making an assumption about the distribution of durations between the moves. Viewed from the latter perspective, reinforcement would suggest that the longer an individual resides at a particular location, the higher is his probability of remaining. Thus, the reinforcement hypothesis is recognizable as the "Axiom of Cumulative Inertia" in the Cornell Mobility Model (McGinnis 1968).

How do we distinguish between heterogeneity and reinforcement? Conceptually, one or the other is likely to be more appropriate to a particular phenomenon. For example, as McFarland (1970) has pointed out, the assumption that attachments grow over time seems more reasonable for geographic mobility than for occupational mobility. Analytically, it may

be possible to distinguish between these alternative processes by examining the change over time in the distribution of moves in successive time units. The reinforcement model suggests that the variance of this distribution should increase for a cohort as some individuals become increasingly prone to move. By contrast, the heterogeneity explanation suggests that the variance should remain constant. Nevertheless, in many social processes both phenomena probably occur, and individual-level attribute data would seem necessary in order to disentangle their separate effects.

APPENDIX A

$\lim_{t \rightarrow \infty} P(t) = M(\infty)$, The Equilibrium Matrix for M

PROOF

We assume M is ergodic and therefore

$$\lim_{v \rightarrow \infty} M^v = M(\infty)$$

exists (Feller 1957, p. 356). By definition,

$$P(t) = \sum_{v=0}^{\infty} r_v(t) M^v,$$

where $r_v(t)$ is specified by equation (7). For each v ,

$$\lim_{t \rightarrow \infty} r_v(t) = \lim_{t \rightarrow \infty} \left(\alpha + \frac{v-1}{t} \right) \left(\frac{t}{\beta + t} \right)^v \left(\frac{\beta}{\beta + t} \right)^a = 0$$

since $\frac{\beta}{\beta + t} \rightarrow 0$. This implies that for $\epsilon > 0$ and L , an arbitrary integer, there exists a T such that for $t > T$,

$$\sum_{v=0}^{L-1} r_v(t) < \epsilon/2. \tag{A1}$$

Since $\lim_{v \rightarrow \infty} M^v = M(\infty)$, we choose L such that for $v > L$

$$\max_{i,j} \left| m_{ij}^{(v)} - m_{ij}(\infty) \right| < \epsilon/2, \tag{A2}$$

where $m_{ij}^{(v)}$ is the (i,j) entry of M^v . Then, for $t > T$,

$$\max_{i,j} \left| p_{ij}(t) - m_{ij}(\infty) \right| = \max_{i,j} \left| \sum_{v=0}^{\infty} r_v(t) m_{ij}^{(v)} - \sum_{v=0}^{\infty} r_v(t) m_{ij}(\infty) \right|,$$

where $p_{ij}(t)$ is the (i,j) entry of $P(t)$, and the last sum follows from the relation,

$$\sum_{v=0}^{\infty} r_v(t) = 1.$$

Regrouping terms,

$$\begin{aligned} & \max_{i,j} \left| \sum_0^{L-1} r_v(t) [m_{ij}^{(v)} - m_{ij}(\infty)] + \sum_L^{\infty} r_v(t) [m_{ij}^{(v)} - m_{ij}(\infty)] \right| \\ & \leq \sum_0^{L-1} r_v(t) \max_{i,j} |m_{ij}^{(v)} - m_{ij}(\infty)| + \sum_L^{\infty} r_v(t) \max_{i,j} |m_{ij}^{(v)} - m_{ij}(\infty)| \\ & \leq \sum_0^{L-1} r_v(t) \cdot 1 + \frac{\epsilon}{2} \sum_L^{\infty} r_v(t), \end{aligned}$$

since, in the middle line, the maximum in the first summation is less than or equal to one because the matrices are stochastic, and the second maximum has been reduced using (A2). Using (A1) with the first sum in the final inequality and noting that

$$\sum_L^{\infty} r_v(t) \leq 1$$

in the second sum, we conclude that

$$\max_{i,j} |p_{ij}(t) - m_{ij}(\infty)| < \epsilon,$$

and therefore

$$\lim_{t \rightarrow \infty} P(t) = M(\infty).$$

APPENDIX B

The Truncated Negative Binomial Distribution

Let P_v be the negative binomial probability for v events:

$$P_v = \binom{\alpha + v - 1}{v} q^v p^\alpha, \quad v = 0, 1, 2, \dots \tag{B1}$$

For any positive integer-valued distribution, we have

$$1 = \sum_0^{\infty} P_v = P_0 + \sum_1^{\infty} P_v,$$

and therefore,

Extensions of the Mover-Stayer Model

$$\frac{1}{1 - P_0} \sum_1^{\infty} P_v = 1.$$

This is a truncated distribution in that the P_0 term is lacking. For the negative binomial distribution, we obtain

$$R_v = \frac{1}{1 - P_0} P_v = \frac{1}{1 - p^\alpha} \binom{\alpha + v - 1}{v} q^v p^\alpha, \quad v \geq 1, \tag{B2}$$

which is the truncated negative binomial distribution. Note for reference that

$$R_v = \frac{\alpha + v - 1}{v} q R_{v-1}. \tag{B3}$$

TO OBTAIN α

$$\begin{aligned} \mu &= \sum_1^{\infty} v R_v = R_1 + \sum_2^{\infty} v R_v \\ &= R_1 + \sum_2^{\infty} (\alpha + v - 1) q R_{v-1} \quad (\text{by [B3]}) \\ &= R_1 + q \sum_1^{\infty} (\alpha + v) R_v \\ &= R_1 + q\alpha + q\mu, \end{aligned} \tag{B4}$$

and $\alpha = (1/q)(\mu p - R_1)$.

TO OBTAIN q, β

$$\begin{aligned} \text{var}(v) &= \sum_1^{\infty} (v - \mu)^2 R_v \\ &= R_1 - \mu^2 + \sum_2^{\infty} v^2 R_v. \end{aligned}$$

Using (B3), we obtain for the last term,

$$\sum_2^{\infty} v^2 R_v = \frac{q}{1 - q} [\alpha\mu + \alpha + \mu + R_1].$$

Substituting this value into the expression for $\text{var}(v)$, and using (B4) to simplify, yields

$$(1 - q) \text{var}(v) = \mu(1 - R_1),$$

and
$$q = 1 - \frac{\mu(1 - R_1)}{\text{var}(v)}.$$

American Journal of Sociology

Finally, since $q = 1/(\beta + 1)$ (from the definition of $q(1)$, eq. [8] in text), we obtain for β , $\beta = (1/q) - 1$.

REFERENCES

- Bartholomew, D. J. 1967. *Stochastic Models for Social Processes*. New York: Wiley.
- Bellman, Richard E. 1960. *Introduction to Matrix Analysis*. New York: McGraw-Hill.
- 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, no. 6. Ithaca, N.Y.: Cornell University Press.
- Chiang, Chin L. 1968. *Introduction to Stochastic Processes in Biostatistics*. New York: Wiley.
- Coleman, James S. 1964. *Introduction to Mathematical Sociology*. New York: Free Press.
- Feller, William. 1957. *An Introduction to Probability Theory and Its Applications*. New York: Wiley.
- Gantmacher, F. R. 1959. *The Theory of Matrices*. Vol. 1. New York: Chelsea.
- Goldstein, Sidney. 1964. "The Extent of Repeated Migration: An Analysis Based on the Danish Population Register." *Journal of the American Statistical Association* 59 (December): 1121-32.
- Goodman, L. A. 1961. "Statistical Methods for the Mover-Stayer Model." *Journal of the American Statistical Association* 56 (December): 841-68.
- Hodge, R. W. 1966. "Occupational Mobility as a Probability Process." *Demography* 3:19-34.
- Lieberson, S., and G. V. Fuguitt. 1967. "Negro-White Occupational Differences in the Absence of Discrimination." *American Journal of Sociology* 73:188-200.
- Lipset, Seymour M., and R. Bendix. 1959. *Social Mobility in Industrial Society*. Berkeley: University of California Press.
- McFarland, David D. 1970. "Intra-generational Social Mobility as a Markov Process: Including a Time-stationary Markovian Model That Explains Observed Declines in Mobility Rates over Time." *American Sociological Review* 35 (June): 463-76.
- McGinnis, R. 1968. "A Stochastic Model of Social Mobility." *American Sociological Review* 33 (October): 712-22.
- Morrison, Peter A. 1967. "Duration of Residence and Prospective Migration: The Evaluation of a Stochastic Model." *Demography* 4:553-61.
- Palmer, Gladys L. 1954. *Labor Mobility in Six Cities*. New York: Social Science Research Council.
- Rogers, Andrei. 1966. "A Markovian Analysis of Migration Differentials." *Proceedings American Statistical Association*, Social Science Section. Washington, D.C.: American Statistical Association.
- Spilerman, Seymour. 1970. "The Causes of Racial Disturbances: A Comparison of Alternative Explanations." *American Sociological Review* 35 (August): 627-49.
- . 1972. "The Analysis of Mobility Processes by the Introduction of Independent Variables into a Markov Chain." *American Sociological Review* 37 (June): 277-94.
- Tauber, Karl E., L. Chiazze, Jr., and W. Haenszel. 1968. *Migration in the United States*. Washington, D.C.: Government Printing Office.
- Tarver, James D., and William R. Gurley. 1965. "A Stochastic Analysis of Geographic Mobility and Population Projections of the Census Divisions in the United States." *Demography* 2:134-39.