297-75

INSTITUTE FOR
RESEARCH ON
POVERTY DISCUSSION
PAPERS

IDENTIFYING STRUCTURAL PARAMETERS OF SOCIAL PROCESSES USING FRAGMENTARY DATA

Burton Singer
and
Seymour Spilerman

UNIVERSITY OF WISCONSIN ~ MADISON

Identifying Structural Parameters of Social Processes Using Fragmentary Data*

Burton Singer
Columbia University

Seymour Spilerman
University of Wisconsin

*The work reported here was supported by grants NSF-GP-31505X and NSF-GS-38574 at Columbia University, and by funds granted to the Institute for Research on Poverty of the University of Wisconsin by the Office of Economic Opportunity pursuant to the Economic Opportunity Act of 1964. The conclusions are the sole responsibility of the authors.

ABSTRACT

In this paper we present examples illustrating five important issues involved in the identification of structural information about social processes using the fragmentary data which is usually available for this purpose. These issues are listed below.

(i) It is of considerable importance to use continuous-time models to describe processes for which there is no substantive basis for a choice of basic unit time interval between state changes. Failure to observe this point can result in conclusions about compatibility of data with a discrete-time model which are dependent simply on an ad-hoc choice of unit time interval.

(ii) The trade-off between parsimony and realism in the selection of base-line models is closely linked to a consideration of the currently as well as potentially available data bases which can be used for the purpose of identifying structural mechanisms. Examples of some major sociological data bases are described in order to explicitly illustrate the fragmentary nature of the observations. In addition, special mixtures of Markov and semi-Markov processes are presented as examples of models which are both moderately realistic and simple enough to allow for identification of structural parameters using rather limited data.

(iii) Very little attention has been paid to the development of strategies for discriminating between several apriori plausible models all fitted to the same data. We present an example of testing data for compatibility with the class of continuous-time Markov chains (null hypothesis) vs. a restricted class of mixtures of Markov chains as an alternative.

(iv) The evenly-spaced sampling of Markov semi-groups can lead to the aliasing of intensity matrices. This phenomenon is analogous to the familiar aliasing of frequencies in the evenly-spaced sampling of stationary time series.

(v) Two strategies are presented for assessing the sensitivity of conclusions about structural parameters compatible with a given data set to noise and other uncontrollable sources of variability.

(iii) specification of strategies for discriminating among several apriori plausible models, all fitted to the same data. Section 4 treats an example of testing data for compatibility with time homogeneous Markov chains (the null hypothesis) vs. a restricted class of mixtures of Markov chains as an alternative.

(iv) the possible presence of non-uniqueness in the identification of intensity matrices for Markov chains with stationary transition probabilities. Section 5 illustrates this phenomenon, which is analogous to the familiar aliasing of frequencies in the evenly-spaced sampling of stationary time series.

(v) assessment of the sensitivity of conclusions about structural parameters compatible with a given data set to "noise" and other uncontrollable sources of variability. Two strategies for this kind of investigation are illustrated in Section 6.

2. DISCRETE VS. CONTINUOUS TIME

Even though most empirical processes in sociology and economics evolve continuously, it is not uncommon for the dynamic models which are fitted to data to be of the difference equation type. It is thereby presumed that the process under investigation can be described adequately by a discrete-time structure defined on integer multiples of a basic unit time interval, call it Δ, which may be a week, month, quarter, year, several years, etc. However, in many situations where these models have been used (e.g., studies of occupational mobility, geographic migration) there is no substantive basis for establishing a particular time interval as a natural unit for the empirical process. If we denote by $t_0 = 0$, t_1, $2t_1$, $3t_1, \ldots$, etc. the evenly spaced time points at which observations usually are taken (e.g., quarterly or yearly), a common practice is to associate these observation times with the natural spacing unit for the process under study. This practice can lead to serious difficulties in attempting to identify underlying structural mechanisms which could have generated the data at hand.

Example 1:

Suppose, for simplicity, that you have observations collected yearly (i.e., t_1 = one year) on a population which simply switches back and forth between two distinct states. Suppose, further, that you entertain discrete-time Markov chains with stationary transition probabilities as an initial base-line class of models to be compared with your data. (Non time-homogeneous chains, or models incorporating longer range dependence, can be viewed as alternatives to the null hypothesis of Markov chains with stationary transition probabilities.) Now assume that you have three empirically determined stochastic matrices given by[1]

$$\left.\begin{aligned} \hat{P}(0,t_1) &= \hat{P}(t_1,2t_1) = \begin{pmatrix} 1/4 & 3/4 \\ 5/8 & 3/8 \end{pmatrix} \\ \hat{P}(0,2t_1) &= \begin{pmatrix} 17/32 & 15/32 \\ 25/64 & 39/64 \end{pmatrix} \end{aligned}\right\} \tag{2.1}$$

where $\hat{P}(kt_1,\ell t_1)$, $0 \leq k < \ell$, has entries $\hat{p}_{ij}(kt_1,\ell t_1)$ = {proportion of individuals in state i at time kt_1 who are in state j at time ℓt_1}.

If you identify t_1 with the natural time unit Δ for the evolution of the population, then a test of compatibility of the data (2.1) with a discrete-time Markov chain structure (stationary transition probabilities) consists of asking whether there exists at least one stochastic matrix M such that

$$\left.\begin{aligned} &\hat{P}(0,t_1) = \hat{P}(t_1,2t_1) = M \\ \text{and} \quad &\hat{P}(0,2t_1) = M^2 \end{aligned}\right\} \tag{2.2}$$

The entries m_{ij} in M represent structural information about the population and have the interpretation, "propensity of an individual in state i to move to state j in one unit of time." Clearly, (2.2) is satisfied with the obvious choice

$$M = \begin{pmatrix} 1/4 & 3/4 \\ 5/8 & 3/8 \end{pmatrix}$$

and you might tentatively settle for the discrete-time Markov chain model

$$P(kt_1, \ell t_1) = M^{(\ell-k)t_1}, \qquad k < \ell \tag{2.3}$$

as a description of the data.

Now suppose that you consider the basic time interval Δ in the evolution of the population under study to be six months. Then your annual data remains the same, while a test for model compatibility reduces to asking whether there is a stochastic matrix M such that

$$\hat{P}(0,t_1) = \hat{P}(t_1,2t_1) = M^2$$

and $$\hat{P}(0,2t_1) = M^4 .$$

Since the matrix

$$\begin{pmatrix} 1/4 & 3/4 \\ 5/8 & 3/8 \end{pmatrix}$$

has no stochastic square roots,[2] there is no discrete-time Markov chain with stationary transition probabilities which can describe the data in (2.1).[3] The essential point to be noted is that a substantively based judgment simply on the natural time unit for the underlying process can make a considerable difference in the conclusions which are reached about model compatibility.

Some still more striking features of the present example are:

1. $\hat{P}(0,t_1)$ has no even stochastic roots of any order.
2. $\hat{P}(0,t_1)$ has a stochastic cube root, a stochastic fifth root, but no odd stochastic roots of order greater than 5.

This means that by choosing a natural time unit for the underlying process equal to four months or 2.4 months, the data would be compatible with a Markov chain having one-step transition probabilities given by

$$\sqrt[3]{\hat{P}(0,t_1)} = \begin{pmatrix} .0611 & .9389 \\ .7824 & .2176 \end{pmatrix} \qquad \text{(for } \Delta = \text{4 month interval)}$$

and $$\sqrt[5]{\hat{P}(0,t_1)} = \begin{pmatrix} .0063 & .9937 \\ .8281 & .1719 \end{pmatrix} \qquad \text{(for } \Delta = \text{2.4 months)}$$

Any other time unit of the form $\frac{t_1}{n}$ shorter than one year leads to a conclusion of model incompatibility.

This phenomena is characteristic of any application of discrete-time models that lacks a solid substantive basis for a choice of unit-time interval. A more natural strategy would be to test the data for compatibility with one or more continuous-time models in which the waiting times between transitions by individuals are viewed as random variables having distributions belonging to special parametrized families. We note, in this connection, that the data (2.1) would not be compatible with any continuous-time Markov chain having stationary transition probabilities. This is a consequence of the fact that $\hat{P}(0,t_1) = \hat{P}(t_1,2t_1)$ cannot be written as $e^{t_1 Q}$, and $\hat{P}(0,2t_1)$ cannot be written as $e^{2t_1 Q}$, where $Q \varepsilon \underline{\underline{Q}}$, $\underline{\underline{Q}} = \left\{Q: q_{ii} \leq 0, q_{ij} \geq 0 \text{ for } i \neq j, \sum_j q_{ij} = 0\right\}$. A necessary and sufficient condition for 2 x 2 matrices to have the representation e^Q with $Q \varepsilon \underline{\underline{Q}}$ is that trace $(\hat{P}) > 1$ (see e.g., Kingman [1962] and Singer and Spilerman [1975]). The point to be made here, however, is that this conclusion about compatibility of the data with a continuous-time model requires no ad-hoc decision about unit time intervals, and is also invariant under a change in the spacing between observations, t_1.

3. MIXTURES AND FRAGMENTARY DATA

A family of base-line models which have been informative about underlying structural mechanisms in social mobility studies are mixtures of independent Markov and semi-Markov processes. The role which these mathematical structures play in empirical sociological investigations is to provide a simple, yet moderately realistic framework, against which data may be compared. The most useful substantive information obtained in these comparisons has arisen from pronounced and highly structured sets of residuals (or violations) from the models. These violations often serve to suggest alternative conceptualizations of the directly observable data, as well as special forms of more realistic models. A well known instance of the revealing nature of residuals from discrete-time Markov models was the large discrepancy (first observed by Blumen, Kogan, and McCarthy [1955]) between diagonal entries in empirically determined stochastic matrices, which were based on observations taken at widely spaced time points, and predictions based on the powers of a one-step transition matrix fitted to observations taken at closely spaced time points.

Among the multiple explanations which could plausibly account for this kind of discrepancy, the treatment of a socially heterogeneous population as though it was homogeneous has been singled out as not only an important explanation, but also revealing of a basic weakness in many mobility studies. These observations have led to the investigation of meaningful bases for the classification of a heterogeneous population into homogeneous sub-populations (e.g., classification according to rate of movement or classification according to propensity to move in a particular pattern (Spilerman [1972a], [1972b]; McFarland [1970]; Singer and Spilerman [1974])). In addition, when detailed individual histories cannot be obtained for persons in the sub-populations (a typical situation in empirical sociology), mixtures of Markov and semi-Markov models provide a parsimonious representation of the observed population-level data in terms of the non-directly observable movement which occurs in the individual sub-populations.

We illustrate these ideas with a class of concise models which are well adapted to the classification of sub-populations according to the <u>rate</u> at which transitions occur.

<u>Example 2</u>:

Consider a homogeneous sub-population whose evolution is described by two independent processes: (i) a discrete-time Markov chain with stationary transition probabilities which characterizes movements between states at their times of occurrence, and (ii) a sequence of independent, identically distributed positive random variables which describe the waiting times between transitions.

A convenient construction of the sample paths for these processes is based on first introducing the sequence $\tau_1^{(\gamma)}, \tau_2^{(\gamma)},\ldots$ of independent identically distributed random variables such that $\text{Prob}(\tau_1^{(\underline{\gamma})} \leq u) = F_{\underline{\gamma}}(u)$, $[\underline{\gamma} = (\gamma_1,\ldots,\gamma_\sigma)$ is a vector of real parameters], where $F_{\underline{\gamma}}(u)$ is a parametrized family of continuous distribution functions which represent moderately realistic waiting time distributions. The sequence $\tau_1^{(\underline{\gamma})}, \tau_2^{(\underline{\gamma})},\ldots$ denotes the waiting times between transitions for what we will call type-$\underline{\gamma}$ individuals. Now, introduce the discrete-time Markov chain $\{X(k)\}_{k=0,1,\ldots}$ having one-step transition matrix M, and define the continuous-time process

$$Y_{\underline{\gamma}}(t) = X(T_{\underline{\gamma}}(t)) \tag{3.1}$$

where $T_{\underline{\gamma}}(t) = \max\{n: \sum_{i=1}^{n} \tau_i^{(\underline{\gamma})} \leq t\}$ = {total number of transitions by a type-$\underline{\gamma}$ individual up to time t}. $Y_{\underline{\gamma}}(t)$ is a special type of semi-Markov process in which the kernals

$$Q_{ik}(u) = \text{Prob}\left(X(n) = k, \ \tau_n^{(\underline{\gamma})} \leq u \,\middle|\, X(0);\ X(1),\ \tau_1^{(\underline{\gamma})};\ldots;X(n-1) = i,\ \tau_{n-1}^{(\underline{\gamma})}\right)$$

have the form (see Pyke [1961])

$$Q_{ik}(u) = m_{ik}\, F_{\underline{\gamma}}(u) \tag{3.2}$$

<u>Mixtures</u> of stochastic processes having the structure (3.1) can now be defined by introducing a random vector $\underline{Z} = (Z_1,\ldots Z_\sigma)$ taking values $\underline{\gamma}$ in the parameter set associated with a specific class of distributions $F_{\underline{\gamma}}(u)$ and independent of the Markov chain $\{X(k)\}_{k=0,1,\ldots}$. Then define the mixture process as

$$Y_{\underline{Z}}(t) = X(T_{\underline{Z}}(t)) \tag{3.3}$$

where Prob $(Z_1 \leq \gamma_1,\ldots,Z_\sigma \leq \gamma_\sigma)$ is specified by a mixing measure μ defined on the parameter space.

Important examples of mixing measures μ and waiting time distributions $F_{\underline{\gamma}}(u)$ adapted to social mobility studies are:

(i) $F_{\gamma}(u) = 1 - e^{-\gamma u}, \qquad \gamma > 0, \quad u \geq 0$

$\mu\{0\} = \text{Prob } (Z = \{0\}) = s$

$\mu\{\gamma_0\} = \text{Prob } (Z = \{\gamma_0\}) = 1 - s$

$Y_Z(t)$ is then a continuous-time analog of the classical mover-stayer model where s represents the proportion of stayers in the total population.

(ii) $F_{\gamma}(u) = 1 - e^{-\gamma u}, \qquad \gamma > 0, \quad u \geq 0$

$$\mu([0,y]) = \text{Prob } (Z \leq y) = \int_0^y \frac{\beta^{\alpha}\gamma^{\alpha-1} e^{-\beta\gamma}}{\Gamma(\alpha)} d\gamma \qquad (3.4)$$

$$\beta > 0, \quad \alpha > 0$$

Then $Y_Z(t)$ is a mixture of continuous-time Markov chains where type-γ individuals have an expected waiting time until a transition given by $\frac{1}{\gamma}$, and they occur in the total population with a frequency governed by the Gamma distribution (3.4) with parameters (α,β).

(iii) $F_{\underline{\gamma}}(u) = 1 - e^{-\gamma_1 u^{\gamma_2}}, \qquad \gamma_1 > 0, \quad 0 < \gamma_2 < 1$

$$\mu([0,y)\times[0,w)) = \text{Prob } (Z_1 \leq y, \ Z_2 \leq w)$$

$$= \int_0^y \frac{\beta^{\alpha}\gamma^{\alpha-1}e^{-\beta\gamma}}{\Gamma(\alpha)} d\gamma \int_0^w dz$$

$$y\varepsilon[0,\infty), \ w\varepsilon[0,1]$$

$Y_{\underline{Z}}(t)$ is a mixture of semi-Markov processes where type-$\underline{\gamma} = (\gamma_1,\gamma_2)$ individuals have duration-dependent waiting times such that the longer an individual stays in a particular state the less likely he is to move in the immediate future. This is a formalization of R. McGinnis' [1968] idea of cumulative inertia (see also Ginsberg [1971] and Singer and Spilerman [1974]). Then the type-$\underline{\gamma}$ individuals are distributed in the total population with γ_1 having a Gamma density, and γ_2 following a uniform density on [0,1].

The trade-offs between parsimony and realism in the selection of base-line models is closely linked to a consideration of the currently, as well as potentially, available data bases which can be used for the purpose of identifying structural mechanisms. The most desirable form of observations for this purpose is a collection of continuous individual histories. These records would be identified with finite sections of sample paths in the stochastic process models, and would provide the ideal basis for testing and discriminating among apriori plausible models as well as for parameter estimation. The vast majority of literature on statistical inference for Markov and semi-Markov processes concerns itself with records of this type (Anderson and Goodman [1957]; Billingsley [1961]; Albert [1962]) and emphasizes large sample properties of testing and estimation procedures. The issues involved in model identification and comparison (on both substantive and numerical grounds) of several models fitted to the same data has received considerably less attention, even for sample-path data. For a notable exception, however, see Bush and Mosteller [1961].

Even though accurate continuous individual histories would be desirable, they are with rare exception (e.g., J. Coleman [1975]; A. Sørensen [1972]) not presently available; and even current plans for large-scale demographic and social accounting systems (see R. Stone [1972]) do not envision the collection of sample-path data, or involve extensive consideration of sampling schemes which would facilitate efficient comparison and discrimination among multiple plausible models. We therefore focus on strategies for model estimation and model discrimination which require only the more limited sorts of data that census bureaus and allied agencies normally produce.

In the following discussion we will refer to any collection of observations which contains less information than long finite sections of sample paths as fragmentary. Three of the most common data gathering situations which are currently encountered in empirical sociology are exhibited in examples 3-5.

Example 3:

Let $Y^{(i)}(t)$ represent the state of the i^{th} individual (person, sub-population, size of organization, etc.) at time t. Then, given a spacing $\Delta > 0$ and times $t_0 = 0$, $t_1 = \Delta$, $t_2 = 2\Delta, \ldots$, we observe $\{Y^{(i)}(k\Delta)\}$ where $1 \leq i \leq N$, N = {number of individuals in a closed population under study}; and $0 \leq k \leq n$, n = {total number of time points after t_0 at which observations are obtained}.

This is precisely the sampling situation in Blumen, Kogan, and McCarthy [1955] with $\Delta = 3$ months. Here the states are industrial categories, and $Y^{(i)}(k\Delta)$ is the category of the i^{th} individual at time $k\Delta$ as recorded in the Social Security Administration's Work History File [1972].

Example 4:

Consider the same kind of observations as in example 3 but augmented by $T^{(i)}(0,k\Delta)$ = {number of transitions by the i^{th} individual between time 0 and time $k\Delta$}. This kind of information was obtained in the social mobility studies of Palmer [1954], Lipset and Bendix [1963], and in the much larger study by Parnes [1972]. It is fragmentary relative to sample path data in that the full set of transition times, as well as a complete list of states which are visited between the sampling times, is missing.

Example 5:

Observations are taken retrospectively on current residence, first and second prior residence, and birthplace of individuals in particular age cohorts. This kind of data was collected in Taeuber's Residence History Study (Taeuber, et. al. [1968]), and represents an instance of fragmentary information about a migration process in that gaps are present in the residence histories.

The basic problem which confronts a researcher who desires structural information about a population, and who must settle for data such as in the above examples, is the need for numerical strategies which allow matrices such as M in example 2 to be recovered using the data at hand. An illustration of such a model identification strategy forms the content of section 4.

4. TESTING OF MODELS AND IDENTIFICATION OF STRUCTURAL PARAMETERS

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 time-homogeneous Markov chains (the null hypothesis) vs. a restricted class of mixtures of Markov chains. To fix the ideas, assume that observations

$$\{Y^{(i)}(t_k),\ T^{(i)}(0,t_k)\}, \qquad 1 \le i \le N, \quad 0 \le k \le n \tag{4.1}$$

have been collected at the evenly spaced time points $0 = t_0 < t_1 <\ldots< t_n$, where $t_{k+1} - t_k = \Delta > 0$, $k = 0,\ldots,n-1$; $Y^{(i)}(t_k)$ denotes the state of the i^{th} individual in the survey at time t_k; and $T^{(i)}(0,t_k)$ equals the total number of transitions by the i^{th} individual in the time interval $(0,t_k)$. Also, consider $n < 10$ to be a typical situation for empirical sociology.

Now introduce the Markov models with sample path representations of the form (3.1) and $F_\gamma(u) = 1 - e^{-\gamma u}$, $\gamma > 0$, $u > 0$. This class of models is to be tested on data of the form (4.1) in competition with mixture models having sample path representations of the form:

$$Y_Z(t) \;=\; X(T_Z(t)) \tag{4.2}$$

where $F_\gamma(u) = 1 - e^{-\gamma u}$, $\quad \gamma > 0,\ u > 0$

and $\text{Prob}(Z \le y) \;=\; \int_0^y \frac{\beta^\alpha \gamma^{\alpha-1} e^{-\beta\gamma}}{\Gamma(\alpha)}\, d\gamma$

Identification of structural information in this setting means:

(a) deciding which of the models, if any, provides a good description of the data,

(b) calculating the transition matrix (or possibly several matrices) M whose entries represent propensities of individuals to move between any pair of states. Diagonal entries of M represent within-state mobility, such as within-industry job change.

A strategy for fitting data of the form (4.1) to a Markov model, and simultaneously testing it for compatibility with this structure, consists of the following steps:

(1) Calculate the separate exponential parameter estimates

$$\gamma^{(k)} = \frac{1}{k\Delta N} \sum_{i=1}^{N} T^{(i)}(0,k\Delta) \qquad \text{for } k = 1,\ldots,n.$$

These values should be roughly equal if the data are compatible with the proposed model. A final summary estimate of γ, call it $\hat{\gamma}$, can then be produced by averaging $\{\gamma^{(k)}\}_{k=1,\ldots,n}$ possibly with weights.

(2) Form the matrices $\hat{P}(i\Delta,j\Delta)$, $0 \leq i < j \leq n$, with entries

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

Then calculate

$$\frac{1}{\hat{\gamma}(j-i)\Delta} \log \hat{P}(i\Delta,j\Delta) \qquad \text{for } 0 \leq i < j \leq n \qquad (4.3)$$

and check whether or not there is at least one branch of the logarithm which can be written as M-I, where M is a stochastic matrix. (See Singer and Spilerman [1975] for details on these calculations). If the data are compatible with a Markov model $\left[\text{i.e., representable as } \hat{P}(i\Delta,j\Delta) = e^{\hat{\gamma}(j-i)\Delta(M-I)}\right]$, then at least one branch of the logarithm of any given matrix in the list (4.3) should be roughly equal to some branch of the logarithm of any other matrix in the list. In addition, this common logarithm should be representable as M-I, with M stochastic. A final summary estimate of M can be calculated by averaging the separate estimates

$$I + \frac{1}{\hat{\gamma}(j-i)\Delta} \log \hat{P}(i\Delta,j\Delta), \qquad 0 \leq i < j \leq n \qquad (4.4)$$

In mobility data for which estimates of the form (4.3) are appropriate (e.g., Blumen, Kogan, and McCarthy [1955]), it has sometimes been found that

$$I + \frac{1}{\hat{\gamma}\Delta} \log \hat{P}(i\Delta,[i+1]\Delta), \qquad i = 0, 1,\ldots,n-1 \qquad (4.5)$$

is roughly constant, but that the common value of these matrices differs substantially from the terms in (4.4) for (j-i) large (i.e., widely spaced time points). An average of the matrices based on closely spaced data for a few time points can then be viewed as a reasonable estimate of M, but with the proviso that the Markov model only provides an adequate description of the data for relatively short time stretches. In this situation, a useful procedure for detecting possible compatibility of the <u>full</u> data set with a mixture of the form (4.2) is to compare $\hat{P}(0,k\Delta)$ for k = n, n-1 and n-2 (i.e., matrices at widely spaced time points) with forecasts of these matrices by $e^{k\Delta\bar{\gamma}(\bar{M}-I)}$, where $\bar{\gamma}$ and $\bar{M}$ are estimates based on closely spaced observations. Blumen, Kogan, and McCarthy [1955] found that this kind of comparison yielded a pronounced discrepancy of the form

$$\hat{p}_{ii}(0,k\Delta) > \left(e^{k\Delta\bar{\gamma}(\bar{M}-I)}\right)_{ii} \qquad \text{for } k = n,\ n-1,\ n-2 \qquad (4.6)$$

and observed that one possible explanation for this inequality was that a socially heterogeneous population was being treated as though it was homogeneous.

The mixture (4.2) incorporates a form of heterogeneity in which the observed population is viewed as being comprised of type-γ individuals with frequency given by the 2-parameter family of distributions (3.4). In order to assess whether data which deviates from a Markov model according to (4.6) is actually compatible with a mixture of the form (4.2), the following strategy can be utilized:

(i) Under the hypothesis that (4.2) is the correct model, $T_Z(t)$ has a negative binomial distribution

$$\text{Prob}(T_Z(t)=k) = \binom{\alpha+k-1}{k}\left(\frac{t}{\beta+t}\right)^k\left(\frac{\beta}{\beta+t}\right)^\alpha$$

We estimate the parameters (α,β) using the observations $\{T^{(i)}(0,j\Delta)\}_{1\le i\le N}$ and call any such estimates $(\hat{\alpha}_j,\hat{\beta}_j)$. If the data are compatible with the mixture model, then we should find $(\hat{\alpha}_1,\hat{\beta}_1) \approx (\hat{\alpha}_2,\hat{\beta}_2) \approx \cdots \approx (\hat{\alpha}_n,\hat{\beta}_n)$.

(ii) With the estimates $(\hat{\alpha}_j,\hat{\beta}_j)$ at hand, observe that $P(0,j\Delta) = \int_0^\infty e^{j\Delta\gamma(M-I)}d\mu(\gamma)$ can be written as

$$P(0,j\Delta) = \left(\frac{\beta}{\beta+j\Delta}\right)^\alpha\left[I - \frac{j\Delta}{\beta+j\Delta}M\right]^{-\alpha}, \qquad j = 1,2,\ldots,n$$

Now solve the equations

$$\hat{P}(0,j\Delta) = \left(\frac{\hat{\beta}_j}{\hat{\beta}_j + j\Delta}\right)^{\hat{\alpha}_j}\left[I - \frac{j\Delta}{\hat{\beta}_j + j\Delta}M\right]^{-\hat{\alpha}_j}, \qquad 1\le j\le n$$

for M and check whether each calculation yields roughly the same matrix and that it is, in fact, a stochastic matrix. If the empirically determined matrices $\hat{P}(0,k\Delta)$, $1 \le k \le n$, satisfy this requirement, then the n stochastic matrices computed above may be averaged to produce a final summary estimate of M. In addition, passage of these tests supports the contention that the full set of data is compatible with the mixture model (3.3). Detailed technical recommendations concerning the estimation and computation procedures used in this model identification strategy are currently in preparation.

5. NON-UNIQUENESS AND ALIASING

An unavoidable consequence of estimating models involving some form of stationarity from data collected at evenly spaced time points is the presence of alias patterns. Aliasing is a widely recognized phenomenon arising in the evenly spaced sampling of stationary time series, and it leads to difficult questions in the interpretation of power spectral estimates (see Blackman and Tukey [1958] pp. 30-35). The essence of aliasing in that context is simply that several distinct frequencies may all be compatible with a given set of evenly spaced data. This places a researcher in the position of having to provide external substantive arguments in order to rule out some of those frequences as being physically meaningless, and thereby identifying the structural information supported by the data. These judgments are often difficult to render and represent a form of non-uniqueness in the identification of structural parameters (i.e., the primary frequencies) which is entirely analogous to the aliasing of intensity matrices in mobility investigations.

We illustrate this phenomenon in the simple context of evenly spaced sampling from Markov chains with stationary transition probabilities. For a detailed discussion of the computational issues as well as interpretation problems in empirical sociology, the reader should consult Singer and Spilerman [1975]. In the present discussion, however, we simply wish to emphasize the possible presence of aliasing in model identification situations where Markov chains, as well as their mixtures, are utilized.

Example 6:

Suppose observations are taken at times $t_0 = 0 < t_1 < t_2 < \ldots.$ with $t_{k+1} - t_k = \Delta > 0$, and consist of matrices $P(i\Delta, j\Delta)$ having the representation $e^{(j-i)\Delta Q_0}$ where Q_0, the intensity matrix of a 3-state Markov chain, has the special form

$$Q_0 = \begin{pmatrix} -(q_1+q_2) & q_1 & q_2 \\ q_2 & -(q_1+q_2) & q_1 \\ q_1 & q_2 & -(q_1+q_2) \end{pmatrix} \qquad (5.1)$$

where $q_1 > 0$, $q_2 > 0$ (i.e., Q_0 is a 3 x 3 circulant matrix).

The question we wish to address is whether the matrix (5.1) could be uniquely identified from a knowledge of the matrices $\{P(i\Delta, j\Delta)\}$, $0 \leq i < j \leq n$. Since the extent of possible non-uniqueness increases with the spacing between observation times (see e.g., Cuthbert [1973]), it is sufficient to check whether or not

$$\frac{1}{\Delta} \log P(i\Delta,[i+1]\Delta) = \frac{1}{\Delta} \log e^{\Delta Q_0} \tag{5.2}$$

has a unique logarithm, namely Q_0, in the class of intensity matrices $\underline{\underline{Q}} = \{Q: q_{ii} \leq 0;\ q_{ij} \geq 0,\ i \neq j;\ \sum_j q_{ij} = 0\}$. Assume, for illustrative purposes, that $q_1 > q_2$. Then (5.2) would uniquely identify Q_0 if and only if the spacing between observation times satisfied

$$\Delta < \frac{2\pi}{\sqrt{3}\, q_1} \tag{5.3}$$

For all spacings $\Delta \geq \frac{2\pi}{\sqrt{3}\, q_1}$, the calculations (5.2) would yield not only Q_0 as in (5.1), but also additional intensity matrices which play the role of aliases. Figure 1 shows the relation between $L(\Delta)$ = (number of branches of $\log e^{\Delta Q_0}$ which are intensity matrices) and Δ = (spacing of the observation times).

<u>Figure 1 about here</u>

The critical spacings[4] Δ^- and Δ^+ at which $L(\Delta)$ increases are given by

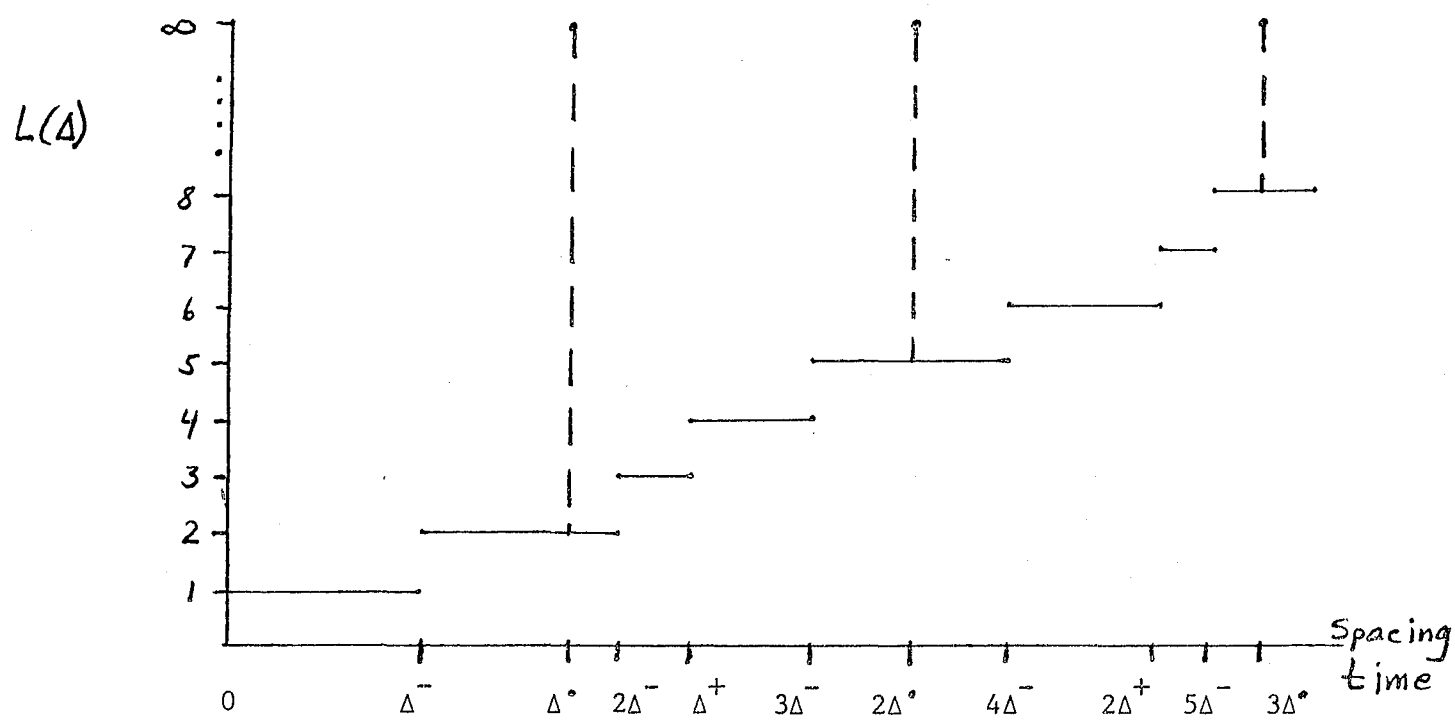
$$\Delta^- = \frac{2\pi}{\sqrt{3}\, q_1}, \qquad \Delta^+ = \frac{2\pi}{\sqrt{3}\, q_2}$$

All aliases except those occurring at $\Delta = k\Delta^*$ are of the form

$$Q_0 + \frac{2\pi \ell}{\Delta\sqrt{3}} \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$$

where ℓ can assume all integer values such that

Figure 1. Number of Branches of the Logarithm which are Intensity Matrices, as a Function of the Spacing Interval



$$\frac{-q_1 \Delta\sqrt{3}}{2\pi} \leq \ell \leq \frac{q_2 \Delta\sqrt{3}}{2\pi}$$

A researcher confronted with this list could only choose among the candidates on external substantive grounds. For an indication of the variety of critical spacings (Δ^-, Δ^+) in $L(\Delta)$ which can occur just within the class of 3 x 3 circulant intensity matrices Q_0 (with $q_1 > q_2$), observe that

$$\frac{\Delta^+}{\Delta^-} = \frac{q_1}{q_2}$$

and the only constraint on this ratio is

$$1 < \frac{\Delta^+}{\Delta^-} < +\infty$$

The isolated spacings, $k\Delta^* = \dfrac{2\pi k}{\sqrt{3}\,(q_1 - q_2)}$, $k = 1,2,\ldots$, at which $L(k\Delta^*) = +\infty$, correspond to a situation in which $e^{k\Delta^* Q_0}$ has repeated eigenvalues with non-distinct elementary divisors. This, in turn, leads to a continuum of branches of $\log e^{k\Delta^* Q_0}$ which are intensity matrices. For a detailed discussion of this phenomenon, first explicitly displayed by J. Cuthbert [1973], see Singer and Spilerman [1975].

The non-uniqueness illustrated in this example is typical of what can happen in evenly spaced sampling of a Markov semi-group whose intensity matrix has at least one pair of complex conjugate eigenvalues. However, a detailed understanding of the behavior of the function $L(\Delta)$ for a variety of special semi-groups arising in sociology, demography, and economics lies in the future.

6. ERROR STRUCTURES AND SENSITIVITY

Our previous discussion of testing and identification strategies has treated observations as though they were error-free. However, in any of these data collection situations, the observations are contaminated by a variety of errors and other uncontrollable sources of variability. Among the most common influences of this sort are: misclassification of individuals, non-response to a survey, persons dropping out of a study before its completion, and variation in background profile of individuals.[5] Realistic models of these sources of variability and formal specification of their relationships to the structural mechanisms discussed earlier are currently in the preliminary development stage. Nevertheless, some useful perturbation strategies can be recommended as a means of assessing the sensitivity of identified structural parameters to noise in the data.

The importance of such sensitivity analyses is heightened, for example, by the fact that a given set of empirically determined matrices may be representable in the form $\left\{e^{(j-i)\Delta Q_0}\right\}$, $0 \leq i < j \leq n$, while matrices within "error distance" of the data may be expressed in the form $\left\{e^{(j-i)\Delta Q_1}\right\}$, $0 \leq i < j \leq n$, where Q_0 and Q_1 represent very different structural mechanisms. Furthermore, matrices $\hat{P}(i\Delta,j\Delta)$ may not be embeddable in <u>any</u> Markov semi-group, while small perturbations[6] $\tilde{P}(i\Delta,j\Delta) = \hat{P}(i\Delta,j\Delta)\cdot$(small perturbation) may well be representable as $\tilde{P}(i\Delta,j\Delta) = e^{(j-i)\Delta Q}$ for some $Q \in \underline{\underline{Q}}$.

<u>Strategy 1</u>:

Suppose observations are collected at evenly spaced time points $t_0 = 0 < t_1 < t_2 < \ldots t_n$, with $t_{i+1} - t_i = \Delta > 0$, and that they yield matrices $\hat{P}(i\Delta,j\Delta)$ representable in the form $e^{(j-i)\Delta Q_0}$, $0 \leq i < j \leq n$, for $Q_0 \varepsilon \underline{\underline{Q}}$. Futhermore, assume that there is only <u>one</u> $Q_0 \varepsilon \underline{\underline{Q}}$ which generates <u>all</u> of the matrices $\hat{P}(i\Delta,j\Delta)$. In particular, this means that $e^{\Delta Q_0}$ has a <u>unique</u> logarithm in $\underline{\underline{Q}}$, even though more widely spaced observations (those taken at intervals such as $k\Delta = 2\Delta, 3\Delta, \ldots,$ etc.) may yield matrices

$e^{k\Delta Q_0}$ whose logarithms have several branches in $\underset{=}{Q}$. Now compute $e^{(\Delta+h)Q_0}$ for a suitably chosen[7] small h. These calculations generate the matrix $\tilde{P} = e^{\Delta Q_0} \cdot e^{hQ_0}$, where e^{hQ_0} is viewed as a perturbation array. $\tilde{P}$ may now have two branches of its logarithm in $\underset{=}{Q}$.

To make the discussion more specific, recall the circulant example in section 5 and let Δ be a spacing such that the function $L(\Delta) = 1$ (see figure 1), but $L(\Delta+h) = 2$. Then $\tilde{P}$ may be written as $e^{(\Delta+h)Q_0} = e^{(\Delta+h)Q_1}$ with $Q_0 \neq Q_1$. Futhermore, setting $h = 0$ in the semi-group generated by Q_1 would yield $e^{\Delta Q_1}$, which could be viewed as a small perturbation of $e^{\Delta Q_0}$. The point to be made here is that unique identifiability of structural parameters (i.e., Q_0) need not persist under small perturbations of the data. In addition, Q_1 may represent a qualitatively very different mobility mechanism from Q_0. The virtue of the perturbation strategy is that it gives a researcher some insight into the structure of a neighborhood of his data, and thereby can act as a caution on the strength of conclusions which may be reached about Q_0.

<u>Strategy 2</u>:

Suppose that observations are again collected at evenly spaced time points and that they yield matrices $\{\hat{P}(i\Delta,j\Delta)\}$, $0 \leq i < j \leq n$, such that

$$\frac{1}{(j-i)\Delta} \log \hat{P}(i\Delta,j\Delta) \in \underset{=}{Q} \tag{6.1}$$

for most pairs (i,j). However, assume that several matrices, say $\hat{P}(0,\Delta)$, $\hat{P}(3\Delta,4\Delta)$, and $\hat{P}(0,3\Delta)$, do not satisfy (6.1). These violations may be attributable simply to sampling variability and other sources of error, and we are interested in knowing whether small perturbations in $\hat{P}(0,\Delta)$, $\hat{P}(3\Delta,4\Delta)$, and $\hat{P}(0,3\Delta)$ would be consistent with (6.1) <u>and</u> coincide with the logarithms of those matrices which already satisfy (6.1). To this end, compute the matrices Q_1, Q_2, and Q_3 which are solutions of the respective variational problems

$$\text{(i)} \quad \min_{Q \in \underline{Q}} \left\| \frac{1}{\Delta} \log \hat{P}(0,\Delta) - Q \right\|$$

$$\text{(ii)} \quad \min_{Q \in \underline{Q}} \left\| \frac{1}{\Delta} \log \hat{P}(3\Delta,4\Delta) - Q \right\|$$

$$\text{(iii)} \quad \min_{Q \in \underline{Q}} \left\| \frac{1}{3\Delta} \log \hat{P}(0,3\Delta) - Q \right\|$$

for a suitable choice of norm $||\cdot||$. Then check whether or not $R_1 = \hat{P}(0,\Delta) - e^{\Delta Q_1}$, $R_2 = \hat{P}(3\Delta,4\Delta) - e^{\Delta Q_2}$, and $R_3 = \hat{P}(0,3\Delta) - e^{3\Delta Q_3}$ may each be regarded as matrices with small enough entries to be classified as "noise." If the answer, based on context dependent judgments of smallness, is affirmative and $Q_1 \approx Q_2 \approx Q_3$, then we would still regard the full set of data as compatible with a continuous-time Markov model. If the answer is negative, then the structure of the residual matrices R_i, $i = 1, 2, 3$, can often suggest other models and conceptualizations of the data.

Strategies 1 and 2 represent only the most rudimentary kinds of sensitivity investigations suited to mobility studies. Much remains to be done in this direction, including consideration of formal error models and intensive examination of residuals such as the matrices $\{R_i\}$ in the above illustration, using some of the data sets mentioned in examples 3, 4, and 5 of section 3.

7. CONCLUSIONS

We have presented simple examples of five issues which are central to the effective analysis of sociological data. However, each of these topics is only in a preliminary stage of development, and extensive investigation of a variety of data sets will be necessary before testing and model identification strategies, suited to empirical sociology, can be placed on a firm foundation. Among the more important open problems related to the issues raised in this paper are:

(i) testing and identification of models incorporating continuous-state spaces (which would be applicable, for example, to the analysis of income dynamics), using rather granular data.

(ii) development of an effective methodology to test observations for compatibility with models such as those in section 3, but which prohibit transitions between particular pairs of states, as specified by a substantive theory.

(iii) testing and identification of models of open heterogeneous social systems, using fragmentary data from several sources and of differing quality.

(iv) specification of designs for observational studies which will facilitate comparisons among several models. This should include a detailed assessment of the appropriateness of retrospective data collection versus a panel study for particular sorts of problems, and a consideration of the optimal frequency of reinterviewing when the panel format is used.

NOTES

1 The symbol "$\wedge$" over a quantity means that it is estimated from data.

2 See F. R. Gantmacher [1960] for a lucid discussion of roots of matrices.

3 In this instance, the empirical process might follow a non time-homogeneous Markov chain (e.g., $\hat{P}(0,t_1) = M_1 \cdot M_2$ where M_1 and M_2 are distinct stochastic matrices regulating successive <u>annual</u> transitions).

4 The symbol Δ^- (respectively, Δ^+) corresponds to the minimal spacing for which a negative (respectively, positive) multiple of $2\pi i$ can be added to the complex conjugate eigenvalues of Q_0 and thereby define eigenvalues of a different intensity matrix, call it Q_1, such that $e^{\Delta^- Q_0} = e^{\Delta^- Q_1}$.

5 Considerations of parsimony may lead to only a few of the major sources of variability in a population being explicitly introduced in a mathematical model. The remaining variables can be informative with respect to residuals from this base-line structure.

6 The symbol "$\bullet$" can mean either matrix addition or matrix multiplication depending on the application.

7 By "suitably chosen" we mean small enough so that $\tilde{P}$ may be considered as being within error distance of the original matrix $\hat{P}$.

REFERENCES

Albert, Arthur. 1962. "Estimating the infinitesimal generator of a continuous-time, finite state Markov process." Annals of Mathematical Statistics 33 (June): 727-753.

Anderson, T. W. and Goodman, L. A. 1957. "Statistical inference about Markov chains." Annals of Mathematical Statistics 28:89-109.

Billingsley, P. 1961. Statistical Inference for Markov Processes. Chicago: University of Chicago Press.

Blackman, R. and Tukey, J. 1958. The Measurement of Power Spectra. New York: Dover Publications.

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.

Bush, R. and Mosteller, F. 1961. "A comparison of eight models." In R. Bush and W. Estes, Studies in Mathematical Learning Theory. Stanford: Stanford University Press.

Coleman, James S. 1975. "Analysis of occupational mobility by models of occupational flow." In K. Land and S. Spilerman, Social Indicator Models. New York: Russell Sage Foundation.

Cuthbert, J. R. 1973. "The logarithm function for finite-state Markov semi-groups." Journal of the London Mathematical Society 6:524-532.

Gantmacher, F. R. 1960. The Theory of Matrices (vol. 1). New York: Chelsea.

Ginsberg, R. 1971. "Semi-Markov processes and mobility." Journal of Mathematical Sociology 1:233-263.

Kingman, J. F. C. 1962. "The imbedding problem for finite Markov chains." Zeitschrifft für Wahrscheinlichkeitstheorie 1:14-24.

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." American Sociological Review 33 (October):712-22.

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.

Pyke, R. 1961. "Markov renewal processes: Definitions and preliminary properties." Annals of Mathematical Statistics 32:1231-1242.

Singer, Burton and Seymour Spilerman. 1974. "Social mobility models for heterogeneous populations." In Herbert L. Costner (ed.), Sociological Methodology 1973-1974. San Francisco: Jossey-Bass.

________. 1975. "The representation of social processes by Markov models." Institute for Research on Poverty, University of Wisconsin, Discussion Paper 272-75.

Social Security Administration. 1972. "Basic statistical data files available to outside researchers." U. S. Department of Health, Education, and Welfare. 3 pages (mimeo).

Sørensen, A. 1972. "The organization of activities in time." Center for Demography and Ecology, University of Wisconsin-Madison, Working Paper 72-1.

Spilerman, Seymour. 1972a. "The analysis of mobility processes by the introduction of independent variables into a Markov chain." American Sociological Review 37 (June):277-94.

________. 1972b. "Extensions of the mover-stayer model." American Journal of Sociology 78:599-627.

Stone. R. 1972. "A Markovian education model and other examples linking social behavior to the economy." Journal of Royal Statistical Society. Series A (vol. 135, Part 4), pp. 511-543.

Taeuber, Karl E., L. Chiazze, Jr., and W. Haenszel. 1968. Migration in the United States. Washington, D. C.: Government Printing Office.