79-70

INSTITUTE FOR RESEARCH ON POVERTY DISCUSSION PAPERS

EXTENSIONS OF THE MOVER-STAYER MODEL

Seymour Spilerman

UNIVERSITY OF WISCONSIN ~ MADISON

79-70

EXTENSIONS OF THE MOVER-STAYER MODEL

by

SEYMOUR SPILERMAN

EXTENSIONS OF THE MOVER-STAYER MODEL

Seymour Spilerman
University of Wisconsin

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 Steve Robinson 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.

November 1970

ABSTRACT

A major drawback to the traditional Markov formulation is that it assumes population homogeneity with respect to transition behavior. This assumption, clearly, is 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. 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 inter-regional migration.

EXTENSIONS OF THE MOVER-STAYER MODEL

Seymour Spilerman

University of Wisconsin

1. 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 required by the Markov model, transitions from a state of origin do not conform to the assumption of population homogeneity. Some persons simply move more often, or differently, from others. This has been found with industrial mobility data (Blumen, Kogan, and McCarthy 1955), with inter-generational and intra-generational occupational mobility (Hodge 1966; Lieberson 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 state changes by a single object in mind. In the analysis of social mobility, however, the movements by 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.

Attempts to relax the homogeneity requirement while retaining the essential Markov framework have led to research in two directions. In one approach, interest has focussed on the construction of subpopulation matrices, and on ways for categorizing 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 constructing 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 the spirit of this approach McFarland (1970) has reported on an analytic method for combining subpopulation or individual-level transition matrices and projecting from these to the k-step population matrix, and Spilerman (in press) has 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 with 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, in press, for a discussion of the convergence of these two perspectives.) Work in this direction has led to the development of the "mover-stayer" model (Blumen, Kogan, and McCarthy 1955). Under the specifications of this process, heterogeneity is contended with 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 intragenerational 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 transitions are available [Goodman 1961]). Since much of our mobility data lack 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 heterogeneity can generally be attributed to two types of persons. In fact, in their concluding chapter BKM (Blumen, Kogan, and McCarthy 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.

2. THE MOVER-STAYER MODEL AND BKM'S COMMENTS ON AN EXTENSION

<u>The mover-stayer model</u>. In their study of industrial mobility BKM (1955) 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

$$P(1) = \begin{pmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & & \vdots \\ p_{m1} & \cdots & p_{mm} \end{pmatrix}$$

is the observed one-step transition matrix and if

$$P(k) = \begin{pmatrix} p_{11}^{(k)} & \cdots & p_{1m}^{(k)} \\ \vdots & & \vdots \\ p_{m1}^{(k)} & \cdots & p_{mm}^{(k)} \end{pmatrix}$$

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 $(\hat{p}_{11}^{(k)} \ldots \hat{p}_{mm}^{(k)})$ which commonly have the property that $\hat{p}_{ii}^{(k)} < p_{ii}^{(k)}$ for $i = 1, \ldots, m$.

Although one might suspect that overtime change in the p_{ij} elements of P(1) are 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 $t = 1, \ldots, k$ are <u>observed</u> one-step transition matrices for successive time intervals, and if $\hat{\hat{P}}(k)$ is the predicted k-step array,

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

then the relationship between the main diagonal elements of $\hat{\hat{P}}(k)$ and P(k) can 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 overtime 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,

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

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 which indicates 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 never moves, and (b) the population which is mobile is homogeneous in its pattern of movement and follows a Markov process. 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 parameters 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 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.

BKM address this problem in their concluding chapter (1955:138-146). 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 each person to make a fixed number of transitions in each time interval, we assume that transitions are random occurrences[1] and that the rate of movement by an individual refers to his expected number of transitions, not to the actual number. This is reasonable since an individual with rate equal to, say, three moves per unit time interval will not necessarily make this 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/time unit. A formal way of stating this is to assume that individuals move in accordance with a Poisson process[2] with parameter value (expected number of moves) $\lambda = 3$.

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

$$P(1) = \sum_{k=0}^{\infty} r_k(1)M^k \qquad (2)$$

where $r_k(1)$ is a Poisson probability, $r_k(1) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$ (for t=1), and indicates the proportion of individuals who are expected to make k transitions during t = 0-1 from among those having a rate equal to λ, and M is the transition matrix followed at each move. For equation (2) to hold it is necessary to also assume that the $r_k(t)$ values are the same for persons in all states of the process.

With the above specification it is easy to show[3] that $P(1) = e^{-\lambda t[I-M]}$, for t = 1 and any matrix M. This result is important because the k-step transition matrix P(k) is now given by

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

Therefore, assuming that individuals are homogenous in their rates of transition, the Markov requirement will be satisfied. Thus, the fact that transitions occur according to a Poisson process does not violate the Markov property.[4]

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 of the population q_1 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 $r_k(t)$ equal the expected proportion of the total population who make k transitions during 0-t, irrespective of the individual mobility rates. Then, combining the coefficients of M^k from the separate processes, we have

$$r_k(1) = \sum_{i=1}^{g} q_i \frac{(\lambda_i t)^k e^{-\lambda_i t}}{k!} \qquad \text{for } t=1 \qquad (3)$$

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

$$r_k(1) = \int_0^\infty \frac{(\lambda t)^k e^{-\lambda t}}{k!} f(\lambda)\, d\lambda \qquad \text{for } t=1 \qquad (4)$$

Equation (4) says that the expected proportion of individuals who make k transitions in the unit time interval 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 k transitions. The summation is taken over all possible values of λ, which is assumed to have a continuous density function.

BKM develop the generalization of the mover-stayer model to this point (as does Bartholomew 1967:27-37 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 the conditions under which the assumptions necessary for the extention to apply will be met.

3. AN EXTENSION OF THE BASIC MODEL

Assumptions and the derivation. In order to extend the mover-stayer model it is necessary to specify the form of $f(\lambda)$ in equation (4). Since we have little a priori knowledge 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 either unimodal or decline exponentially. This seems reasonable since several studies of mobility (Goldstein 1964; Lipset and Bendix 1959:158; Palmer 1954:50; Taeuber et al. 1968:46) 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,

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

where $\Gamma(\alpha) = \int_o^\infty y^{\alpha-1}e^{-y}dy$. 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 1968:49 for details on the integration)--

$$r_k(1) = \int_o^\infty \frac{(\lambda t)^k e^{-\lambda t}}{k!} \frac{\beta^\alpha}{\Gamma(a)} \lambda^{\alpha-1}e^{-\beta\lambda}d\lambda$$

$$= \frac{\Gamma(k+\alpha)}{k!\,\Gamma(\alpha)} \beta^\alpha t^k (\beta+t)^{-(k+\alpha)} \qquad \text{for } t=1 \qquad (6)$$

Using the relation $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$, this result becomes

$$r_k(1) = \binom{\alpha + k - 1}{k}\left(\frac{t}{\beta + t}\right)^k\left(\frac{\beta}{\beta + t}\right)^\alpha \qquad \text{for } t=1 \qquad (7)$$

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 k-moves will satisfy a negative binomial distribution.

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

$$P(1) = \sum_{k=0}^{\infty} r_k(1)M^k$$

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

where $p(t) = \frac{\beta}{\beta + t}$, $q(t) = \frac{t}{\beta + t}$, and $t=1$.

Conditions for a closed form solution to P(1). Recall that for S, a scalar, and any real number α, $(1-S)^{-\alpha}$ has the binomial expansion

$$(1-S)^{-\alpha} = 1 + \alpha S + \frac{\alpha(\alpha+1)}{2!}S^2 + \frac{\alpha(\alpha+1)(\alpha+2)}{3!}S^3 + \ldots = \sum_{o}^{\infty} \binom{k+\alpha-1}{k} S^k \tag{9}$$

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

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

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

Analogous to the condition on S in the scalar case (equation 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. Since M is a transition matrix it is stochastic and its largest eigenvalue equals one. However, $q(t) = \frac{t}{\beta + t} < 1$ for finite t (since $\beta>0$), in particular for $t=1$, and this ensures that the eigenvalues of $q(1)M$ will be strictly less than one.

A second consideration relates to the meaning of $[I - q(t)M]^{-\alpha}$ when α is an arbitrary real number.[5] For an integer k and a non-singular 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 non-singular A (Gantmacher 1959:240).

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

$$A = HDH^{-1} \tag{11}$$

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.[6] In this circumstance $A^{-\alpha}$ can be written[7] as

$$A^{-\alpha} = e^{-\alpha \log A} = He^{-\alpha \log D} H^{-1} = HD^{-\alpha}H^{-1} \tag{12}$$

where $D^{-\alpha}$ is a diagonal matrix with elements of the form $\mu^{-\alpha}$.

Now, substituting the result of equation (10) into equation (8), we obtain for P(1),

$$P(1) = [p(t)]^{\alpha} [I - q(t)M]^{-\alpha} \qquad \text{for } t=1 \tag{13}$$

By the preceding discussion, this result will hold for $q(t) = \frac{t}{\beta + t} < 1$, and in particular for $q(1)$. In practice, letting $[I-q(1)M] = A$, $P(1)$ can be found from M by the diagonalization procedure of equation (12).

Obtaining the M matrix from P(1). Equation (13) 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 (13) may be used to project to $P(t)$ for some $t > 1$.

Fortunately, no difficulty arises in solving equation (13) for M so long as $P(1)$ is non-singular. We obtain

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

Although there is no guarantee that an observed $P(1)$ matrix will be non-singular, this condition is unlikely to be violated in practice. The only instance where a singular $P(1)$ matrix is likely to occur is when the unit time interval is very long with respect to the behavior of the process, so that one's destination state no longer depends upon his origin and the rows of $P(1)$ are therefore identical.[8] In this case, however, $P(t)=P(1)$ for all values of t and the process is not problematic.

In the more common situation where $P(1)$ is non-singular, M is obtained by diagonalizing $P(1)$ in the manner discussed previously:

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

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 $\bar{k}$ and S_k^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 (Chiang 1968:50). This yields

$$\left.\begin{aligned} \hat{\beta} &= \frac{\bar{k}}{S_k^2 - \bar{k}} \\ \hat{\alpha} &= \hat{\beta}\bar{k} \end{aligned}\right\} \qquad (15)$$

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

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

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 1) suggest that $\lim P(t)=M(\infty)$, the equilibrium matrix for M, which may be found by the usual Markov methods.[9]

If we view the process as embedded in continuous time so that non-integer 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 backwards in time. Starting with P(1), for example, we can find P(1/2). 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 $r_k(t)$, the proportion of the population expected to make k transitions by time t, is the same for all system states. BKM (1955:139) 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 an individual is in. The assumption can be tested directly by computing the $r_k(1)$ distribution separately for individuals originating in each of the system states and comparing these distributions.

There are two other basic 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 (4) should be altered and the Poisson compounded with a different curve. 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 inter-arrival times, but they assume an absence of heterogeneity. If a large number of inter-arrival times were available for each person we could test the Poisson assumptions 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, then the comparison between the actual distribution of moves and the negative binomial estimates would provide a test of the Poisson specification. In practice, the form of $f(\lambda)$ would seem to be more problematic and the more interesting question.

4. 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 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 would usually not be available.

Table 1 about here

We further assume that the six types of persons in the population (Panel B of Table 1) move in accordance with a Poisson process which is specified by the indicated λ value for each subpopulation. Consequently, the Poisson distribution was used to generate an expected proportion of each subpopulation who make k = 0, 1, 2,... moves during t_0-t_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 as column (1) of Table 2. Each of the $r_k(1)$ values has been multiplied by 1000 so we can refer to the number of persons making a specified number of moves. These "observed" data were then used with the equation

$$P(1) = \sum_{k=0}^{10} r_k(1)M^k$$

to generate an "observed" transition matrix P(1). This process was repeated[10] for t = 3 and t = 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 (column 1 of Table 2) are the kinds of data which can be collected.[11] We assume now that only this information is available.

Table 1. Structure of the Simulated Data

(a) Individual-level transition matrix

$$M = \begin{pmatrix} .600 & .200 & .100 & .100 \\ .150 & .700 & .100 & .050 \\ .100 & .100 & .750 & .050 \\ .050 & .050 & .100 & .800 \end{pmatrix}$$

(b) Distribution of the population by rate of mobility

λt (t=1)	Proportion of the population with this λ value
0.1	.25
1.0	.35
2.0	.20
3.0	.10
4.0	.06
5.0	.04
	1.00

Tables 2 and 3 about here

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 which was built into the data the main diagonal elements from this projection are, as expected, consistently smaller than the corresponding observed values reported in Table 3. Moreover, the discrepancy increases over time.

Table 4 about here

Turning to the model proposed in this paper, $\bar{k}$ and S_k^2 (from column 1 of Table 2) were first used with equation (15) to estimate α and β, $\hat{\alpha} = 1.371$, $\hat{\beta} = 0.915$. Having estimates of these parameters, the negative binomial formula (equation 7) can be used to generate a predicted distribution of moves. This distribution is presented as column 2 of Table 2, alongside the observed values. While there are some sizable deviations between expected and observed values, this method will usually produce a superior fit to simply dichotomizing the population into stayers and movers,[12] 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),

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

Table 2. Distribution of Number of Moves from Observed (Simulated) Data and from Negative Binomial Estimates

k	(1) [1000 $r_k(1)$]	(2) [1000 $\hat{r}_k(1)$]
Number of moves	Number of persons with k moves (observed data)	Number of persons with k moves (calculated from negative binomial α=1.371, β=.915)
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
1000 Σ $r_k(1)$	996	997
$\overline{k}$	1.498	
s_k^2	3.133	

Table 3. Observed Transition Matrices
for $t_0 - t_1$, $t_0 - t_3$, and $t_0 - t_6$

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

$$P(3) = \begin{pmatrix} .448 & .218 & .178 & .155 \\ .165 & .529 & .178 & .128 \\ .143 & .177 & .555 & .125 \\ .110 & .135 & .178 & .577 \end{pmatrix}$$

$$P(6) = \begin{pmatrix} .365 & .235 & .215 & .184 \\ .177 & .441 & .215 & .166 \\ .164 & .210 & .462 & .164 \\ .141 & .181 & .215 & .463 \end{pmatrix}$$

Table 4. Predicted Population Transition Matrices Using a Stationary Markov Process

---

$$\hat{P}(1) = P(1) = \begin{pmatrix} .650 & .155 & .101 & .093 \\ .118 & .719 & .101 & .063 \\ .090 & .102 & .747 & .061 \\ .056 & .064 & .101 & .779 \end{pmatrix}$$

$$\hat{P}(3) = [P(1)]^3 = \begin{pmatrix} .346 & .261 & .209 & .183 \\ .199 & .446 & .209 & .147 \\ .170 & .209 & .478 & .143 \\ .126 & .154 & .209 & .511 \end{pmatrix}$$

$$\hat{P}(6) = [P(1)]^6 = \begin{pmatrix} .231 & .280 & .264 & .224 \\ .211 & .317 & .264 & .207 \\ .200 & .260 & .337 & .203 \\ .174 & .224 & .264 & .336 \end{pmatrix}$$

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. $\hat{M}$, $\hat{\alpha}$, and $\hat{\beta}$ can now be used with equation (16) to project to P(t) for any value of t. Estimates for P(1), P(3), and P(6) are presented in Table 5.

Table 5 about here

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. 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 concentration of the simulated data (from Panel B of Table 1). 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 (cf. Spilerman 1970).

Figure 1 about here

Table 5. Predicted Population Transition Matrices Using the Generalized Mover-Stayer Model[a]

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

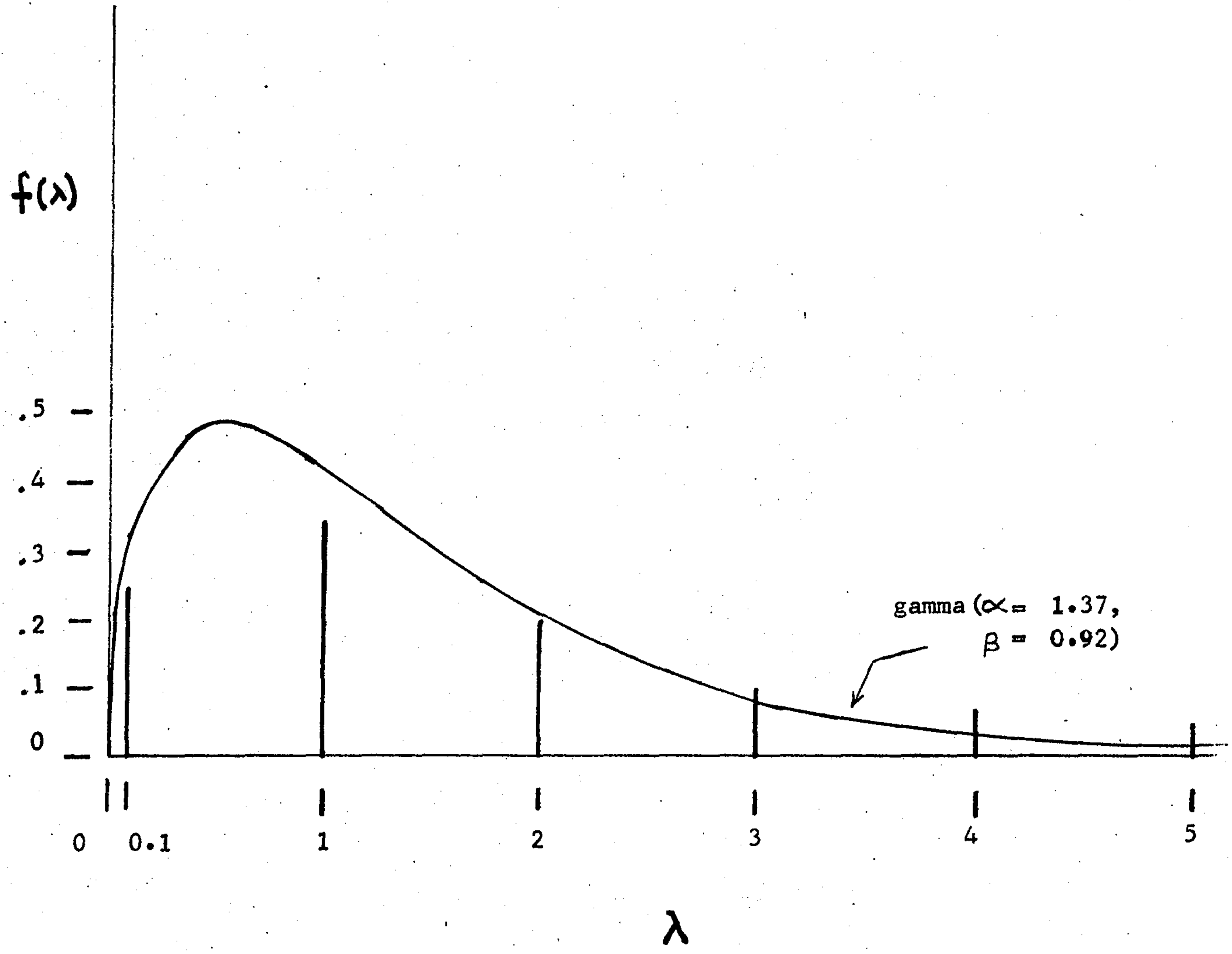
$$\hat{P}(3) = \begin{pmatrix} .423 & .232 & .184 & .162 \\ .175 & .510 & .184 & .131 \\ .150 & .184 & .539 & .127 \\ .113 & .138 & .184 & .565 \end{pmatrix}$$

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

---

[a] Estimates from equation 16.

Figure 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 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 non-movers, 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 genuine 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 according to a gamma density.[13] 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 mover-stayer methods, can be obtained by comparing the mean and variance of the distribution of movers. For a Poisson distribution, $\mathrm{var}(k) = E(k)$, while if heterogeneity is present in the distribution of λ, $\mathrm{var}(k) > E(k)$. Thus, by comparing $\bar{k}$ and S_k^2 for movers the likely adequacy of a Poisson (mover-stayer) assumption for the movers can be ascertained. (See Spilerman 1970 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 who <u>fail</u> to move which minimizes the deviations of the observed values from the expected distribution of movers. Thus, we would use the gamma to estimate the number of movers who happen to make zero transitions during $t_0 - t_1$. The advantage of this approach is that it will allow a "best fit" of the gamma to the distribution of movers who make $k \geq 1$ moves to be obtained, since the term for $k = 0$ does not influence the parameter estimates.

The procedure here is to fit a negative binomial to the observed distribution of moves (Panel A of Table 2) except that information about the $k = 0$ term is not used. The "truncated" negative binomial[14] must be employed to estimate α and β when the zero term of the observed distribution is missing. This probability distribution is defined by the equation[15],

$$R_k = \frac{1}{1 - p^\alpha} \binom{\alpha + k - 1}{k} q^k \, p^\alpha \qquad \text{for } k \geq 1 \qquad (17)$$

where the two independent parameters α and q are estimated by (see Appendix 2),

$$\hat{q} = 1 - \frac{\bar{k}(1-R_1)}{S_k^2}$$

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

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

$$\hat{\beta} = \frac{1}{\hat{q}} - 1$$

These estimates of α and β are now used with the regular negative binomial formula (equation 7) to estimate the number of <u>movers</u> who failed to move during $t_0 - t_1$. The difference between these observed and calculated numbers, appropriately standardized,[16] provides an estimate for the size of the spike at $\lambda = 0$. The calculations using the alternate method were carried out for the data in column 1 of Table 2. The results are presented in column 2 of Table 6 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.

Table 6 about here

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):

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

In this equation S is a diagonal matrix containing as entries the proportion of the initial population in a state who are stayers,[17] I-S is a corresponding matrix for movers, and the remaining parameters are estimated as before, but

Table 6. Distribution of Number of Moves from Observed Data and from Negative Binomial, with Parameters Estimated from Truncated Negative Binomial

k Number of moves	(1) $[1000\ r_k(1)]$ Number of persons with k moves (observed data)	(2) $[1000\ \hat{r}_k(1)]$ Number of persons with k moves (α=2.231, β=1.251 calculated from truncated negative binomial)
0	388	388[a]
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
$1000\ \Sigma\ r_k(1)$	996	997
$\bar{k}$	2.448[b]	
s_k^2	2.776[b]	

---

[a] This value includes 162 persons in the spike.

[b] k = 0 observation is excluded from the calculation.

using now only information on movers. The improved predictions for $\hat{P}(t)$, using equation 18, are presented in Table 7.

Table 7 about here

5. GEOGRAPHIC MIGRATION

Using data made available by Karl Taeuber from his analysis of residential mobility in the United States (Taeuber et al., 1968), the model of this paper was applied to inter-regional transitions by males. The Taeuber data were collected in 1958 from retrospective reports about prior residences and are described in detail elsewhere (Taeuber et al., 1968). For the purpose of this study four geographic regions were defined as states of the process: (1) Northeast, (2) North Central, (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 was 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 residences 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 that the heterogeneity in proneness to move was being artificially reduced.

A second difficulty with these data stems from the little inter-regional migration which appears to take place (see Panel A of Table 8). In part,

Table 7. Predicted Transition Matrices
Using the Spiked Gamma Model[a]

$$\hat{P}(1) = \begin{pmatrix} .650 & .156 & .102 & .093 \\ .118 & .718 & .102 & .063 \\ .090 & .103 & .746 & .061 \\ .056 & .064 & .102 & .778 \end{pmatrix}$$

$$\hat{P}(3) = \begin{pmatrix} .446 & .219 & .179 & .156 \\ .166 & .529 & .178 & .128 \\ .144 & .177 & .555 & .124 \\ .111 & .135 & .177 & .577 \end{pmatrix}$$

$$\hat{P}(6) = \begin{pmatrix} .371 & .232 & .215 & .183 \\ .175 & .448 & .213 & .165 \\ .163 & .207 & .468 & .162 \\ .142 & .180 & .211 & .468 \end{pmatrix}$$

---

[a] Estimates from equation 18.

this derives from the truncation of the distribution at k = 4 moves since the individuals with many residence changes are most likely to have 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 about here

Table 8 presents the observed transition matrix for the population during 1937-44 (Panel A) and the observed and predicted distributions of the population by number of moves for this period (Panel 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 (column 1 of Panel B) with the distribution predicted from the negative binomial (column 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--

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

Table 8. Observed Population Transition Matrix and Distribution of Moves for Geographic Migration Data, Together with Distribution Predicted From Negative Binomial

(a) Observed Population Transition Matrix (1937-44)

$$P(1) = \begin{pmatrix} .970 & .010 & .012 & .008 \\ .007 & .947 & .015 & .030 \\ .011 & .028 & .938 & .023 \\ .003 & .015 & .017 & .966 \end{pmatrix} \quad \begin{array}{c} n_k \\ \hline 3437 \\ 4160 \\ 4110 \\ 1341 \end{array}$$

(b) Distribution of the Population by Number of Moves during 1937-44

	(1)	(2)
(k)	(n_k)	($\hat{n}_k$)
Number of moves	Number of persons with k moves	Number of persons with k moves (calculated from negative binomial, α=1.771, β=6.382)
0	10120	10082
1	2328	2419
2	507	454
3	93	77
4		12
5	---[a]	1
Σn_k	13048	13045
$\bar{k}$	.278	
s_k^2	.321	

---

[a] Number not available.

$\hat{M}$ therefore indicates how individuals transfer each time they move. It shows, in particular, that the Northeast is most successful in retaining its residents <u>when they move</u>, 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 by using equation (8) with the upper limit of the sum set to four, the maximum number of moves by an individual retained for 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 about here

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 very poor. One reason, incidentally, why the predictions from both models are not even 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 1967:558-559), we have a transition matrix which is changing over time. BKM, by comparison, did not have this concern since their data covered only a three year-period.

Table 9. Observed and Predicted Transition Matrices for 1937-51 and 1951-58

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

---

[a] $\hat{P}(t) = \sum_{k=0}^{4} \hat{r}(t)M^k$

While the geographic data preclude our demonstrating the suitability of this model in situations of considerable population heterogeneity, we can investigate the reason why the Markov estimates are reasonably good. Comparing $\overline{k}$ and S_k^2 from Table 8 we see that the difference between these values, equal to .043, is not very large. For a Poisson process in which λ, the parameter of the distribution, is varying, an estimate of the variance of λ is given by

$$\hat{\sigma}_\lambda^2 \doteq S_k^2 - \overline{k} = .043$$

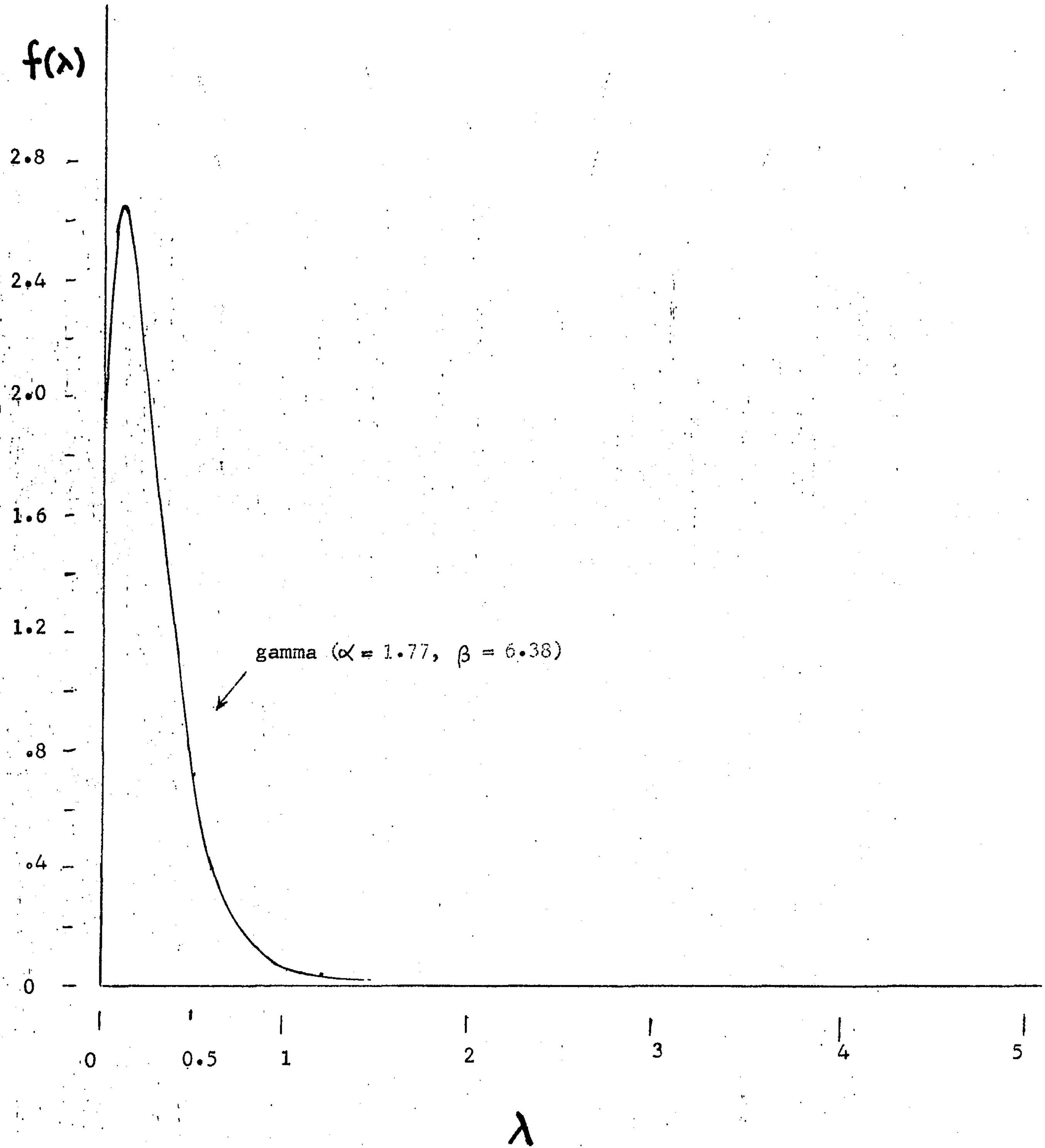
Thus, primarily because of the deletion of individuals with more than four moves, little heterogeneity is present in the population, and the Markov chain model, which formally requires all persons to have an identical parameter value, now provides reasonably good projections. In fact, having estimated 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 unit on the Y-axis is one-half the size of the corresponding unit in Figure 1. It is therefore evident that the population is indeed highly concentrated over a narrow range of λ values.

Figure 2 about here

6. CONCLUSIONS

The extensions developed in this paper, like the original mover-stayer model, cast the burden of explaining the heterogeneity onto variations in the rate of mobility since, by assumption, all persons follow an identical

Figure 2. Distribution of the Population by Expected Rate of of Movement, From Geographic Migration Data



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) 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 non-stationarity is attributable primarily to changes in the M matrix, which would suggest a change in the manner of selection of destination states at a transition, or to alterations in the gamma distribution, which would be indicative of a change 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 1964:300) 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 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 over-time change in the distribution of moves in successive time units. The reinforcement model suggests that the variance of this distribution should increase over time 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 data would seem necessary in order to disentangle their separate effects.

APPENDIX 1. $\lim_{t \to \infty} P(t) = M(\infty)$, THE EQUILIBRIUM MATRIX FOR M

Let t be variable in equation (8). Taking the limit as t → ∞ we have

$$\lim_{t \to \infty} P(t) = \lim \sum_{k=0}^{\infty} r_k (t) M^k$$

For every finite k,

$$\lim_{t \to \infty} r_k(t) = \lim \binom{\alpha + k-1}{k} \left(\frac{t}{\beta + t}\right)^k \left(\frac{\beta}{\beta + t}\right)^\alpha$$

$$= r_k = 0$$

since $\frac{\beta}{\beta + t} \to 0$. However, $r_k(t)$ is a probability distribution for all t and therefore $\sum_{k=0}^{\infty} r_k = 1$. This implies that the non-zero values of the distribution are concentrated at $k = \infty$ and consequently $\sum_{k=L}^{\infty} r_k = 1$ for arbitrarily large L. Choose L sufficiently large so that $M^L = M(\infty)$ to any desired degree of accuracy. We then have $\lim_{t \to \infty} P(t) = \sum_{k=L}^{\infty} r_k M^k = M^L \Sigma r_k = M(\infty)$.

APPENDIX 2. THE TRUNCATED NEGATIVE BINOMIAL DISTRIBUTION

Let P_k be the negative binomial probability for k events:

$$P_k = \binom{\alpha + k - 1}{k} q^k \; p^\alpha \qquad k = 0,1,2,\ldots \qquad \text{(A2-1)}$$

For any positive integer-valued distribution we have $1 = \sum_o^\infty P_k = P_o + \sum_1^\infty P_k$, and therefore $\frac{1}{1-P_o} \sum_1^\infty P_k = 1$. This is a truncated distribution in that the P_o term is lacking. For the negative binomial distribution we obtain

$$R_k = \frac{1}{1-P_o} P_k = \frac{1}{1-p^\alpha} \binom{\alpha + k - 1}{k} q^k \; p^\alpha, \quad k \geq 1 \qquad \text{(A2-2)}$$

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

$$R_k = \frac{\alpha + k - 1}{k} qR_{k-1} \qquad \text{(A2-3)}$$

<u>To obtain α.</u>

$$\mu = \sum_1^\infty k \; R_k = R_1 + \sum_2^\infty k \; R_k$$

$$= R_1 + \sum_2^\infty (\alpha + k - 1)q \; R_{k-1} \qquad \text{(by A2-3)}$$

$$= R_1 + q \sum_1^\infty (\alpha + k)R_k$$

$$= R_1 + q\alpha + q\mu \qquad \text{(A2-4)}$$

and $\alpha = \frac{1}{q} (\mu p - R_1)$

To obtain q, β.

$$\mathrm{var}(k) = \sum_{1}^{\infty} (k - \mu)^2 R_k$$

$$= R_1 - \mu^2 + \sum_{2}^{\infty} k^2 R_k.$$

Using (A2-3) we obtain for the last term,

$$\sum_{2}^{\infty} k^2 R_k = \frac{q}{1-q} [\alpha\mu + \alpha + \mu + R_1]$$

Substituting this value into the expression for var(k) and using (A2-4) to simplify yields,

$$(1 - q)\ \mathrm{var}(k) = \mu\ (1 - R_1)$$

and $q = 1 - \dfrac{\mu(1 - R_1)}{\mathrm{var}(k)}$

Finally, since $q = \dfrac{1}{\beta+1}$ (from the definition of q(t), equation (8) in text), we obtain for β, $\beta = \dfrac{1}{q} - 1.$

NOTES

[1] A transition from i → i would be considered movement in this terminology. For example, if the states of the system were geographic regions, an i → i transition would represent intra-regional movement. Alternatively, we might speak of the expected number of exposures to movement as do BKM (1955:139), but allow an individual to not move (an i → i transition) at an exposure.

[2] A precise specification of the (time-homogeneous) Poisson process is given by the following four assumptions:

(i) λ is constant over time.

(ii) In an infinitesimal time interval Δt at most one event can occur.

(iii) The probability of an event in Δt equals $\lambda \Delta t$, the probability of no event in Δt equals $1-\lambda \Delta t$.

(iv) 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 1957:400).

[3]

$$P(t) = \sum_{k=0}^{\infty} r_k(t) M^k = \sum_{o}^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!} M^k$$

$$= e^{-\lambda t} \sum_{o}^{\infty} \frac{(\lambda t M)^k}{k!} = e^{-\lambda t} e^{\lambda t M} = e^{-\lambda t [I-M]}$$

The convergence of the infinite sum to $e^{\lambda t M}$ will hold for an arbitrary matrix M.

[4] In fact, the assumption that the occurrence of transitions follows a Poisson process leads to a <u>continuous-time</u> Markov formulation.

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

[6] 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 Jordon form (Bellman 1960:191) 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.

[7] 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.

[8] This statement assumes an absence of stayers (λ=0 individuals).

[9] This assumes an absence of λ=0 individuals. If stayers are present then $\lim P(t)$ is given by the mover-stayer formulation (BKM 1955:111-114), $\lim P(t) = S + (I-S)M(\infty)$.

[10] The upper limit on the summation was set to 30 for $t = 3$, and to 60 for $t = 6$.

[11] We also have available the distribution of $r_k(t)$ for $t = 3, 6$, analogous to column 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 over-time change in the distribution.

[12] The inadequacy of the mover-stayer dichotomization for data analysis can be less severe than would at first appear. As BKM(1955: 142) point out it is not the case that movers must each make a single transition during a time unit, only that they follow a Poisson process with a common λ value. A direct generalization of this conceptualization of the mover-stayer model is discussed shortly in conjunction with the "spiked gamma" (with vodka, please).

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

[14] An analogous procedure, using the truncated Poisson distribution (Coleman 1964:366), may be used to estimate the parameters of the mover-stayer model.

[15] Equation (17) is not written as a function of time since only the values for t=1 are considered.

[16] The negative binomial estimates are standardized by forcing $\sum_1^\infty \hat{N}_k = \sum_1^\infty N_k$, where N_k is the observed number of persons making k transitions during $t_0 - t_1$.

[17] 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.

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, vol. 6. Ithaca, New York: Cornell University.

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 59:1121-1132.

Goodman, L.A.

1961 "Statistical methods for the mover-stayer model," Journal of the American Statistical Association 56:841-868.

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 35:463-476.

McGinnis, R.

1968 "A stochastic model of social mobility." Americal Sociological Review 33:712-722.

Morrison, Peter A.

1967 "Duration of residence and prospective migration: The evaluation of a stochastic model." Demography 4:553-561.

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.

Spilerman, Seymour

1970 "The causes of racial disturbances: A comparison of alternative explanations." American Sociological Review, 35:627-649.

In Press "The analysis of mobility processes by the introduction of independent variables into a Markov chain." Submitted to American Sociological Review.

Taeuber, Karl E., L. Chiazze, Jr., and W. Haenszel

1968 Migration in the United States. Washington, D.C.:U.S. 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 2:134-139.