331-76

INSTITUTE FOR RESEARCH ON POVERTY DISCUSSION PAPERS

SOME METHODOLOGICAL ISSUES IN THE ANALYSIS
OF LONGITUDINAL SURVEYS

Burton Singer and
Seymour Spilerman

UNIVERSITY OF WISCONSIN-MADISON

Contents

1. INTRODUCTION

In recent years there has been a considerable expansion in the
availability of longitudinal data files. Sociological theory has always
had the study of social change as its core, yet the majority of quanti-
tative empirical researches have involved the analysis of cross-sectional
data. Longitudinal studies, in particular multi-wave panel studies, have
not been very common. In part, this is because of the considerable cost
involved in surveying a population sample at multiple points in time. It
is also due to the fact that several years must usually elapse after the
first interview for the longitudinal aspect of the data to become suffi-
ciently detailed so that patterns of change can be detected and studied.
However, stimulated by a recent concern with the development of social
indicators and by a related interest in social experimentation, a number
of large scale studies have been funded, and sufficient time has elapsed
for these investigations to have produced longitudinal files. Indeed, in
comparison with even a decade ago, we appear to be moving into an era
which will be comparatively rich in the existence of multi-wave panel data
on large population samples. Important examples of currently available
data sets of this sort are the Michigan Panel Study on Income Dynamics
(Morgan and Smith, 1969), the National Longitudinal Study of Labor Force
Experience (Parnes Study, 1972), the Sewell-Hauser Panel on Wisconsin
Youth (1975), and files from several negative income tax studies (e.g.,
A. Rees and H. Watts, 1976).

The expansion in availability of these sorts of files raises ques-
tions about proper analytic methodology for exploiting the richness and
unique properties of panel data, especially in instances where more than
two waves of interviews have occurred. Sociologists frequently ask

questions about <u>distributional</u> change and are interested in forecasting
the evolution of a population among system states, as well as in under-
standing the structure of the dynamic process. The most common examples
of such studies concern occupational mobility (e.g., Lieberson and Fuguitt,
1967; Hodge, 1966) and geographic migration (e.g., Tarver and Gurley, 1965;
Rogers, 1966). Some economists (e.g., Smith and Cain, 1967; McCall, 1973)
have viewed income dynamics from the same perspective.

The mathematical framework that has been used in these investigations
is discrete-time Markov chains. We shall discuss a number of limitations of
this structure as a description of social processes; at this point, though, we
wish only to motivate our investigation by focusing on one discrepancy
between forecasts from a Markov model and observations on the empirical
process. In applications of Markov chains to industrial mobility, Blumen,
Kogan, and McCarthy (1955) (hereafter referred to as BKM) discovered an
empirical regularity which has subsequently been observed in many other
sociological investigations and which has motivated a rich and diverse
research effort. In particular, they noted the tendency for the main
diagonal entries of observed stochastic matrices to be underpredicted by
the main diagonal entries in powers of one-step Markov transition matrices.
This has led to the formulation of a variety of alternative stochastic
process models which might plausibly account for the regularity. Futher-
more, there has been a critical reevaluation of the substantive and
statistical issues involved in estimation and comparison of several models
fitted to the rather fragmentary longitudinal data which is usually avail-
able on an empirical process.

The purpose of this paper is to review some of the methodological
developments which were an outgrowth of BKM's pioneering investigation.

Particular attention will be paid to parsimony of models relative to multi-wave panel data, and to the testing and identification of multiple models which may be compatible with a given set of observations. In Section 2 we review BKM's study and describe some conceptual difficulties which can arise when discrete-time structures are applied to social processes that evolve continuously in time. Section 3 contains an overview of the alternative explanations which have been proposed to account for the empirical regularity observed by BKM; namely, the underprediction of diagonal entries in observed transition matrices by diagonal entries in powers of Markov transition matrices. Models of heterogeneous populations which extend BKM's formulation to continuous time, and which incorporate more diverse forms of heterogeneity, are described in Section 4. In Section 5 we illustrate the companion issues of embeddability and identification for continuous-time Markov chains. This is the prototype of a set of methodological problems which are central to the analysis of panel data, and which have received remarkably little attention.

Generally speaking, embeddability tests refer to the task of ascertaining whether or not an empirical process is compatible with the conceptual assumptions (mathematical structure) underlying a particular class of models (e.g., general Markov, mixtures of Markov, semi-Markov). Where the answer is affirmative, identification procedures refer to techniques for recovering the specific set of structural parameters from the model class which should be associated with the empirical process. One indication of the difficulties involved with identification is the fact that empirically determined stochastic matrices based on data collected at evenly spaced time points may be embeddable in the class of continuous-time Markov models, but a unique structure from that class may not be identifiable.

Finally, Section 6 illustrates a rudimentary strategy for discriminating among four classes of stochastic process models using multi-wave panel data. That discussion is intended to illustrate the flavor of the kinds of strategies which are in serious need of development. Indeed this is the place where the greatest methodological challenges lie, and foremost among them is the specification of designs for panel studies which will facilitate discrimination among multiple plausible models.

2. MOVERS AND STAYERS--A REVIEW

2.1 Model specifications and an empirical regularity

In attempting to describe the propensity of persons in particular age and sex cohorts to move between pairs of industrial categories, BKM first fit a discrete-time Markov chain with stationary transition probabilities to quarterly data on the occupations of persons listed in the Social Security Administration's Work History File (1972). By a discrete-time Markov chain we mean a stochastic process $\{X(k),\ k = 0,1,2,\ldots\}$ describing state transitions by an individual where the system states might be geographic regions, occupations, industries, or income categories, depending on the particular substantive problem. Probability statements about the process are governed by the analytical recipe

$$\text{Prob}\{X(k+n) = j \mid X(0),\ X(1),\ldots,X(n-1),\ X(n) = i\}$$
$$= \text{Prob}\{X(k+n) = j \mid X(n) = i\} = m_{ij}^{(k)} \qquad (2.1)$$

for $k = 0,1,2,\ldots$ $n = 0,1,2,\ldots$. Thus, the fundamental assumption of a Markov process is that future system state is not a function of past history, once current state is specified. The element $m_{ij}^{(k)}$ is the $(i,\ j)$

entry in the stochastic matrix M^k (k-fold matrix multiplication of M).

This specifies the k-step transition matrix under a Markov chain, i.e.,

$P(0, k) = M^k$. M is itself a stochastic matrix whose entry m_{ij} has the

interpretation,

m_{ij} = probability that an individual in category i will
move to category j in one unit of time.

This mathematical structure describes the evolution of a <u>homogeneous</u>

population, because it is assumed that all individuals evolve according

to the same transition mechanism (namely, the matrix M).

BKM's estimation method was simply to identify an average of the

observed one-quarter (i.e., three-month interval) transition matrices

with the matrix of one-step Markov chain transition probabilities

$||m_{ij}||$. With this estimate in hand, they tested the model by compar-

ing M^k with $\hat{P}(0, k)$, the empirically determined transition matrix[1]

based on observations taken at the beginning of the initial quarter and

at the end of the kth quarter. BKM carried out this comparison for

k = 4, 8, and 11 and found that

$$\hat{P}_{ii}(0, k) > m_{ii}^{(k)}, \qquad k = 4, 8, 11; 1 \le i \le r = \text{number of states;} \qquad (2.2)$$

that is, the main diagonal elements in the k-step matrix predicted by

a Markov process under-represent the main diagonal elements in the

observed k-step matrix. They also noted that the magnitude of the

inequality increased together with k.

BKM suggested that one plausible explanation for the discrepancy

summarized in (2.2) was that a socially <u>heterogeneous</u> population was

being treated as though it was homogeneous. They proposed an alterna-

tive model to accommodate heterogeneity in which the population was

viewed as consisting of two kinds of individuals. They assumed that a non-directly observable fraction s_i of the individuals in industry category i--called stayers--never moved, and that their evolution was described by the degenerate Markov chain $\{X_1(k), k = 0,1,2,...\}$ with one-step transition matrix given by the identity I. In addition, the evolution of a non-directly observable fraction, $1-s_i$, of the individuals--called movers--who were in industry category i at the beginning of the initial quarter was described by a discrete-time Markov chain $\{X_2(k), k = 0,1,2,...\}$ with one-step transition matrix M. The diagonal entries m_{ii} were not required to be zero, thereby allowing for within-industry job change. It was also assumed that the mover population evolved independently of the stayers, and that the same transition matrix M governed the evolution of movers who started in each category at the beginning of the initial quarter.

The _observable process_ $\{Z(k), k = 0,1,2,...\}$ describing the evolution of individuals who start out in each industry category in the initial quarter is thus a mixture of the components of the bivariate process $(X_1(k), X_2(k))$. Its transition probabilities are given by

$$\text{Prob}\{Z(k) = j \mid Z(0) = i\} = \begin{cases} (1-s_i)m_{ij}^{(k)} & \text{for } i \neq j \\ s_i + (1-s_i)m_{ii}^{(k)} & \text{for } i = j \end{cases} \quad (2.3)$$

$$k = 1,2,...; \quad 1 \leq i, j \leq r = \text{number of states.}$$

In matrix notation, this may be written as

$$P(k) = SI + (I-S)M^k$$

$$\text{where } S = \begin{pmatrix} s_1 & & 0 \\ & \ddots & \\ 0 & & s_r \end{pmatrix}.$$

The formulation (2.3) has come to be known as the "mover-stayer" model, and a variety of simultaneous estimation methods for the structural parameters $(s_1, \ldots, s_r)$ and M are given in a paper by L. Goodman (1961), who improved considerably on BKM's initial procedures. BKM found that this model of a heterogeneous population provided a better description of job mobility, as measured by the quarterly observations, than the original Markov chain model of a homogeneous population. Furthermore, the mover-stayer model accounted for much of the empirical regularity (2.2) and thus has motivated subsequent attempts to develop more refined models of heterogeneous populations.

2.2 A difficulty with discrete-time models

Despite the initial success of the mover-stayer formulation there are conceptual difficulties with the basic strategy of fitting discrete time models to mobility data. In particular, when structural[2] information about a population is the primary goal of an investigation, then the substantive interpretation attached to estimates of the matrix M--in either the pure Markov or mover-stayer model--is

$$m_{ij} = \text{probability that an individual in state } i \text{ will move to state } j \underline{\text{ when a change occurs}} \text{ .}$$

If you regard M as a matrix of structural change parameters and fit discrete-time models to evenly spaced observations, then you are tacitly assuming that the natural time unit between, say, industry or occupational changes coincides with the sampling interval (3 months in the Social Security Administration's Work History File). Since there is no substantive basis for such an identification, the parameters estimated by BKM cannot legitimately be interpreted as structural information

about the population of workers; alternate choices of the sampling

interval will yield different matrices M. Indeed, BKM were aware of

this difficulty and noted that during a given quarter some persons

will have moved twice, others will have moved three times, etc. For

these unidentifiable persons you are really estimating M^2, M^3, etc.

Nevertheless, even by dropping any attempt to identify M as a matrix

of structural parameters and just fitting a discrete-time model to

quarterly data, BKM found an empirical regularity of considerable impor-

tance. In fact, as we will indicate in Section 4, even when continuous-

time Markov models--whose parameters can legitimately be interpreted

as structural coefficients[3]--are fit to a variety of longitudinal data

sets, the regularity observed by BKM still appears.

The ambiguity in specifying an appropriate time scale for intra-

generational mobility processes has also been pointed out by H. White

(1970, pp. 319-320) and Singer and Spilerman (1974, pp. 360-362).

However, a facet of this ambiguity which seems to have been overlooked

by BKM, as well as by subsequent users of the mover-stayer formulation

(e.g., McCall, 1973), is the fact that conclusions about compatibility

of data with a discrete-time model can depend entirely on an ad-hoc

choice of unit-time interval. To see this in the simplest possible

setting, recall BKM's initial fitting of a discrete-time Markov chain

to quarterly observations.

Suppose, for illustrative purposes, that you agree that a natural

time unit for job mobility in a particular population cohort is six

weeks. Then an attempt to fit an observed one-quarter (12-week) transi-

tion matrix $\hat{P}(0, 1)$ to a Markov chain consists of asking whether there

exists a stochastic matrix M such that

$$\hat{P}(0, 1) = M^2.$$

An affirmative answer would require that $\hat{P}(0, 1)$ have at least one stochastic square root, $\sqrt{\hat{P}(0, 1)}$. That this is by no means automatic can be seen if you consider a two-state process with observed one-quarter transition matrix

$$\hat{P}(0, 1) = \begin{pmatrix} 1/4 & 3/4 \\ 5/8 & 3/8 \end{pmatrix}.$$

This matrix has no stochastic square roots, and it is therefore incompatible with a discrete-time Markov structure if the natural time unit is believed to equal six weeks. However, if you use a four-week time unit then you find that $\hat{P}(0, 1)$ does have a stochastic cube root given by

$$\sqrt[3]{\hat{P}(0, 1)} = \begin{pmatrix} .0611 & .9839 \\ .7824 & .2176 \end{pmatrix}$$

More generally, $\hat{P}(0, 1)$ has no stochastic roots of any even order, while it does have a stochastic cube root, a stochastic fifth root, but no odd stochastic root of order greater than five.

A consideration of high order roots (say, greater than four) is not really an issue with quarterly observations of job mobility; however, it certainly could be for annual observations or more widely spaced data. The essential point to be made here, however, is that for processes such as intragenerational occupational mobility which are both intrinsically nonsynchronous[4] and lack any substantive basis for a choice of unit time interval, a more natural strategy is to fit continuous-time models

(in which the waiting times between moves are viewed as random variables) to the data, and carry out systematic discrimination among alternative models in that setting.[5] This kind of extension of the mover-stayer framework was first carried out by S. Spilerman (1972a) with further generalizations indicated in Singer and Spilerman (1974). These developments will be reviewed together with a variety of other models in Section 4.

3. OTHER EXPLANATIONS OF HIGH DIAGONALS

BKM's introduction of the mover-stayer model to explain "clustering on the main diagonal," i.e., the empirical regularity[6]

$$\hat{p}_{ii}(0, k) > m_{ii}^{(k)}, \qquad i = 1,\ldots,r; \quad k = 2,3,\ldots \qquad (3.1)$$

has led to the development of a variety of qualitatively different kinds of models, all capable of accounting for (3.1). The five principal features of social processes which are not taken into account in uni-variate time-stationary Markov models[7] and which have motivated the construction of alternative models are:

  (i)   population heterogeneity

  (ii) time-varying propensities to change system states (e.g., income categories, occupations, industries)

  (iii) non-exponential waiting times between changes of state

  (iv) strong dependence on past history

  (v)   latent variables.

Features (i), (ii), and (iii) have received the most attention in attempts to develop stochastic process models which can account for (3.1) and which also mirror other widely observed empirical phenomena, such as

the increasing propensity with the passage of time for persons in a
particular occupation to remain there. For a nice empirical study of
manpower flows in British labor markets where this behavior occurs,
see Kuhn, Poole, Sales, and Wynn (1973). Since our primary concern in
Sections 4-6 will be with specification, estimation, and identification
issues involving models based on (i)-(iii), a few remarks about (iv)
and (v) are in order.

In a review of BKM's study, W. Feller (1956) suggested that for
processes such as job mobility, dependence on past behavioral patterns
was probably so pronounced that it would be essential to develop detailed
models incorporating past history in order to have a satisfactory
description of the observed empirical patterns. Indeed, Feller suggested
the use of higher order Markov processes for this purpose. As a strategy
for understanding social phenomena such as mobility among occupa-
tion, industry, or income categories, this kind of program has never
been seriously followed up and has in fact been criticized on several
grounds. Coleman (1964a, pp. 9-11), in particular, has emphasized that
the intrinsically heterogeneous nature of most populations is largely
ignored by an introduction of higher order Markov models, and that such
an exercise is more akin to blind curve fitting of successively higher
order polynomials to irregular data.

One might argue that models incorporating both heterogeneity and
long range dependence should be introduced; however, the fragmentary
nature of the data which can be collected in most surveys--particularly
the small number of time points at which persons involved in panel studies
can be re-interviewed--makes judgments as to the relative importance of
phenomena which are to be incorporated in parsimonious models essential.

In fact, a primary reason for the emphasis on population heterogeneity and the neglect of long range dependence is the greater importance for the development of sociological theory attached to an understanding of the components of heterogeneity. The strategies of introducing independent variables into Markov chain models developed by Coleman (1964a), McFarland (1970), and Spilerman (1972b) as well as the mixture models introduced in Spilerman (1972a) and Singer and Spilerman (1974) are all based on considerations of parsimony of models relative to the available data and on the judged importance of population heterogeneity.

Concerning item (v), many of the observed attitudinal responses in panel studies, such as opinions about political issues, career aspirations, etc. are related to a variety of non-directly observable (or latent) social and psychological variables. In addition, there are often several competing theories about the relationships which may exist between latent and manifest (i.e., observable) variables. An important research objective with panel data is to discriminate among dynamic models incorporating a variety of latent and manifest variable relationships. Despite its importance, this aspect of the analysis of longitudinal surveys is largely undeveloped. The major attempts to consider both the substantive and methodological issues have been by Coleman (1964a), Lazarsfeld and Henry (1968), and Wiggins (1973). The last of these contains a superb collection of examples and lucid statements on the enormous range of unresolved mathematical, statistical, and social-theoretic problems. In the remainder of this paper we will concentrate on models which incorporate population heterogeneity, time-varying propensities to change state, and general classes of waiting times between moves. However, it should be noted that the same

methodological issues arise in dealing with latent structure models but with a considerable increase in complexity.

4. PARSIMONIOUS MODELS AND FRAGMENTARY DATA

In the context of panel studies, J. Coleman (1964b) introduced continuous-time Markov chains as an initial baseline class of models. However, in fitting these models to observed data, he noted the same kind of empirical regularity--under-prediction of diagonals of observed matrices--which BKM and others had found using discrete-time models. This finding has motivated the development of a variety of formal models of heterogeneous populations which are both moderately realistic and simple enough so that parameters can be estimated and the models falsified using rather fragmentary data. The strategies for introducing heterogeneity have basically been of two distinct types: individuals (or sub-populations) are classified either according to the rate at which they move (Spilerman, 1972a; Singer and Spilerman, 1974) or according to their propensity to move between pairs of states when a transition occurs (McFarland, 1970; Spilerman, 1972b; Singer and Spilerman, 1974). These sub-populations are not always directly observable, and mixtures of Markov and semi-Markov processes provide simple, readily interpretable models of the observed population-level processes. Explicit descriptions of models of these types, suited to intra-generational mobility studies, are given in sub-sections 4.1, a-d below.

4.1 Model specifications

In order to illustrate some explicit models of heterogeneous populations and clarify the substantive assumptions which accompany their

use, we first recall the basic mathematical structure of continuous-time

Markov chains with stationary transition probabilities. In particular,

consider a stochastic process with a finite number of states whose

transition probabilities are governed by the system of ordinary differ-

ential equations

$$\frac{dP(t)}{dt} = QP(t), \qquad P(0) = I \qquad (4.1)$$

where $P(t)$ and Q are rxr matrices. It is well-known (Coleman (1964b),

pp. 127-130; Chung (1967), pp. 251-257) that if Q has the structure

$$q_{ij} \geq 0 \text{ for } i \neq j, \quad q_{ii} \leq 0, \quad \sum_{j=1}^{r} q_{ij} = 0, \quad i = 1,\ldots,r \qquad (4.2)$$

then the functions $P(t)$, $t > 0$ which are solutions of (4.1) comprise

the transition matrices of continuous-time stationary Markov chains.

A typical element, $p_{ij}(t)$, of $P(t)$ has the interpretation,

$p_{ij}(t)$ = probability that an individual starting in
state i at time 0 will be in state j at
time t.

The Q-arrays, which are known as "intensity matrices," represent

structural information about the population:

(i) $\dfrac{q_{ij}}{-q_{ii}}$ = probability that an individual in state i
will move to state j, given the occurrence
of a transition

(ii) $\dfrac{1}{-q_{ii}}$ = expected length of time for an individual
in state i to remain in that state

We will denote the class of intensity matrices (arrays of the form

(4.2)) by $\underline{Q}$.

Solutions of (4.1) are given by the exponential formula

$$P(t) = e^{Qt}, \quad t > 0 \tag{4.3}$$

where the matrix exponential e^A (A being an arbitrary rxr matrix) is defined by

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

(a) A simple factored representation of Q:

The above general formulation of continuous-time Markov transition matrices has been used in numerous sociological contexts (e.g., Coleman 1964b, pp. 177-182; Bartholomew, 1973). However, the analysis of social processes, particularly in a heterogeneous population, is greatly facilitated by an alternative formulation which provides the basis for a classification of individuals (or sub-populations) according to their rates of movement, their propensities to move to particular states, or both simultaneously. A starting point for this development was S. Spilerman's (1972a) extension of the mover-stayer formulation to continuous-time, with a more general classification of sub-populations than the simple mover-stayer dichotomy. The basis for this extension was simply the introduction of a factored representation for Q-matrices of the special form $Q = \lambda(M-I)$, where λ is a positive constant signifying the expected rate of movement, and M is the transition matrix that each individual in the population follows at a move.

Classification according to <u>rate</u> of movement means assigning a number λ to each individual (or sub-population), thereby designating what we will call type-λ individuals. The value $\frac{1}{\lambda}$ can be interpreted

as an individual's mean waiting time before moving (or before making a decision to possibly move). Similarly, classification according to propensity to transfer to particular states means assigning a stochastic matrix M to an individual, thereby designating what we will call type-M individuals. If persons are to be classified in both of the above ways simultaneously, we would speak of type-(λ, M) individuals.

Using this classification scheme, the random variables $\{Y(t), t > 0\}$ which describe a type-λ individual's history may be constructed from two separate processes: (1) a sequence of independent positive random variables $\tau_0, \tau_1, \ldots$ describing waiting times between moves and satisfying

$$\text{Prob}(\tau_i > t) = e^{-\lambda t}, \quad i = 0,1,2,\ldots \quad t > 0$$

and (2) a discrete-time Markov chain $\{X(k), k = 0,1,2,\ldots\}$ having one-step transition matrix M which describes moves when they occur. You can then think of an individual whose transition probabilities are governed by $e^{t\lambda(M-I)}$ as evolving according to the following prescription:

(i) Starting in state i at time 0, the individual stays there for an exponentially distributed length of time τ_0 with

$$\text{Prob}(\tau_0 > t) = e^{-\lambda t}, \quad t > 0$$

Thus, $Y(t) = X(0) = i$ for $0 \le t < \tau_0$.

(ii) At the end of this time he makes a decision to move to state j with probability m_{ij}. (In general, $m_{ii} \ne 0$.) Thus, $Y(\tau_0) = X(1) = j$.

(iii) Now he waits in state j for an exponentially distributed length of time τ_1 which is independent of τ_0, $X(0)$, and $X(1)$; especially,

$$\text{Prob}(\tau_1 > t \,|\, X(0), \tau_0, X(1)) = \text{Prob}(\tau_1 > t) = e^{-\lambda t}$$

and

$$Y(t) = X(1) \text{ for } \tau_0 \leq t < \tau_0 + \tau_1$$

(iv) Then he makes another decision to move to state h with probability m_{jh}; hence,

$$Y(\tau_0 + \tau_1) = X(2) = h$$

(v) The above sequence is repeated. In general,

$$Y(t) = X(k) \qquad \text{for } \sum_{i=0}^{k-1} \tau_i \leq t < \sum_{i=0}^{k} \tau_i$$

with $\tau_0, \tau_1, \ldots$ independent of $\{X(k), k = 0,1,2,\ldots\}$ and of each other.

S. Spilerman's (1972a) extension of the mover-stayer model was a mixture of Markov processes of the above sort in which individuals associated with the parameter λ were assumed to occur in the total population with a frequency described by the Gamma density

$$g(\lambda) = \frac{\beta^\alpha \, \lambda^{\alpha-1} \, e^{-\beta\lambda}}{\Gamma(\alpha)} \quad \text{where} \quad \alpha > 0, \ \beta > 0, \ \lambda \geq 0 \ .$$

Type-λ individuals are considered to be non-directly observable, and all types of individuals are treated as having the same propensity to move among the states, prescribed by the matrix M. The <u>population-level process</u> $\{Z(t), t > 0\}$, which is observable, then has transition probabilities given by

$$P(t) = \int_0^\infty e^{t\lambda(M-I)} g(\lambda) d\lambda$$

$$= (\frac{\beta}{\beta+t})^\alpha [I - \frac{t}{\beta+t} M]^{-\alpha} \tag{4.4}$$

The choice of a gamma density in this specification is based on the ability of that functional form to describe a variety of unimodal curves, unimodality being a reasonable characterization of the frequency of occurrence of different types of persons, with respect to rate of movement, in heterogeneous populations (Palmer (1954), p. 50; Taeuber, Chiazze, and Haenszel (1968), p. 46).

Two other mixtures of some importance for intra-generational occupational mobility are processes with transition probabilities governed by

$$P(t) = sI + (1-s)e^{t\lambda_0(M-I)} \tag{4.5}$$

and

$$P(t) = sI + (1-s)\int_0^\infty e^{t\lambda(M-I)}g(\lambda)d\lambda$$

$$= sI + (1-s)(\frac{\beta}{\beta+t})^\alpha(I - \frac{t}{\beta+t}M)^{-\alpha} \tag{4.6}$$

Equation (4.5) is a continuous-time analog of the mover-stayer model in which the fraction of stayers is the same for all states, and $\frac{1}{\lambda_0}$ is the expected waiting time between moves in the mover population. Equation (4.6) combines the mover-stayer model with the more general form of heterogeneity in the mover population which was specified in (4.4). Because this mixture adds a concentration of stayers to the gamma density, it is known as the spiked gamma (with vodka please).

(b) A more general factored representation of Q:

From a substantive point of view, a principal defect of the individual-level description in (a) is the requirement that a person's waiting time distribution be the same in every state. It is desirable

to eliminate this constraint and retain the flexibility of the full Markov model, since there are many instances in which rate of movement is a function of system state: for example, if the system states are industry categories we know that industries differ in their rates of employee separation (Blauner 1964, pp. 198-203).

We therefore classify a person according to the diagonal matrix

$$\Lambda = \begin{pmatrix} \lambda_1 & & \underline{0} \\ & \ddots & \\ \underline{0} & & \lambda_r \end{pmatrix}, \qquad \lambda_i \geq 0, \; i = 1,2,\ldots,r$$

where $\frac{1}{\lambda_i}$ has the interpretation, "average waiting time in state i." A type-Λ individual's history $\{Y(t), \; t > 0\}$ is now governed by the transition matrices

$$P(t) = e^{t\Lambda(M-I)}, \qquad t \geq 0 \tag{4.7}$$

and these individuals are viewed as occurring in the total population with a proportion specified by a joint probability density $g(\lambda_1,\ldots,\lambda_r)$. The previous construction of individual histories $\{Y(t), \; t > 0\}$ out of random waiting times $\tau_0, \; \tau_1,\ldots$ and a discrete-time Markov chain $\{X(k), k = 0,1,2,\ldots\}$ must now be modified by allowing the distribution of τ_k to depend on the current state $X(k)$.

In particular, we define

$$Y(t) = X(k) \qquad \text{if} \; \sum_{i=0}^{k-1} \tau_i \; \leq t < \sum_{i=0}^{k} \tau_i$$

where $\text{Prob}(\tau_k > t \mid X(0), \; \tau_0, \; X(1), \; \tau_1,\ldots X(k-1), \; \tau_{k-1}, \; X(k) = i)$

$$= \text{Prob}(\tau_k > t \mid X(k) = i) \; = \; e^{-\lambda_i t} \tag{4.8}$$

for $1 \leq i \leq r; \; k = 0,1,2,\ldots$.

It should be pointed out that this formulation requires more compli-cated estimation techniques than the simple factored representation described in the previous section. However, a full discussion of these issues in the context of panel studies lies outside the scope of the present paper.

(c) More general waiting time distributions than exponential:

Despite the more diverse form of heterogeneity which is formalized in (b), the increasing tendency of persons to remain in a particular state (occupation, geographic region, etc.) the longer they have been there is an empirical regularity which is not captured by any time-stationary Markov model. R. McGinnis (1968) refers to this phenomenon as cumulative inertia, and empirical evidence of its presence in intra-generational mobility is provided, for example, by Land (1969); Myers, McGinnis, and Mesnick (1967) and Kuhn, Poole, Sales, and Wynn (1973). This phenomenon is also known in the demography literature as "duration-dependence," and a nice review of formal models which incorporate it is provided by Hoem (1972).

In order to formalize duration-dependence and simultaneously classify individuals according to rate of movement and propensity to transfer to particular states, it is convenient to retain the decom-position of individual histories $\{Y(t),\ t > 0\}$ discussed in the previous sections. The only modification is the introduction of special non-exponential distributions, $F_i(t)$, $1 \leq i \leq r$, which describe duration-dependent waiting times in state i. In particular, we define

$$Y(t) = X(k) \qquad \text{for } \sum_{i=0}^{k-1} \tau_i \leq t < \sum_{i=0}^{k} \tau_i \qquad (4.9)$$

where τ_0, τ_1,... are positive random variables satisfying

$$\text{Prob}(\tau_k > t \mid X(0), \tau_0, X(1), \tau_1,...,X(k-1), \tau_{k-1}, X(k) = i)$$

$$= \text{Prob}(\tau_k > t \mid X(k) = i)$$

$$= 1 - F_i(t) \qquad\qquad 1 \le i \le r \qquad\qquad (4.10)$$

To incorporate the notion of duration-dependence (or cumulative inertia) we restrict $F_i(t)$ to be of the form

$$F_i(t) = 1 - e^{-\int_0^t h_i(u)du} \qquad\qquad 1 \le i \le r \qquad (4.11)$$

where $h_i(u)$ is a positive decreasing function such that

$$\int_0^\infty h_i(u)du = +\infty \quad .$$

The assumption that h_i be decreasing implies that the longer an individual stays in state i, the less likely he is to move in the immediate future. In particular, the probability that an individual known to be in state i at time t will exit from that state in the next dt units of time is given by

$$h_i(t)dt = \frac{f_i(t)dt}{1 - F_i(t)}$$

where $f_i(t)$ is the probability density corresponding to $F_i(t)$.

The process $\{Y(t), t > 0\}$ defined above is a special form of semi-Markov process[8] whose transition probabilities

$$\text{Prob}(Y(t) = j \mid Y(0) = i) = p_{ij}(t). \qquad 1 \le i \le r$$

are the unique solutions of the system of integral equations

$$p_{ij}(t) = \delta_{ij}[1 - F_i(t)] + \sum_{k=1}^{r} \int_0^t f_i(s)m_{ik}p_{kj}(t-s)ds$$

In this equation,

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ \\ 0 & \text{if } i \neq j \end{cases} \qquad (4.12)$$

and $||m_{ik}|| = M$ is the one-step transition matrix governing the discrete-time Markov chain $X(k)$, $k = 0,1,2,\ldots$ used to specify $Y(t)$ in (4.9).[9] Now classification of an individual evolving according to a semi-Markov process would mean to characterize him by the family of distributions $\mathbb{F} = \{F_1(t),\ldots,F_r(t)\}$ describing the waiting times in any state, and by the stochastic matrix M describing his propensity to move to particular states.

In specifying a population-level process $\{Z(t), t > 0\}$ as a mixture of this kind of semi-Markov process, parametric families of distributions are usually used to define $F_i(t)$ and then a suitable mixing distribution is defined on the parameters. For example,

$$F(t) = 1 - e^{-\gamma_1 t^{\gamma_2}} \qquad \text{with } \gamma_1 > 0, \ 0 < \gamma_2 < 1$$

can be expressed in the form (4.11) with

$$h(u) = \gamma_1 \gamma_2 u^{\gamma_2 - 1}$$

and a reasonable initial choice of mixing distribution can be defined by

$$\int_0^{\gamma_1} \int_0^{\gamma_2} g(u, v)du\, dv = \int_0^{\gamma_1} \frac{\beta^\alpha u^{\alpha-1} e^{-\beta u}}{\Gamma(\alpha)} du \cdot \int_0^{\gamma_2} dv$$

Thus γ_1 and γ_2 are treated as independent parameters with γ_1 being

gamma distributed and γ_2 being uniformly distributed on $[0, 1]$. This mixture specification is meant to be only a suggestion of a reasonable starting point for the fitting of semi-Markov mixtures to multi-wave panel data. A series of empirical investigations comparing a variety of mixture models remains to be carried out.

A final point which should be mentioned concerning the semi-Markov models (4.9) is the basically regenerative nature of these processes. In particular, individuals evolving according to (4.9)-(4.11) have an increasing propensity to remain in each state the longer they are there. However, once a change in state occurs, an individual may be much more likely to move again in the immediate future than he was before the change occurred. Although the cumulative inertia behavior occurs in each state separately, it need not, according to these models, hold throughout a career involving changes of state (i.e., there is no explicit notion of individual aging). This raises the question of finding alternative models to the above semi-Markov formulation in which the propensity to move in the immediate future decreases throughout an individual's history. This is the subject of the next section.

(d) A non-time-stationary Markov model:

Consider a population in which an individual's history $\{Y(t),\ t > 0\}$ is defined by

$$Y(t) = X(k) \qquad \text{for} \quad \sum_{i=0}^{k-1} \tau_i \leq t < \sum_{i=0}^{k} \tau_i \qquad (4.13)$$

where $\{X(k),\ k = 0,1,2,\ldots\}$ is again a discrete-time Markov chain, governed by M and describing moves when they occur. $\tau_0,\ \tau_1,\ldots$ are

waiting times between moves (or decisions to possibly move), and they satisfy

$$\text{Prob}(\tau_k > t \mid X(0), \tau_0, \ldots, X(k-1), \tau_{k-1}, X(k))$$

$$= \text{Prob}(\tau_k > t \mid \tau_0 + \ldots + \tau_{k-1})$$

$$= 1 - \exp\left(-\int_{(\tau_0 + \ldots + \tau_{k-1})}^{t+(\tau_0 + \ldots + \tau_{k-1})} h(u)\,du\right) \qquad (4.14)$$

where $h(u)$ is positive, decreasing, and satisfies

$$\int_0^\infty h(u)\,du = +\infty .$$

The specification (4.14) implies that after each successive move, an individual's propensity to remain in his new state is not only greater the longer he stays, but it is also greater than at any time prior to his last move. In particular, this formulation captures the notion of cumulative inertia throughout a career, such as might result from aging, and seems more appropriate than some of the previous semi-Markov models for investigations of intra-generational occupational mobility. See, in particular, Kuhn, Poole, Sales, and Wynn (1973) for some empirical evidence supporting this position; also see Sørenson (1975) for additional details on this sort of formulation.

The stochastic process specified by (4.13) and (4.14) is a special non-time-stationary Markov process[10] where

$$\text{Prob}(Y(t) = j \mid Y(0) = i) = p_{ij}(0, t)$$

$$= \left(e^{\left(\int_0^t h(u)\,du\right)(M-I)}\right)_{ij} \qquad (4.15)$$

In principle, heterogeneous population models could be constructed
from mixtures of this kind of non-time-stationary model of individual
behavior. However, the fragmentary nature of the data which is usually
available in multi-wave panel studies makes judgments about the rela-
tive importance of non-stationarity vs. heterogeneity essential if
parsimonious models are to be fit to the data. In terms of the dis-
cussion of high diagonals in Section 3, the difficult conceptual
point which such judgments raise is that each of the following quali-
tatively different interpretations is capable of accounting for that
empirical regularity.

(i) A homogeneous population described by the non-stationary
model (4.13) and (4.14).

(ii) A heterogeneous population described by mixtures of sta-
tionary Markov models such as the mover-stayer extensions
(4.4), (4.5), and (4.6).

(iii) A homogeneous population described by a semi-Markov model
such as (4.9)-(4.11).

(iv) A heterogeneous population described by mixtures of (i)
and (iii) above.

A strategy for discriminating among alternative conceptual models,
such as these, in a panel study is outlined in Section 6. The dis-
cussion there is designed to illustrate a general strategy of fitting
several models to the same data, each of which emphasizes a qualita-
tively different behavioral pattern. Highly structured residuals from
such models usually represent the most suggestive information about
factors which have not been formally incorporated in a model. (The
empirical regularity found by BKM is a simple instance of residuals

from a base-line model being suggestive about alternative descriptions of an empirical process.) One of the principal research directions which this approach suggests is the intensive development of fitting and identification procedures for a variety of realistic models using limited longitudinal information.

4.2 Fragmentary Data

From the outset we have emphasized the limited number of time points at which panel data are usually obtained. It is important for a proper understanding of the estimation and identification strategies discussed in Sections 5 and 6 that some explicit instances of longitudinal data be described, together with an indication of precisely what, in each instance, is meant by the phrase "fragmentary."

Example 1:

Let $\{Y^{(i)}(t), 0 \leq t \leq t^*, t^* = $ duration of the study$\}$ represent the history of the ith individual in a panel study (e.g., occupational career pattern, succession of brand preferences, etc.), and let $0 = t_0 < t_1 < ... < t_n$ represent the times at which the waves of the panel are scheduled (i.e., the re-interview times). Although changes of state can occur at any time t, the observed process is

$$\{Y^{(i)}(t_k), \ k = 0,1,2,...,n\}, \qquad 1 \leq i \leq N, \qquad (4.16)$$

where N = number of persons in the closed population under study. Thus, the transitions between sampling instants as well as their times of occurrence are not observed. It is because of this missing information that we refer to data of the form (4.16) as fragmentary.[11] It should

be noted that this was precisely the sampling situation in BKM's study where $t_{i+1} - t_i = \Delta = 3$ months, $i = 0,1,2,\ldots$.

Example 2:

In Taeuber, et al.'s (1968) residence history study, observations are taken retrospectively on current residence, first and second prior residence, and birth place of individuals in particular age cohorts. This kind of data represents an instance of fragmentary information about a migration process in that gaps are present in the residence histories.

Example 3:

Let $T^{(i)}(t) = \{$number of transitions by the ith individual between time 0 and time $t\}$, and consider observations of the form $\left\{ Y^{(i)}(t_k), \right.$ $\left. T^{(i)}(t_k),\ 0 \le k \le n,\ 1 \le i \le N \right\}$. This kind of information was obtained in the social mobility studies of Palmer (1954), Lipset and Bendix (1963), and in the much larger study of Parnes (1972). It is fragmentary due to the fact that the times of occurrence of the transitions are missing.

From the perspective of estimation and identification with any of the mathematical models mentioned previously, the ideal situation would be to have complete histories of moves among states, as well as durations in each state, for a long time interval. However, because of cost considerations in conducting many re-interviews over a long time span, and because of low response reliability when detailed retrospective questions are asked, only fragmentary data have been obtained in such major investigations as Blau and Duncan's OCG1 Survey (1968), Hauser and Featherman's OCG2 Survey (1973), Michigan's Income Dynamics

Panel, and the National Longitudinal Study of Labor Force Experience
(Parnes Study 1972). This raises the question of what sorts of partial
information to gather if the data are to be used to discriminate among
alternative theories using formal mathematical models. For example,
if the study concerns occupational mobility and the collection design
is a retrospective survey, we might collect any of the following
kinds of data:

a) A complete history of all jobs held and durations in the
   jobs.

b) First occupation and current occupation.

c) First occupation, current occupation, and number of
   intervening occupations held.

d) First occupation, current occupation, and previous occupation
   (possibly together with duration times in each occupational
   state).

Clearly, the combinations can be elaborated. What is consequen-
tial about this decision is that once alternative (a)--complete
histories--is rejected as a research design, it becomes crucial as
to which pieces of data one decides to collect. Different estimation
procedures must be employed according to the kinds of information
gathered, and some procedures will yield more efficient estimates of
the parameters than will others. The choice of data collection
strategy must also reflect the classes of mathematical models that a
researcher intends to apply, since certain information which is not
required to fit one model type is crucial to the estimation of another.

The simplest setting in which to illustrate estimation and
identification with fragmentary data is the fitting of continuous-time

Markov chains to data of the form described in Example 1; that is,

observations on individuals' locations at a few points in time. The

essential steps are described in Section 5.

5. EMBEDDABILITY AND IDENTIFICATION

Suppose observations on a closed population have been collected

at the evenly spaced time points $t_0 = 0$, t_1, t_2,...,t_n where $t_{k+1} - t_k$

$= \Delta > 0$, for $k = 0,1,...,n$, and assume that the number of observations

on the population is small, say, $n \leq 8$. Furthermore, consider the

observations to include only information on current system states;

namely, $\left\{ Y^{(i)}(t_k), k = 0,1,2,...,n \right\}$ for $1 \leq i \leq N$, N = number of

persons in the closed population under investigation. This is a standard

data collection situation in multi-wave panel studies (for example, BKM

used this type of data), and it provides the simplest setting in which

to illustrate embeddability and identification issues.

Embeddability refers to the question of whether or not observa-

tions on an empirical process are compatible with the conceptual

assumptions (theoretical structure) underlying a particular class of

mathematical models (e.g., time-homogeneous Markov, mixture of Markov,

semi-Markov). Where the answer is affirmative, identification proce-

dures refer to techniques for recovering the particular set of struc-

tural parameters from the model class which should be associated with

the empirical process. Both issues are central to the analysis of

panel data. Identification, in particular, can be difficult to accom-

plish due to the fact that qualitatively different sets of structural

parameters may be consistent with data from evenly spaced observations.

To fix the ideas in the simplest setting, consider fitting a continuous-time Markov chain with stationary transition probabilities to data of the sort described above. A procedure for carrying out this task consists of two principal steps:

(i)   Form the stochastic matrices $\hat{P}(k\Delta, \ell\Delta)$ with entries

$$\frac{n_{ij}^{(k\Delta,\ell\Delta)}}{n_{i+}^{(k\Delta,\cdot)}} = \frac{\left(\begin{array}{l}\text{number of persons in state } i \text{ at time } k\Delta \\ \text{who are also in state } j \text{ at time } \ell\Delta\end{array}\right)}{\left(\text{number of persons in state } i \text{ at time } k\Delta\right)}$$

and check that

$$\hat{P}(k_1\Delta, k_2\Delta) = \hat{P}(k_3\Delta, k_4\Delta) \tag{5.1}$$

for $k_1 < k_2 \leq n; \; k_3 < k_4 \leq n$

and that

$$\hat{P}(k_1\Delta, k_2\Delta) = \hat{P}(k_1\Delta, \ell\Delta)\hat{P}(\ell\Delta, k_2\Delta) \tag{5.2}$$

where $0 \leq k_1 < \ell < k_2 \leq n$.

Equation (5.1) is a test of time stationarity; and (5.2) is a primitive test of the Markov assumption (i.e., independence of future state from past history, given current state). Formal tests of this kind are described by Anderson and Goodman (1957) and Billingsley (1961).

(ii)   Compute

$$\frac{1}{(\ell-k)\Delta} \log \hat{P}(k\Delta, \ell\Delta), \qquad 0 \leq k < \ell \leq n \tag{5.3}$$

and observe that if the data are compatible with a time stationary Markov model, then at least one branch of the logarithm of any given matrix in the list (5.3) should be roughly equal to some branch of the logarithm of any other matrix in the list. In addition, this common

logarithm should be an intensity matrix (i.e., it should belong to the

class $\underline{Q} = \left\{ Q: q_{ii} \leq 0, q_{ij} \geq 0 \text{ for } i \neq j, \sum_{j=1}^{r} q_{ij} = 0 \right\}$).

The process of verifying that $\hat{P}$ can be represented in the form

e^Q for at least one $Q \in \underline{Q}$ is a test for <u>embeddability</u> of the data in a

continuous-time Markov model. Although this step is seemingly straight-

forward, it should be pointed out that some surprisingly subtle phenomena

are involved in the embeddability test. In particular, due to the

multiple valued nature of the logarithm function, it is not immediately

apparent that one can find an effective computation algorithm to check

for the existence of even one branch of $\log \hat{P}$ which is an intensity

matrix. Indeed, it would appear that infinitely many branches of the

logarithm might have to be checked to decide on embeddability.

Fortunately, however, any matrix which can be represented as e^Q

with $Q \in \underline{Q}$ must have eigenvalues of a rather restrictive nature. In

fact it is the existence of sharp upper and lower bounds on the eigen-

values which lead to a practical computation strategy for deciding

embeddability. The explicit eigenvalue restrictions and associated

computation scheme are outlined below. However, for a detailed dis-

cussion of this point and further indication of its role in the analysis

of panel data, the reader should consult Singer and Spilerman (1976).

Closely related to the problem of deciding embeddability with a

finite number of tests is the fact that in the course of such a compu-

tation, there may be several branches of the logarithm of a stochastic

matrix which are intensity matrices. <u>Identification</u> refers to the task

of deciding which of these intensity arrays should be associated with

the empirical process. In particular, it is possible to have

$$\hat{P}(k\Delta, \ell\Delta) = e^{(\ell-k)\Delta Q_1}$$

and

$$\hat{P}(k\Delta, \ell\Delta) = e^{(\ell-k)\Delta Q_2} \qquad 0 \leq k < \ell \leq n \qquad (5.3)$$

where $Q_1 \neq Q_2$ but $Q_1 \in \underline{Q}$ and $Q_2 \in \underline{Q}$. (See Singer and Spilerman, 1975, 1976 for explicit examples of this behavior.) The phenomenon (5.3) is an instance of aliasing for Markov transition matrices, and it is entirely analogous to the aliasing of structural coefficient matrices in continuous-time econometric models (see, in particular, P.C.B. Phillips, 1973). In this situation, the set of underlying structural parameters (i.e., the unique intensity matrix which should be associated with an empirical process) is not identifiable. A researcher confronted with matrices such as Q_1 and Q_2 would either have to adjudicate between them on substantive grounds or collect additional data at a time which is not a multiple of the sampling interval Δ. Then the underlying transition mechanism could be identified since only one of the matrices (Q_1 or Q_2) could be consistent with the non-evenly spaced observations.

The computation scheme outlined below recovers all branches of the logarithm of a stochastic matrix $\hat{P}$ which are intensity matrices, provided $\hat{P}$ has distinct eigenvalues. This is clearly the situation in most applications. However, it should be noted that repeated eigenvalue matrices do play an important role in sensitivity analyses, and they can be associated with a continuum of intensity matrices (i.e., a stochastic matrix $\overline{P}$ may be within error distance of the observed array $\hat{P}$ and be representable as $\overline{P} = e^{Q}$ for an uncountably infinite set of matrices in $\underline{Q}$: see Singer and Spilerman (1976) and Cuthbert (1973) for details on this point). This raises difficult questions of both

interpretation and reliability of estimates of structural parameters based on evenly spaced data. Extensive re-analyses of data from a variety of panel studies would be necessary in order to assess whether the possible instability described above is in fact a frequently occurring empirical phenomenon.

A Computation Strategy to Decide Embeddability.

Step 1:

Compute the eigenvalues of $\hat{P}$ and check whether or not they each satisfy

$$\pi(\frac{1}{2} + \frac{1}{r}) \leq \arg(\log \lambda) \leq \pi(\frac{3}{2} - \frac{1}{r}) \tag{5.4}$$

where r = order of the matrix, and λ is an eigenvalue of $\hat{P}$. (The inequalities (5.4) were established by J. Runnenberg, 1962.) In particular, he used the inequalities of F. I. Karpelevitch (1951)

$$\pi(\frac{1}{2} + \frac{1}{r}) \leq \arg(\lambda-1) \leq \pi(\frac{3}{2} - \frac{1}{r}) \tag{5.5}$$

--which restrict the eigenvalues, λ, of an arbitrary rxr stochastic matrix--together with the representation $P(t) = e^{tQ}$ for Markov transition matrices, to obtain (5.4) as a restriction on the eigenvalues of e^{Q}. The shaded zone in Figure 1 depicts the region defined by the inequalities (5.4) and exhibits a typical set of logarithms of the eigenvalues associated with an embeddable matrix.

Step 2:

If all eigenvalues of $\hat{P}$ are real and positive, then their logarithms are real and negative and they automatically satisfy (5.4). In this
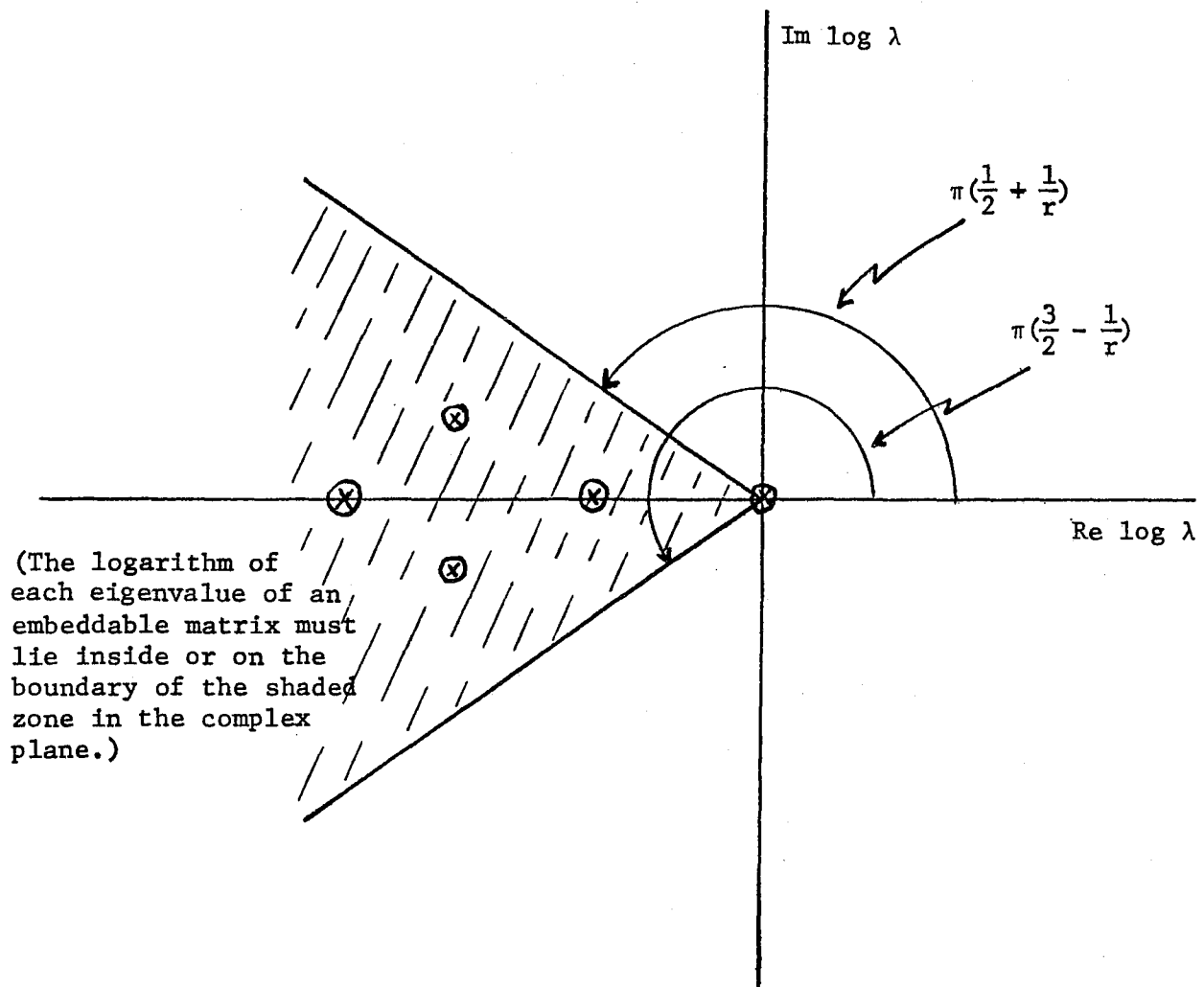
Figure 1. Eigenvalue Restrictions for Embeddable Matrices

situation there can be at most one branch of $\log \hat{P}$ in $\underline{Q}$. To compute
it--and thereby check embeddability--reduce $\hat{P}$ to diagonal form (i.e.,
represent $\hat{P}$ as $\hat{P} = H \Lambda H^{-1}$ where

$$\Lambda = \begin{pmatrix} \lambda_1 & & \underline{0} \\ & \ddots & \\ \underline{0} & & \lambda_r \end{pmatrix}$$

and H is a nonsingular similarity transformation). Then calculate

$$\log \hat{P} = H \log \Lambda H^{-1} \tag{5.6}$$

where

$$\log \Lambda = \begin{pmatrix} \log \lambda_1 & & \underline{0} \\ & \log \lambda_2 & \\ & & \ddots \\ \underline{0} & & \log \lambda_r \end{pmatrix}$$

If the matrix (5.6) is in $\underline{Q}$, then $\hat{P}$ is embeddable in the unique
continuous-time Markov model with intensity matrix given by (5.6).
If (5.6) is not in $\underline{Q}$, then $\hat{P}$ is simply not embeddable in any continuous-
time Markov model.

Step 3:

If $\hat{P}$ has complex eigenvalues they must occur in conjugate pairs.
For each such pair $(\lambda = \rho e^{i\theta}, \bar{\lambda} = \rho e^{-i\theta})$ determine all branches of
their logarithms which satisfy (5.4); especially,

$$\pi(\frac{1}{2} + \frac{1}{r}) \le \arg(\log_k \lambda) \le \pi(\frac{3}{2} - \frac{1}{r})$$

where r = order of the matrix $\hat{P}$,

$$\arg(\log_k \lambda) = \tan^{-1}\left(\frac{\theta + 2\pi k}{\log \rho}\right), \qquad \text{and k specifies}$$

a branch of $\log_k \lambda$ according to

$$\log_k \lambda = \log \rho + i(\theta + 2\pi k); \quad k = 0, \pm 1, \pm 2, \ldots \quad (5.7)$$

$$0 < \theta < \pi$$

Now select one of the branches for each complex conjugate pair, and compute $\log \hat{P}$ using (5.6). Check the resulting matrix for membership in $\underline{Q}$. Then repeat this calculation for all branches satisfying (5.4). The basic importance of Runnenberg's inequalities (5.4) is revealed at this step, because they guarantee that only _finitely many_ branches need be checked. Furthermore, all intensity matrices compatible with the data (the aliases mentioned in (5.3)) are recovered in these calculations. If multiple matrices $Q \in \underline{Q}$ have been found, the researcher should collect additional information to discriminate among them in the manner described in conjunction with (5.3).

6. STRATEGIES FOR DISCRIMINATING AMONG COMPETING MODELS

Many of the issues involved in attempting to discriminate among competing models can be illustrated in the relatively simple setting of testing data for compatibility with one of the following four classes of models:

(i)   time-stationary Markov chains

(ii) a restricted class of mixtures of (i)

(iii) a restricted class of non-time-stationary Markov chains

(iv) a restricted class of semi-Markov processes.

To fix the ideas, assume that observations

$$\left\{ Y^{(i)}(t_k), \ T^{(i)}(t_k) \right\}, \quad 1 \leq i \leq N, \ 0 \leq k \leq n \quad (6.1)$$

have been collected at the evenly spaced time points $0 = t_o < t_1 < \ldots < t_n$ where $t_{k+1} - t_k = \Delta =$ (spacing between successive observations) > 0; $k = 0, \ldots, n - 1$; $Y^{(i)}(t_k)$ denotes the state of the ith individual in the survey at time t_k; and $T^{(i)}(t_k)$ equals the total number of transitions by the ith individual in the time interval $(0, t_k)$. This is precisely the data collection situation described in Example 3 of Section 4.2.

Now introduce models in which individual histories are represented in the form

$$Y(t) = X(T(t)), \qquad t \geq 0 \qquad\qquad (6.2)$$

where $X(k)$ is a discrete-time Markov chain with stationary transition probabiliiies having one-step transition matrix M, and $T(t)$ is one of the four kinds of stochastic processes listed below.

(A) a time-stationary Poisson process with parameter $\gamma > 0$ [special case of (i)]

(B) a mixture of (A) [special case of (ii)]

(C) a non-time-stationary Poisson process with expected number of jumps in the time interval (s, t) given by $\int_s^t h(u)du$ where h is a continuous, positive, decreasing function such that $\int_0^\infty h(u)du = +\infty$ [special case of (iii)]

(D) a renewal process defined by

$$T(t) = \max\left\{ n: \sum_{i=0}^{n-1} \tau_i \leq t \right\} \qquad \text{if } \tau_o \leq t$$

$$T(t) = 0 \qquad \text{if } \tau_o > t$$

where $\left\{\tau_i\right\}_{i=0,1,\ldots}$ are independent identically distributed positive random variables such that

$$\text{Prob}\{\tau_i > t\} = 1 - e^{-\int_0^t h(u)du}$$

and h satisfies the same hypotheses as in (C) [special case of (iv)].

In each of these models mobility between states is governed by a single stochastic matrix M. The models differ only in the assumptions which are made about the waiting times between moves. It should also be observed that the representation (6.2) provides an alternative description of some of the models presented in Sections 4.1 a, c, and d.

In particular, when $T(t)$ is a time-stationary Poisson process with parameter $\gamma > 0$ (model A), then $X(T(t))$ is simply the time-stationary Markov chain described in Section 4.1a where the intensity matrix Q has the factored representation $Q = \gamma(M-I)$. The advantage of the representation (6.2) in the present context (i.e., with fragmentary data of the special form (6.1)) is that it explicitly describes the relationship between the observable quantities $(Y(t_i), T(t_i))$, $i = 0,1,2,\ldots,n$ and the non-directly observable process $\{X(k), k = 0,1,2,\ldots\}$. The latter process describes transitions when they occur, and is governed by the matrix of structural parameters M.

When $T(t)$ is a mixture of time-stationary Poisson processes (model B), then $X(T(t))$ can be any one of the mixtures (4.4)-(4.6) depending on the choice of mixing distribution. Population heterogeneity is introduced only through a classification of persons according to their rate of movement, and $T(t)$ describes the number of moves by a type-γ individual. Such individuals are assumed to occur in the total population with a probability specified by the mixing distribution.

When $T(t)$ is a non-stationary Poisson process (model C), $X(T(t))$ is the non-stationary Markov chain constructed in Section (4.1d). The following intuitive description is intended to clarify the manner in which this kind of process evolves.

Consider a homogeneous population in which an individual starting in state i at time zero stays there for a random length of time τ_1 with

$$\text{Prob}(\tau_1 > t) = e^{-\int_0^t h(u)\,du}$$

The assumption that h be decreasing implies that the longer an individual stays in state i, the less likely he is to move in the immediate future. At the end of the initial waiting time, the individual moves to state j with probability m_{ij}. Then he stays in his new state for a random length of time τ_2 whose distribution depends on τ_1 according to

$$\text{Prob}(\tau_2 > t \mid \tau_1 = s) = e^{-\int_s^{s+t} h(u)\,du}$$

Since h is decreasing, the propensity of the individual to remain in this new state is not only greater the longer he stays, but it is also greater than at any time prior to his first move. At time $\tau_1 + \tau_2$, the individual moves again according to M, and waits there a length of time τ_3 governed by

$$\text{Prob}(\tau_3 > t \mid \tau_1 + \tau_2 = s) = e^{-\int_s^{s+t} h(u)\,du}$$

This process is repeated, and with each change of state the individual has less and less propensity to move than at any previous time.

Finally, we consider processes of the form X(T(t)) where T(t) is a renewal process (model D). With this specification, X(T(t)) is a special semi-Markov process as defined in Section 4.1c. In order to clarify the manner by which these processes evolve, consider a homogeneous population in which an individual's _initial_ move is regulated

exactly as in the non-stationary Markov model described above. However, his waiting time τ_2 is assumed to be independent of τ_1 and governed by

$$\text{Prob}(\tau_2 > t) \ = \ e^{-\int_0^t h(u)du}$$

After each successive move, the individual's new waiting time is governed by the same probability law as τ_1 and τ_2. The assumption that h is decreasing still implies that the longer the individual remains in a particular state the less likely he is to move in the immediate future. However, in contrast to the non-stationary Markov model, each time a move is made the propensity to move again starts over at a high value and then decreases. In particular, the <u>continual</u> decrease in propensity of the non-stationary Markov model no longer holds for the present semi-Markov processes. Thus, while the former process may be identified with "aging of an individual," the latter is akin to "cumulative inertia in an occupation," as described by McGinnis (1968).

In attempting to identify which of the above four kinds of models--if any--is compatible with data of the form (6.1), the following strategy may be utilized.

Plot cumulative number of moves vs. t and check whether this is approximately linear (Figure 2) or concave downward (Figure 3). It is the case that models (A) and (B) are consistent with the linear picture where the principal trend is described by a regression line through the origin. Model (D) is consistent with the linear picture but with the main trend--(away from t = 0)--described by a straight line having a possibly non-zero intercept. Only model (C) is consistent with a pattern of the form described by Figure 3.

Figure 2. Cumulative Number of Moves Versus t for Data consistent
with Models (A), (B), and (D)*

---

*The cross above t_i reports the median number of moves by all
persons in the panel study during the time interval $(0, t_i)$.
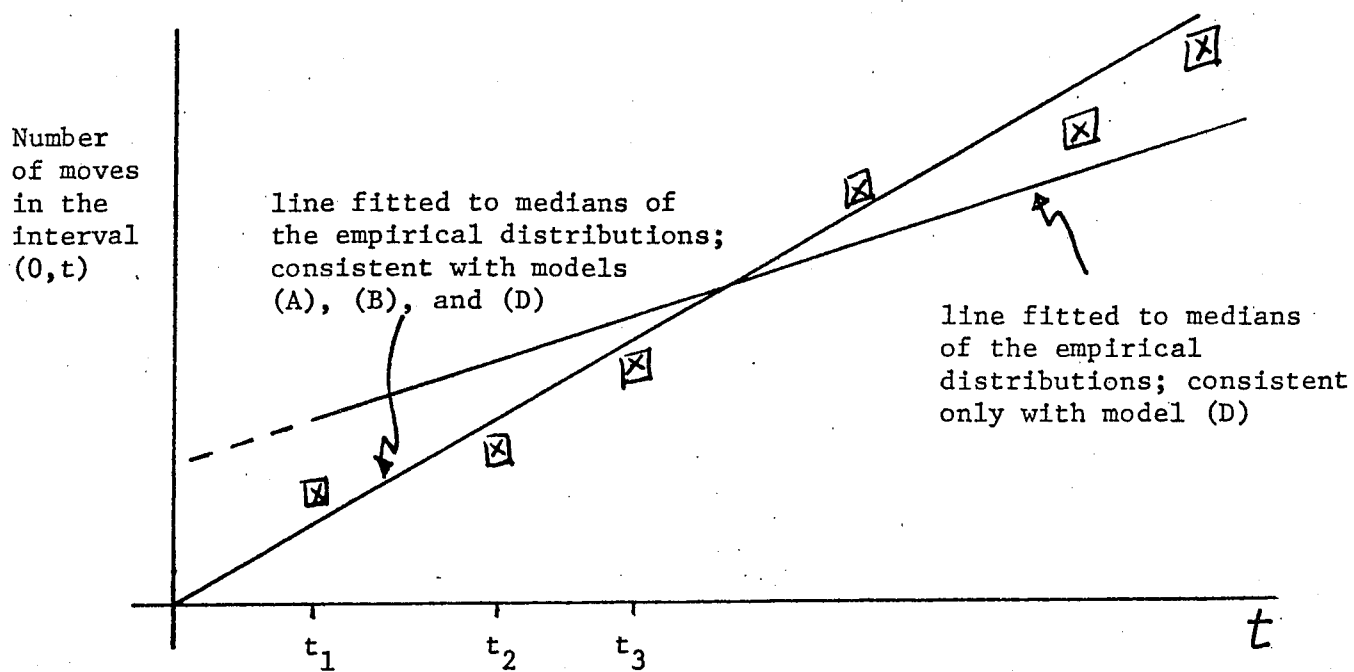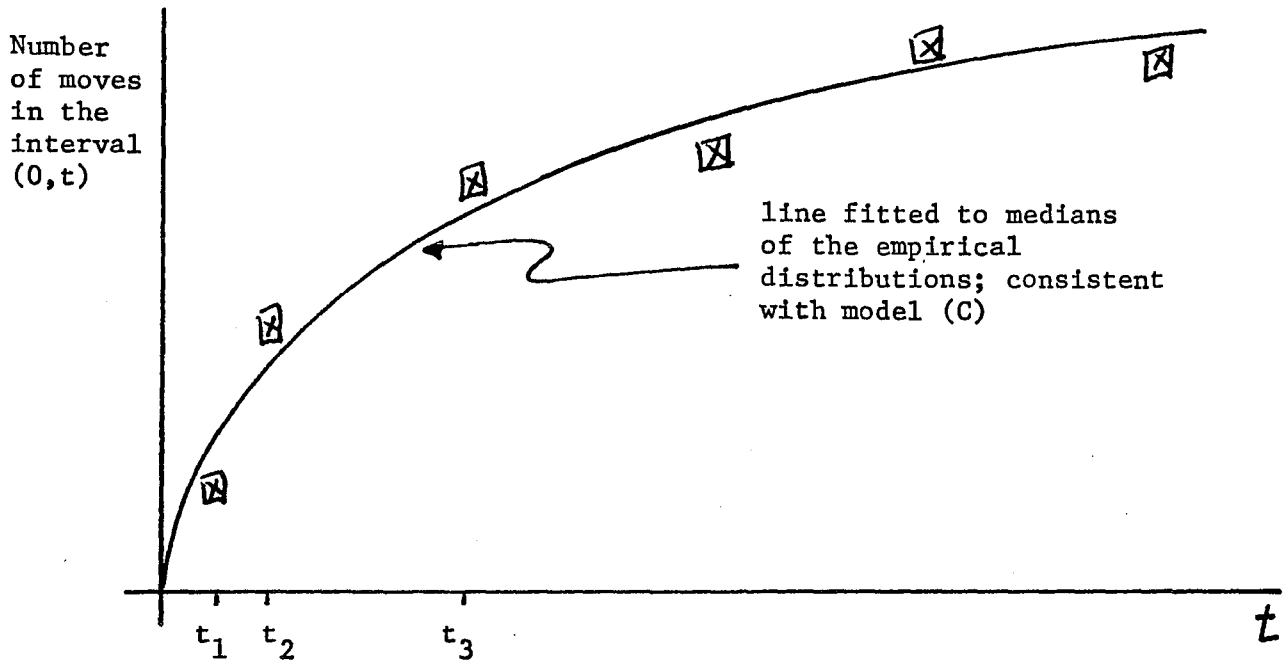
Figure 3. Cumulative Number of Moves Versus t for Data Consistent with Model (C)*

---

*The cross above t_i reports the median number of moves by all persons in the panel study during the time interval $(0, t_i)$.

If the empirical picture corresponds to Figure 2, then the slope, b, of a straight line fitted to the linear pattern would have the following __alternative__ interpretations on the basis of the above data:

(1) $b = \gamma =$ time-homogeneous Poisson parameter

(2) $\int_0^\infty \gamma d\mu(\gamma) = b$ for the mixture of Poisson models[12]

(3) $b = \dfrac{1}{\int_0^\infty (e^{-\int_0^t h(u)du})dt} = \dfrac{1}{\text{(expected waiting time between moves in a renewal process)}}$

If this linear picture is observed, we would solve--using numerical inversion formulas--the following equations for M:

(1) $\hat{P}(0, t_1) = e^{t_1 \gamma (M-I)}$

Call the solution M_1; it corresponds to model (A).

(2) $\hat{P}(0, t_1) = \int_0^\infty e^{t_1 \gamma (M-I)} d\mu(\gamma)$

Call the solution M_1'; it corresponds to model (B).

(3) $\hat{P}(0, t_1) = \sum_{n=0}^\infty (F_n(t_1) - F_{n+1}(t_1))M^n$

Call the solution M_1^*; it corresponds to model (D). (Note: $F_n(t)$ denotes the n-fold convolution of the waiting time distribution

$F(t) = 1 - e^{-\int_0^t h(u)du}$ with itself.)

Now check whether the M-matrix obtained in each case is a bonafide stochastic matrix. This is really an embeddability test for all three model types. If any one of the above calculations yields a matrix which is __not__ stochastic then that model is inconsistent with

the data $\hat{P}(0, t_1)$. If one or more of these calculations yields a stochastic matrix, then we test its ability to predict the observed matrices $\hat{P}(0, t_2)$, $\hat{P}(0, t_3)$ ... using the appropriate equation. In particular, prepare tables of the form

$$(1') \quad \hat{P}(0, t_2) - e^{t_2\gamma(M_1-I)}, \qquad\qquad \hat{P}(0, t_3) - e^{t_3\gamma(M_1-I)}, \ldots$$

$$(2') \quad \hat{P}(0, t_2) - \int_0^\infty e^{t_2\gamma(M'_1-I)} d\mu(\gamma), \qquad \hat{P}(0, t_3) - \int_0^\infty e^{t_3\gamma(M'_1-I)} d\mu(\gamma), \ldots$$

$$(3') \quad \hat{P}(0, t_2) - \sum_n [F_n(t_2)-F_{n+1}(t_2)](M^*_1)^n, \qquad \hat{P}(0, t_3) - \sum_n [F_n(t_3)-F_{n+1}(t_3)](M^*_1)^n, \ldots$$

which represent residuals of observed matrices from predictions based on models (A), (B), and (D), respectively. One instance of the informative nature of such comparisons is the fact that many data sets reveal a discrepancy in comparison (1') in that the diagonal entries in the observed matrices $\hat{P}(0, t_2)$, $\hat{P}(0, t_3)$,..., etc., are substantially larger than the time-homogeneous Markov predictions. Both of the model classes (B) and (D) can account for this kind of discrepancy, despite the fact that they have very different substantive interpretations. Further discrimination requires a more detailed consideration--on both substantive and numerical grounds--of the residual matrices.

Finally, if the concave picture, Figure 3, occurs, then prepare the comparisons

$$\hat{P}(0, t_2) - e^{(\int_0^{t_2} h(u)du)(M_+-I)} \qquad\qquad \hat{P}(0, t_3) - e^{(\int_0^{t_3} h(u)du)(M_+-I)}, \ldots$$

where M_+ is a solution of the equation

$$\hat{P}(0, t_1) = e^{(\int_0^{t_1} h(u)du)(M-I)}$$

Sharp discrepancies here might be revealing about alternative classes of models which should be added to the original list as candidates to describe the empirical process.

We emphasize that the preceding discussion is by no means complete, and is designed only to communicate to the reader, in concise form, the flavor of the sorts of considerations which seem appropriate for the analysis of multi-wave panel data.

7. CONCLUSIONS

We have described a number of issues which arise in fitting models of distributional change to fragmentary data, and in attempting to discriminate among alternative structures fitted to the same fragmentary data. The univariate Markov framework has been applied to advantage in the physical sciences in situations where the notion of population heterogeneity is not especially pertinent, and where the number of observations in time available to a researcher is reasonably large. However, both of these factors are crucial considerations in modeling social phenomena, and we have therefore focused on some variants of the univariate Markov framework that were developed for the expressed purpose of incorporating assumptions about the nature of social processes into mathematical models.

There are additional important issues, closely related to the ones we have discussed, which must be addressed if a routine methodology is to be developed regarding the application of these model types to social

processes. These issues have not been considered in the present review because they are largely undeveloped research areas.

(i) The introduction of substantive theories into Markov models and their variants via restrictions on the structural parameters, such as by prohibiting certain transitions (e.g., occupation or industry shifts) from occurring directly. Some exploratory work on this issue may be found in Coleman (1964a, 1964b).

(ii) Strategies for sensitivity analyses to assess the nature of the dependence of parameter estimates on small perturbations in the data. This topic is discussed within the context of time-stationary Markov models in Singer and Spilerman (1976), but must be extended to other model types.

(iii) Specification of formal error structures, and the development of techniques for setting confidence limits on parameter estimates which derive from fragmentary data.

(iv) The specification of data collection designs for panel studies which will facilitate discrimination among several models fitted to the same fragmentary data. This should include a detailed consideration of the reliability of retrospective interrogation versus the cost and time delays attendant upon reinterview in a panel study. Furthermore, an investigation of the optimal frequency of reinterview in a panel study is required in order to optimize the amount of useful information about change collected per research dollar.

Footnotes

[1] Empirically determined stochastic matrices will be designated by $\hat{P}(u, v)$ with entries

$$\frac{n_{i,j}^{(u,v)}}{n_{i,+}^{(u,\cdot)}} = \frac{\left(\begin{array}{l}\text{number of individuals starting in state } i \\ \text{at time } u \text{ who are in state } j \text{ at time } v\end{array}\right)}{\left(\begin{array}{l}\text{number of individuals starting in state } i \\ \text{at time } u\end{array}\right)}$$

where $u < v$. BKM's estimate of M can thus be written as $\frac{1}{12} \sum_{k=0}^{11} \hat{P}(k, k+1)$ where the unit of time is three months (= 1 quarter).

[2] By "structural information" we mean quantities which characterize a population, irrespective of the observation interval used for data collection.

[3] As the reader will see, these parameters are independent of the sampling interval.

[4] By "non-synchronous" we mean that persons do not all change state simultaneously.

[5] In instances where a substantively meaningful unit time interval exists, a discrete-time model would indeed be appropriate (e.g., explaining presidential election outcomes).

[6] $\hat{p}_{ii}(0, k)$ is a diagonal entry in the observed k-step matrix, and $m_{ii}^{(k)}$ is the corresponding entry in the k-step matrix predicted by a discrete-time Markov chain.

[7]We have replaced the usual mathematical terminology "time-
homogeneous Markov chain" by the phrase "time-stationary Markov
chain." This change of terminology has been incorporated in
order to avoid confusion with our use of the word "homogeneous"
to describe a population of individuals possessing a common set
of transition probabilities. It should also be emphasized
that we do not assume that the initial distribution of indi-
viduals among system states is the equilibrium distribution for
a Markov process. Such an assumption would imply that the Markov
process is also a "stationary" process in the usual mathematical
sense of the word.

[8]For a rigorous mathematical discussion of the special semi-
Markov construction defined by equations (4.9) and (4.10), see
Kurtz (1971).

[9]The specification of semi-Markov processes in equations (4.9),
(4.10), and (4.12) does not describe the most general process of this
kind as treated in the mathematics literature. In particular, the
original semi-Markov framework allowed for waiting time distributions
that could depend on the next future state as well as on the current
state of the process. In order to utilize models incorporating this
kind of detail, a more extensive data base would be required than is
currently available in most multi-wave panel studies. Hence, consid-
erations of parsimony have led us to restrict our attention to a sub-
class of semi-Markov processes which requires the estimation of fewer
parameters.

[10]For a nice mathematical treatment of non-time-stationary Markov chains, see Goodman and Johansen (1973).

[11]Another reason why we might consider data to be fragmentary is if the duration of the study is too brief for significant amounts of movement to have occurred. We do not address this issue in the present discussion.

[12]When the density function $\mu'(\gamma)$ exists, this expression reduces to the familiar formula for a weighted average, $\int_0^\infty \gamma\mu'(\gamma)d\gamma$. By the text expression, however, we mean integration with respect to a general probability measure.

References

Anderson, T. W. and Goodman, L. A.

   1957   "Statistical inference about Markov chains," Annals of

            Mathematical Statistics, 28, pp. 89-109.

Bartholomew, D. J.

   1973   Stochastic Models for Social Processes, 2nd Edition.

            New York: John Wiley and Sons.

Billingsley, P.

   1961   Statistical Inference for Markov Processes. Chicago:

            University of Chicago Press.

Blau, P. and Duncan, O. D.

   1967   The American Occupational Structure. New York: John

            Wiley and Sons.

Blauner, Robert.

   1964   Alienation and Freedom. Chicago: University of Chicago

            Press.

Blumen, I., Kogan, M., and McCarthy, P. J.

   1955   The Industrial Mobility of Labor as a Probability Process.

            Cornell Studies of Industrial and Labor Relations, vol. 6.

            Ithaca, New York: Cornell University.

Chung, Kai Lai.

   1967   Markov Chains with Stationary Transition Probabilities.

            Berlin: Springer.

Coleman, James S.

   1964a   Models of Change and Response Uncertainty. Englewood Cliffs,

            New Jersey: Prentice-Hall.

Coleman, James S.

1964b   Introduction to Mathematical Sociology. New York: Free

         Press.

Cuthbert, J. R.

1973   "The logarithm function for finite-state Markov semi-

         groups," Journal of the London Mathematical Society, 6:

         524-532.

Feller, William.

1956   Book Review of Blumen, Kogan, and McCarthy (1955),

         Psychometrika, 21: 217.

Goodman, G. S. and Johansen, S.

1973   "Kolmogorov's differential equations for non-stationary,

         countable state Markov processes with uniformly continuous

         transition probabilities," Proceedings of the Cambridge

         Philosophical Society, 73:119-138.

Goodman, Leo A.

1961   "Statistical Methods for the Mover-Stayer Model," Journal

         of the American Statistical Association, vol. 56, no. 296:

         841-868.

Hauser, R. and Featherman, D.

1973   "Trends in the occupational mobility of U. S. men, 1962-

         1970," American Sociological Review, 38, (June):302-310.

Hauser, R. and Sewell, W.

1975   Education, Occupation, and Earnings. New York: Academic Press.

Hodge, R. W.

1966   "Occupational mobility as a probability process," Demography,

         3:19-34.

Hoem, Jan.

1972   "Inhomogeneous semi-Markov processes, select actuarial
        tables, and duration-dependence in demography" in T. N. E.
        Grevelle, (ed.), Population Dynamics, Academic Press,
        pp. 251-296.

Karpelevitsch, F. I.

1951   "On the characteristic roots of a matrix with non-negative
        elements," Isvestija, Serie Mathematique, 15:361-383.

Kuhn, A., Poole, A., Sales, P., and Wynn, H. P.

1973   "An Analysis of Graduate Job Mobility," British Journal of
        Industrial Relations, 11:124-142.

Kurtz, Thomas G.

1971   "Comparison of Semi-Markov and Markov Processes," Annals
        of Mathematical Statistics, vol. 42, no. 3:991-1002.

Land, K.

1969   "Duration of Residence and Prospective Migration," Demography,
        6:133-140.

Lazarsfeld, Paul F. and Neil W. Henry.

1968   Latent Structure Analysis. Boston: Houghton Mifflin.

Lipset, S. M. and Bendix, R.

1959   Social Mobility in Industrial Society. Berkeley: Univer-
        sity of California Press.

McCall, John J.

1973   Income Mobility, Racial Discrimination, and Economic
        Growth. Lexington, Mass.: Heath.

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," American Socio-
        logical Review, 33, (October):712, 722.

Morgan, James N. and Smith, James D.

1969   A Panel Study of Income Dynamics, Institute for Social
        Research, Ann Arbor, Michigan.

Myers, G. C., McGinnis, R., and Masnick, R.

1967   "The duration of residence approach to a dynamic stochastic
        model of internal migration: a test of the axiom of cumu-
        lative inertia," Eugenics Quarterly, 14:121-126.

Palmer, G.

1954   Labor Mobility in Six Cities. New York: Social Science
        Research Council.

Parnes, Herbert S.

1972   "Longitudinal Surveys: prospects and problems," Monthly
        Labor Review (February):11-15.

Phillips, P. C. B.

1973   "The problem of identification in finite parameter con-
        tinuous time models," Journal of Econometrics, 1:351-362.

Rees, A. and Watts, H.

1976   The Urban Negative Income Tax Experiment. New York:
        Academic Press.

Rogers, A.

1966   "A Markovian analysis of migration differentials," Proceed-
        ings of the American Statistical Association, Social Science
        Section, Washington, D.C.: American Statistical Association.

Runnenberg, J. Th.

    1962   "On Elfving's problem of imbedding a time-discrete Markov

            chain in a continuous time one for finitely many states,"

            Kroninklijke Nederlandse Akademie van Wetenschappen Pro-

            ceedings, Series A, Mathematical Sciences, vol. LXV, no. 5:

            536-541.

Singer, B. and Spilerman S.

    1974   "Social Mobility Models for Heterogeneous Populations."

            In Herbert Costner (ed.), Sociological Methodology 1973-74.

            San Francisco: Jossey-Bass, pp. 356-401.

    1975   "Identifying Structural Parameters of Social Processes Using

            Fragmentary Data," Proceedings, 40th Session of International

            Statistical Institute, Warsaw Poland.

    1976   "The Representation of Social Processes by Markov Models,"

            American Journal of Sociology, 81, (July).

Smith, J. O. and Cain, G.

    1967   "Markov chain applications to household income distribu-

            ttions," mimeographed, University of Wisconsin.

Social Security Administration

    1972   "Basic statistical data files available to outside researchers,"

            U. S. Department of Health, Education, and Welfare, 3 pages

            (mimeo).

Sørenson, A.

    1975   "Growth in Occupational Achievement: Social Mobility or

            Investments in Human Capital" in K. Land and S. Spilerman,

            editors, Social Indicator Models, New York: Russell Sage

            Foundation.

Spilerman, S.

 1972a "Extensions of the mover-stayer model," <u>American Journal</u>

     <u>of Sociology</u>, 78:599-627.

 1972b "The analysis of mobility processes by the introduction of

     independent variables into a Markov chain," <u>American Socio-</u>

     <u>logical Review</u>, 37, (June):277-294.

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

 1968 <u>Migration in the United States</u>. Washington, D. C.:

     Government Printing Office.

Tarver, J. O. and Gurley, W. R.

 1965 "A stochastic analysis of geographic mobility and population

     projections of the census divisions in the United States,"

     <u>Demography</u>, 2:134-139.

White, Harrison C.

 1970 "Stayers and Movers," <u>American Journal of Sociology</u>, vol. 70,

     no. 2:307-324.

Wiggins, Lee M.

 1973 <u>Panel Analysis--Latent Probability Models for Attitude and</u>

     <u>Behavior Processes</u>. San Francisco: Jossey-Bass.