SENSITIVITY OF ESTIMATES OF MARKOV TRANSITION MATRICES TO PERTURBATIONS AND SAMPLING ERROR

by

Stanley Wasserman

Technical Report No. 393
September 1981

Department of Applied Statistics
School of Statistics
University of Minnesota
St. Paul, MN 55108

Support granted, in part, by National Science Foundation grant #SES80-08573. Jack Litzau provided valuable computational assistance, and Christopher Bingham, useful suggestions.

SENSITIVITY OF ESTIMATES OF MARKOV TRANSITION MATRICES TO PERTURBATIONS AND SAMPLING ERROR

ABSTRACT

The effect of small perturbations of the elements of a probability transition matrix on estimated quantities (such as eigenvalues and Q intensity matrices) is studied. The approach taken here utilizes either known sampling or measurement error models or when no error model is available, an exploratory strategy of probing the neighborhood of the transition matrix. Dirichlet distributions for the errors are introduced to ascertain how much transition matrix roots vary. Concentration ellipsoids are recommended as convenient summaries of this variability.

Key words and phrases: Continuous-time Markov chain; Embeddability; Longitudinal data; Eigenvalues; Mover-Stayer Model; Dirichlet distribution; Concentration ellipsoid.

Estimation of transition and intensity matrices for continuous-time Markov chains is an important but neglected area of investigation. Maximum likelihood estimates, with their known shortcomings, are extensively used, and the problem of multiple solutions to maximum likelihood equations has only recently been addressed. Two virtually ignored but significant issues recognized by Singer and Spilerman (1976b) and others, are:

1) Sensitivity analysis of the dependence of estimated quantities (such as eigenvalues, and intensity matrices) on small perturbations in the data;

and

2) Computing confidence limits on estimates to ascertain variability.

The analysis of longitudinal data and exploration of fitted Markov models will remain incomplete without an understanding of (1) and some attempt to address (2).

In this note, we offer one approach to those undeveloped research areas, using perturbation analysis of transition matrices and associated spectral decompositions. The likelihood of sampling or measurement error in a set of longitudinal records is quite large, and such errors will change estimates to some extent. What we would like to know is how much additional variability is introduced. By assuming that these errors affect an empirical transition matrix in a specific way, we can utilize the perturbation approximations to assess the impact of errors on eigenvalues, empirical intensity matrices, and hence, on the embeddability of the original data as a Markov chain.

In the next section, we present the necessary introduction to finite-state, continuous-time, stationary Markov chains, and follow this with a

brief discussion of perturbation theory. Our postulated chain is stationary and Markov, and our assumed data are just two-wave observations on N replicates of the basic chain, for the obvious reason that we must understand the simplest situation before moving on to more complicated ones.

After these introductory sections, we present several models for sampling errors and show that these models can be put into our linear perturbation framework; consequently, the effect of the errors can be studied further. We also describe an exploratory sensitivity analysis that can be conducted when no specific error model is relevant. In section 4, we introduce a Dirichlet probability model for sampling errors in order to place a confidence ellipsoid around the perturbed eigenvalues. Several examples illustrate the usefulness of the ellipsoid. We conclude with a discussion of the implication of these models and perturbations for the embeddability of a set of longitudinal replications as a Markov chain, and for the identification of all intensity matrices that could have generated the observed data.

1. Background

Consider a family of sxs stochastic matrices $P(t)$, $t \geq 0$, with elements $p_{ij}(t)$, that describes the behavior of a stationary, continuous-time, finite-state Markov chain. The quantity $p_{ij}(t)$ is the transition probability from state i at time u to state j at time $u+t$, for all $u \geq 0$.
These transition probabilities obey the following differential equations:

$$\frac{d\,P(t)}{dt} = Q\,P(t), \quad P(0) = I. \tag{1.1}$$

The sxs intensity matrix $Q = (q_{ij})$ has structure

$$q_{ij} \geq 0, \quad i \neq j\ ; \tag{1.2}$$

$$q_{ii} = -\sum_{j \neq i} q_{ij} \leq 0, \quad i=1,2,\ldots,s\ .$$

As is well known, conditional transition probabilities and mean waiting times between transitions are simple functions of the elements of Q. Any array with structure (1.2) is a member of the class of intensity matrices $\mathcal{Q}$.

Solutions of (1.1) are $P(t) = e^{tQ}$, $t > 0$, an equation that expresses the transition probabilities as a function of Q. Furthermore, given P(t), we have the relation $Q = \frac{1}{t} \log P(t)$. If $\log P(t) \in \mathcal{Q}$, it is <u>embeddable</u> as a continuous-time Markov chain; one then attempts to <u>identify</u> all such $Q \in \mathcal{Q}$ that obey (1.1). Much is known about the embeddability problem-specifically, characterizations of embeddable stochastic matrices (Johansen, 1973a, 1973b; Goodman, 1970; Kingman, 1962; Frydman and Singer, 1979), and necessary and sufficient conditions for embeddability for $s = 2$ and 3 (Johansen and Ramsey, 1979; Frydman, 1980). Singer and Spilerman (1974, 1976a, 1977) and Singer and Cohen (1980) have extensively researched the identification problem, presenting discrimination techniques useful for choosing among competing models fitted to the same longitudinal data.

Assume that we have N replications of the basic continuous-time Markov chain, each moving through the s states over time with transitions governed by P(t). The simplest observational scheme, and the one used here, is a record of state occupied at time u and at time u+t, for each replicate. These records are arrayed into a sxs matrix of frequency counts $F = (f_{ij})$. The maximum likelihood estimate of P(t) (Anderson and Goodman, 1957) is

$$\hat{P} = D^{-1}F$$

where $D = \text{diag}(f_{i+})$, a diagonal matrix of row sums. We define λ and h as an eigenvalue and corresponding unit-length (right) eigenvector of $\hat{P}$. The s eigenvalues of $\hat{P}$ are usually arrayed as elements of the diagonal array Λ, and the associated s eigenvectors as columns of the array H; thus

$$\hat{P} = H\Lambda H^{-1} = \sum_{i=1}^{s} \lambda_i h_i h_i^*$$

is the standard Jordan canonical representation of $\hat{P}$, assuming distinct or simple eigenvalues. We note that the rows of H^{-1}, which we denote by $\{h_i^*\}$, are left eigenvectors of $\hat{P}$; and that the largest (in norm) eigenvalue of $\hat{P}$ is unity. The remaining (s-1) eigenvalues may be complex, because of the non-symmetry of $\hat{P}$; if so, they will occur in $r \leq (s-1)/2$ conjugate pairs, and be represented as $\alpha_j \pm i\beta_j$, $j=1,2,\ldots,r$. We reason that the chance of observing multiple roots is quite small, and are confident that nearly all F's encountered in practice yield $\hat{P}$'s with simple eigenvalues.

Since $\hat{P}$, and consequently all estimates of intensity matrices which are simply multiples of the matrix function $\log \hat{P}$, are linear combinations of scalar functions of the eigenvalues of $\hat{P}$, we concentrate on how sampling and/or measurement errors, viewed as perturbations, affect these roots. In the concluding section, we study the potential embeddability of observed data, which depends on the nature and multiplicity of the roots, and comment on possible identification problems caused by errors.

The remainder of this paper is an attempt to answer the questions: What happens to λ and h if $\hat{P}$ is changed slightly by a perturbation due to sampling or measurement error? And which roots are most sensitive to such changes? An error for us is simply a faulty recording of the starting state i and/or the ending state j for some subset of replicates. The answers to these questions have an impact on the embeddability of $\hat{P}$ and the identification of all $Q \in \mathcal{Q}$ compatible with this $\hat{P}$. The answer to the first question is given in the next section, and is provided by Taylor-series expansions of the perturbed eigenvalues and eigenvectors of $\hat{P}$. Justification of this technique can be found in Wilkinson (1965) or Stewart (1973). We address the second question in section 4 by introducing a set of Dirichlet

distributions for our errors, and calculating the moments of the vector of perturbed eigenvalues under this particular probability model. This allows us to place a concentration ellipsoid around the s-dimensional mean vector. We first present a very brief introduction to perturbation theory.

2. Perturbation Theory

Consider the linear perturbation of $\hat{P}$

$$\hat{P}(\varepsilon,G) = \hat{P} + \varepsilon G + O(\varepsilon^2) \tag{2.1}$$

and let $\lambda(\varepsilon,G)$ and $h(\varepsilon,G)$ be the perturbed values of λ and h, where

$$\lambda(\varepsilon,G) = \lambda + \varepsilon\mu + O(\varepsilon^2)$$

$$h(\varepsilon,G) = h + \varepsilon k + O(\varepsilon^2) \tag{2.2}$$

where μ may be complex and k may be $\varepsilon\ \mathbb{C}^s$, if λ is a complex eigenvalue. In order to retain the stochastic property for $\hat{P}(\varepsilon,G)$, we require that G have zero row sums.

Lemma 1: Given the perturbation (2.1), the effect of this perturbation to terms of order ε^2 on a simple eigenvalue λ and corresponding eigenvector h (as specified by 2.2) is

$$\begin{aligned} \mu &= h^*Gh \\ k &= -(\hat{P}-\lambda I)^{-}Gh \end{aligned} \tag{2.3}$$

where $(\ )^{-}$ denotes a generalized inverse.

One proof of the lemma can be found in section 6.4 of Stewart (1973). We could have written down all terms up to order ε^3, but such an accurate approximation will not be required. The matrix $\hat{P}(\varepsilon,G)$ has (approximate) eigenvalues $1,\lambda_2(\varepsilon,G),\ldots,\lambda_s(\varepsilon,G)$, and corresponding (approximate) eigenvectors $h_1,h_2(\varepsilon,G),\ldots,h_s(\varepsilon,G)$. (Note that $k_1=\underset{\sim}{0}$, since h_1 is proportional to $\underset{\sim}{1}$). We can write

the matrix of perturbed eigenvectors as

$$H(\varepsilon,G) = (h_1, h_2(\varepsilon,G), \ldots, h_s(\varepsilon,G))$$

$$= H - \varepsilon H(G)$$

where $H(G)$ is an sxs array with columns $(\hat{P}-\lambda_j I)^{-}Gh_j$.

To find $H(\varepsilon,G)^{-1}$, and thus the spectral decomposition of $\hat{P}(\varepsilon,G)$, we use the following approximation (Stewart, 1973, page 291):

$$(A-E)^{-1} \approx A^{-1}(I+EA^{-1}).$$

Consequently

$$H(\varepsilon,G)^{-1} = H^{-1}[I+\varepsilon \sum_{i=1}^{s} (\hat{P}-\lambda_i I)^{-}Gh_i h_i^*]$$

$$= H^{-1}[I+\varepsilon \sum_{i=1}^{s} \Big(\sum_{\substack{j=1 \\ j\neq i}}^{s} (\lambda_j-\lambda_i)^{-1} h_j h_j^* \Big) G\, h_i h_i^*)]$$

with rows

$$h_i^*(\varepsilon,G) = h_i^* + \varepsilon h_i^* \left\{ \sum_{j=1}^{s} \sum_{\substack{k=1 \\ j\neq k}}^{s} \frac{h_k h_k^* G h_j h_j^*}{(\lambda_k-\lambda_j)} \right\}$$

and one can check that $h_i^*(\varepsilon,G)h_j(\varepsilon,G) = \delta_{ij} + O(\varepsilon^2)$. Finally, the spectral decomposition of $\hat{P}(\varepsilon,G)$ is

$$H(\varepsilon,G)\Lambda(\varepsilon,G)H(\varepsilon,G)^{-1} = \sum_i (\lambda_i + \varepsilon h_i^* G h_i) h_i(\varepsilon,G) h_i^*(\varepsilon,G) + O(\varepsilon^2)$$

(2.4)

$$= \sum_i \lambda_i h_i h_i^* + \varepsilon \sum_i \sum_j (h_i^* G h_j) h_i h_j^* + O(\varepsilon^2)$$

using the fact that

$$h_i(\varepsilon,G)h_i^*(\varepsilon,G) \approx h_i h_i^* + \varepsilon \left\{ \sum_{j\neq i} (\lambda_i-\lambda_j)^{-1} [h_i h_i^* G h_j h_j^* + h_j h_j^* G h_i h_i^*] \right\}$$

Equations (2.1)-(2.4) will be invaluable in our sensitivity analysis. It is unfortunate that simple expressions do not exist when eigenvalues have

multiplicity greater than one. But realistically, multiple roots should occur with a probability near zero. In the following sections, we use these approximations of the perturbed quantities to study the dependence of estimates of Q on slight fluctuations, and attempt to assess the variability of the perturbed roots.

3. Sampling Error Models

There are two situations that should be distinguished. The first sampling/measurement error situation arises when a specific model for the errors, recorded as entries in an array G, is available. In this case, we propose the following structure for errors in $\hat{P}$:

$$\hat{P}(\varepsilon,G) = \hat{P} o(\varepsilon G) + O(\varepsilon^2) \tag{3.1}$$

where the symbol "o" refers to matrix multiplication, Hadamard or element-by-element multiplication (which we denote by $*$), or matrix addition. With addition, we obtain the standard (linear) perturbation scheme as discussed previously. If $o=*$, then we have

$$\hat{P}(\varepsilon,G) = \varepsilon P * G + O(\varepsilon^2) \tag{3.2}$$

a model that applies for example when the replicates can be classified as "movers" or "stayers". If $o=x$, we obtain

$$\hat{P}(E) = \hat{P}E + O(\varepsilon^2) \tag{3.3}$$

where $E = \varepsilon G$, a proposal of Singer and Spilerman (1976a). Such an error structure can be used when $\hat{P}$ represents the "true" or "error-free" transition matrix, $\hat{P}(E)$, the observed array, and E, an array containing misclassification probabilities. Since an example of this last structure is given by Singer and Spilerman (1976a, page 46), we concentrate here just on the additive and the Hadamard multiplicative proposals.

The second situation obtains when no specific error structure is known, but the researcher is interested in exploring the "neighborhood" of the $\hat{P}$ array to determine how sensitive the computed $\hat{Q} = \frac{1}{t}\log \hat{P}$ arrays are to slight perturbations in t. We show in the last part of this section that this sensitivity analysis reduces to a standard perturbation problem, as long as the changes in t remain small.

A. Addition of stayers

In the application of Markov chains to social mobility, many researchers, beginning with Blumen, Kogan, and McCarthy (1955), have noted that individuals are of two types: those that change states according to a Markov chain (the "movers") and those that remain in their starting states (the "stayers"). One can estimate the number of stayers by studying the movers--the sxs F array has a missing diagonal $\{f_{ii}\}$ which must be estimated--and hence the probability $p_{ii}(t)$ that a replicate (either a mover or a stayer) does not transit to another stated during the period (u,u+t) can be calculated.

It may happen that after further study of the population, $\hat{p}_{ii}(t)$ is found to underestimate $p_{ii}(t)$ since the actual number of stayers in state i exceeds $\hat{f}_{ii}$. Thus, if $\hat{P}(E)$ represents the actual probability transition matrix, and $\hat{P}$ the matrix computed prior to the additional information on the population, we find that

$$\hat{p}_{ii}(E) \geq \hat{p}_{ii}, \quad i=1,2,\ldots,s$$

$$\hat{p}_{ij}(E) \leq \hat{p}_{ij} \quad i\neq j \ .$$

Define f_i^* as the additional number of stayers in state i. Then

$$\hat{p}_{ii}(E) = \frac{f_{ii}+f_i^*}{f_{i+}+f_i^*} \tag{3.4}$$

and

$$\hat{p}_{ij}(E) = \frac{f_{ij}}{f_{i+}+f_i^*} \qquad (i\neq j) \tag{3.5}$$

so that

$$\hat{p}_{ii}(E)-\hat{p}_{ii} = \frac{(f_{i+}-f_{ii})f_i^*}{f_{i+}(f_{i+}+f_i^*)}$$

$$= \frac{(f_{i+}-f_{ii})f_i^*}{f_{ii}(f_{i+}+f_i^*)}\ \hat{p}_{ii} \tag{3.6}$$

$$= \varepsilon_{ii}^*\,\hat{p}_{ii}$$

must be added to the diagonal transition probability estimates and

$$\hat{p}_{ij}(E)-\hat{p}_{ij} = -\frac{f_{ij}f_i^*}{f_{i+}(f_{i+}+f_i^*)}$$

$$= -\varepsilon_i^*\,\hat{p}_{ij} \tag{3.7}$$

must be added to the off-diagonal estimates. Because the change in probabilities is proportional to $\hat{P}$, it is natural to represent this sampling error model as the Hadamard multiplicative scheme as in (3.2), with $E = \varepsilon G$, and

$$E = \begin{pmatrix} (1+\varepsilon_{11}^*) & 1-\varepsilon_1^* & \cdots & 1-\varepsilon_1^* \\ 1-\varepsilon_2^* & (1+\varepsilon_{22}^*) & \cdots & 1-\varepsilon_2^* \\ \vdots & \vdots & & \vdots \\ 1-\varepsilon_s^* & 1-\varepsilon_s^* & \cdots & (1+\varepsilon_{ss}^*) \end{pmatrix}$$

If we write this model as $\hat{P}(E) = \hat{P} + (\hat{P}*E-\hat{P}) = \hat{P} + G$, we can use perturbation theory to evaluate the effect of additional stayers on $\hat{P}$. Thus

$$\hat{P}(E) \approx \sum_{i=1}^{s} \lambda_i h_i h_i^* + \sum_{i=1}^{s}\sum_{j=1}^{s} (h_i^*[\hat{P}*E-\hat{P}]h_j)h_i h_j^*$$

$$= h_1 h_1^* + \sum_{i=2}^{s} \lambda_i h_i h_i^* + \sum_i \sum_j h_i^*\hat{P}*Eh_j h_i h_j^* - \sum_i \sum_j h_i^*\hat{P}h_j h_i h_j^* \tag{3.8}$$

$$= \sum_{i=1}^{s}\sum_{j=1}^{s} h_i^*\hat{P}*Eh_j h_i h_j^*$$

since $h_i^{*}\hat{P}h_j = \lambda_i$, when i=j, and zero otherwise. This double sum is not difficult to evaluate and requires only the eigenvectors of $\hat{P}$. It is not true, however, that $\log \hat{P}(E) = \log \hat{P} + \log E$, since Hadamard matrix multiplication is not associative with respect to matrix multiplication; we give an approximate formula for $\log \hat{P}(E)$ in section 5.

Example 1:

Consider the hypothetical observed transition array (from Singer and Spilerman 1978)

$$\hat{P} = \begin{pmatrix} 0.0722 & 0.0347 & 0.8931 \\ 0.0375 & 0.0361 & 0.9264 \\ 0.9264 & 0.0014 & 0.0722 \end{pmatrix}$$

that governs the movement of a heterogeneous population. It is known additionally that this population contains 93.75% movers. If, for simplicity, we take all the $f_{i+} = 10000$, then the addition of the 6.25% stayers yields $f_1^{*} = f_2^{*} = f_3^{*} = (10000/.9375) - 10000 = 666\frac{2}{3}$. Note that the actual transition array for the entire population is $\hat{P}^{*} = .0625\, I + .9375\, \hat{P}$, since the 6.25% stayers move with transition matrix I.

With these f_i^{*}, we find $\varepsilon_{11}^{*} = \varepsilon_{33}^{*} = .8035$, $\varepsilon_{22}^{*} = 1.6672$ and $\varepsilon_1^{*} = \varepsilon_2^{*} = \varepsilon_3^{*} = .0625$, via equations (3.4) and (3.5). Thus,

$$E = \begin{pmatrix} 1.8035 & 0.9375 & 0.9375 \\ 0.9375 & 2.6672 & 0.9375 \\ 0.9375 & 0.9375 & 1.8035 \end{pmatrix}$$

and

$$G = \hat{P}^{*}E - \hat{P} = \begin{pmatrix} 0.0580 & -0.0021 & -0.0559 \\ -0.0023 & 0.0602 & -0.0579 \\ -0.0579 & -0.0001 & 0.0580 \end{pmatrix}$$

is the change in $\hat{P}$ due to the addition of stayers.

Using (3.8), we find that

$$\hat{P}(E) \approx \begin{pmatrix} 0.13020 & 0.03260 & 0.83720 \\ 0.03519 & 0.09629 & 0.86852 \\ 0.86850 & 0.00130 & 0.13020 \end{pmatrix}$$

which is virtually identical to $\hat{P}^*$. Thus, one can study the effect that variable stayer fractions, f_i^* or E has on $\hat{P}(E)$, with only knowledge of $\hat{P}$.

B. Misclassification

Suppose $k \geq 1$ individuals with recorded transitions $i \to j$ in actuality moved from state i to state $j', j' \neq j$. Then $\hat{p}_{ij}$ decreases by kf_{i+}^{-1}, $\hat{p}_{ij'}$ increases by kf_{i+}^{-1}, and the other (s-2) entries in the i^{th} row remain constant. Since the amount of the change is independent of the magnitudes of $\hat{p}_{ij}$ and $\hat{p}_{ij'}$, the effect of these k identical misclassifications on $\hat{P}$ is

$$\hat{P}(\varepsilon) = \hat{P} + \varepsilon G$$

where $\varepsilon = f_{i+}^{-1}$, $g_{ij} = -k$, $g_{ij'} = +k$, and the remaining elements of G are null.

Any misclassifications, regardless of starting state, in which the actual transitions are known, can be modeled in this standard way. Since the row sums of F are likely to differ, a more general model is

$$\hat{P}(E) = \hat{P} + EG$$

where $E = \text{diag}(f_{i+}^{-1})$ and g_{ij} = net inflow(+)/outflow(-) from $i \to j$ due to these misclassifications. Such perturbations, with general E, are utilized in the next section.

Example 2:

The array, with t = 1,

$$\hat{P} = \begin{pmatrix} 0.50 & 0.25 & 0.20 & 0.05 \\ 0.20 & 0.60 & 0.10 & 0.10 \\ 0.20 & 0.20 & 0.50 & 0.10 \\ 0.05 & 0.05 & 0.40 & 0.50 \end{pmatrix}$$

is not embeddable as a continuous-time Markov chain since

$$\hat{Q} = \log \hat{P} = \begin{pmatrix} -0.8598 & 0.4308 & 0.4010 & 0.0280 \\ 0.3796 & -0.6033 & 0.0488 & 0.1749 \\ 0.3796 & 0.3130 & -0.8675 & 0.1749 \\ -0.0612 & -0.0404 & 0.8722 & -0.7705 \end{pmatrix}$$

where q_{41} and q_{42} are less than zero. Hence we ask if this non-embeddability might be due to some misclassification.

There are, of course, many possible G arrays to study. For simplicity, we take a constant $f_{i+} = 1000$; thus $\varepsilon = 0.001$. If we let the first three rows of G be identically zero, and the last row equal (30,25,-30,-25), then, using equation (2.4),

$$\hat{P}(E) = \begin{pmatrix} 0.5000 & 0.2500 & 0.2000 & 0.0501 \\ 0.1999 & 0.5999 & 0.1001 & 0.0999 \\ 0.1999 & 0.1999 & 0.5001 & 0.0999 \\ 0.0800 & 0.0751 & 0.3699 & 0.4751 \end{pmatrix}$$

so that only the last row of $\hat{P}$ is affected. (This array is trivially easy to calculate). What is interesting here, is that now $\log \hat{P}(E) \in \mathcal{Q}$; viz.,

$$\log \hat{P}(E) = \begin{pmatrix} -0.8584 & 0.4306 & 0.3986 & 0.0291 \\ 0.3726 & -0.6069 & 0.0555 & 0.1789 \\ 0.3726 & 0.3094 & -0.8608 & 0.1789 \\ 0.0114 & 0.0054 & 0.8088 & -0.8256 \end{pmatrix}$$

The eigenvalues of $\hat{P}$ are perturbed from $\{1, 0.25, 0.45, 0.40\}$ to (using equations (2.2) and (2.3) $\{1, 0.2515, 0.4236, 0.4000\}$.

C. Exploring the neighborhood of $\hat{P}$

When no specific error structure is available, one can study how sensitive the computed $Q^* = \frac{1}{t} \log \hat{P}(t)$ is to slight perturbations in t. Compute $\hat{P}(k)$ from the representation $\hat{P}(k) = e^{(t+k\Delta)Q^*}$ where $k \in K = \{-k_1, -k_1+1, \ldots, 0, \ldots, k_2-1, k_2\}$, and k_1 and k_2 are positive integers chosen so that $\hat{P}(k)$ is

"within sampling or measurement error" of $\hat{P}(t)$. This is a suggestion of Singer and Spilerman (1976a, pages 46-8), who also advise a close examination of the set of eigenvalues of each array $\hat{P}(k)$, $k\epsilon K$, to determine the stability of Q^*. We show how this sensitivity investigation can be viewed as a standard perturbation problem, provided $k\Delta$ remains small.

We let $\hat{P} = e^{tQ^*} = e^{Q_o}$ be our observed transition matrix, and consider $\hat{P}(k) = \exp\{(t+k\Delta)Q^*\}$, as the perturbation. Let $\epsilon = \Delta/t$, and note that

$$\begin{aligned}\hat{P}(k) &= \exp\{Q^*(t+k\Delta)\} \\ &= \exp\{tQ^*\}\exp\{k\Delta Q^*\} \\ &= \hat{P}\exp\{k\epsilon Q_o\}\end{aligned}$$

If we expand $\exp\{k\epsilon Q_o\} = I + k\epsilon Q_o + \frac{1}{2}(k\epsilon)^2 Q_o^2 + \ldots$, then

$$\hat{P}(k) = \hat{P} + \epsilon(k\hat{P}Q_o) + \epsilon^2(\frac{k^2}{2}\hat{P}Q_o^2) + O(\epsilon^3)$$

so we have a perturbation of $\hat{P}$, with $G = k\hat{P}Q_o$, and the perturbation approximations can be used to simplify the calculation of eigenvalues, and $\hat{P}(k)$ itself. Note that $\hat{P}Q_o$ has the necessary zero row-sum property. The equation (2.4) will be computationally easier, since it involves only one set of initial matrix multiplications (only k varies), than the exact computation $\hat{P}(k) = H\text{diag}\{\exp(1+k\Delta/t)\lambda_i\}H^{-1}$.

Example 3:

The following example illustrates the advantages of using the approximation. Consider, with t = 1,

$$\hat{P} = \begin{pmatrix} 0.6 & 0.2 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.1 & 0.1 \\ 0.1 & 0.05 & 0.8 & 0.05 \\ 0.02 & 0.04 & 0.04 & 0.9 \end{pmatrix}$$

which can be represented as e^{Q^*} for

$$Q_o = Q^* = \begin{pmatrix} -0.5469 & 0.3080 & 0.1235 & 0.1154 \\ 0.1462 & -0.3851 & 0.1235 & 0.1154 \\ 0.1416 & 0.0459 & -0.2360 & 0.0485 \\ 0.0206 & 0.0464 & 0.0432 & -0.1103 \end{pmatrix}$$

Table 1 gives matrices $e^{(1+0.1k)Q_o}$ for $k\in\{-2,-1,-0.5,0.5,1,2\}$ (the "exact" column) and the perturbation approximation, using

$$G_k = k\hat{P}Q_o = k \begin{pmatrix} -0.2827 & 0.1170 & 0.0795 & 0.0862 \\ 0.0639 & -0.2296 & 0.0795 & 0.0862 \\ 0.0669 & 0.0506 & -0.1681 & 0.0506 \\ 0.0191 & 0.0344 & 0.0369 & -0.0904 \end{pmatrix}$$

(the "approximation" column). As can be seen, there is good agreement.

4. Analysis

We now consider the statistical analysis of the effect of sampling/measurement error on the estimation of P using our additive perturbation model (2.1). We assume that $\hat{P}$ and ε are known and that G is a matrix of s^2 random quantities, subject to the constraints that $g_{i+}=0$. For simplicity, we treat the s rows of G as independent vectors. The parameter ε is constrained to insure that

$$0 \le p_{ij} + \varepsilon g_{ij} \le 1$$

for all s^2 elements of $\hat{P}(\varepsilon,G)$.

A convenient and plausible model for g_i, the i^{th} row of G, is the Dirichlet, with parameters $\beta_i = (\beta_{i1},\beta_{i2},\ldots,\beta_{is})^T$, and $K_i = \Sigma_j \beta_{ij}$. We let $g^*_{ij} = g_{ij} + 1/s$ so that $g^*_i = (g^*_{i1},\ldots,g^*_{is})^T$ is a Dirichlet random vector, with mass function

$$(4.1) \qquad f(g^*_{i1},\ldots,g^*_{is}) = \frac{\Gamma(K_i)}{\prod_{j=1}^{s}\Gamma(\beta_{ij})} \prod_{j=1}^{s} (g^*_{ij})^{\beta_{ij}-1}$$

Table 1

Comparison of $e^{(t+k\Delta)Q^*}$ with perturbation approximation.

	EXACT				APPROXIMATION			
k=-2	0.6604	0.1740	0.0834	0.0823	0.6565	0.1766	0.0841	0.0828
	0.0861	0.7483	0.0834	0.0823	0.0872	0.7459	0.0841	0.0828
	0.0855	0.0398	0.8348	0.0399	0.0866	0.0399	0.8336	0.0399
	0.0161	0.0329	0.0325	0.9184	0.0162	0.0331	0.0326	0.9181
k = -1	0.6292	0.1876	0.0919	0.0913	0.6283	0.1883	0.0920	0.0914
	0.0933	0.7235	0.0919	0.0913	0.0936	0.7229	0.0920	0.0914
	0.0930	0.0449	0.8171	0.0449	0.0933	0.0449	0.8168	0.0449
	0.0181	0.0365	0.0363	0.9091	0.0181	0.0366	0.0363	0.9090
k = -½	0.6144	0.1940	0.0960	0.0957	0.6141	0.1941	0.0960	0.0957
	0.0967	0.7116	0.0960	0.0957	0.0968	0.7115	0.0960	0.0957
	0.0966	0.0475	0.8085	0.0475	0.0966	0.0475	0.8084	0.0474
	0.0190	0.0383	0.0382	0.9045	0.0190	0.0383	0.0382	0.9045
k = ½	0.5861	0.2057	0.1039	0.1043	0.5859	0.2058	0.1040	0.1043
	0.1031	0.6887	0.1039	0.1043	0.1032	0.6885	0.1040	0.1043
	0.1033	0.0525	0.7917	0.0525	0.1033	0.0525	0.7916	0.0525
	0.0210	0.0417	0.0418	0.8955	0.0210	0.0417	0.0419	0.8955
k = 1	0.5726	0.2111	0.1078	0.1085	0.5717	0.2117	0.1080	0.1086
	0.1061	0.6776	0.1078	0.1085	0.1064	0.6770	0.1080	0.1086
	0.1064	0.0550	0.7835	0.0551	0.1067	0.0550	0.7832	0.0551
	0.0219	0.0434	0.0437	0.8910	0.0219	0.0434	0.0437	0.8910
k = 2	0.5470	0.2210	0.1152	0.1168	0.5435	0.2234	0.1159	0.1172
	0.1117	0.6563	0.1152	0.1168	0.1128	0.6541	0.1159	0.1172
	0.1124	0.0600	0.7675	0.0601	0.1134	0.0601	0.7664	0.0601
	0.0238	0.0467	0.0473	0.8823	0.0238	0.0469	0.0474	0.8819

where $0 \leq g^*_{ij} \leq 1$ and $g^*_{i+} = 1$, so that $-1/s \leq g_{ij} \leq 1-1/s$ and $g_{i+} = 0$. We note that with (4.1),

$$E\{g_{ij}\} = K_i^{-1}(\beta_{ij}-K_i/s) \tag{4.2}$$

$$\text{Var}\{g_{ij}\} = [K_i^2(K_i+1)]^{-1}\ \beta_{ij}(K_i-\beta_{ij}) \tag{4.3}$$

and

$$\text{Cov}\{g_{ij},g_{i'j'}\} = \begin{cases} 0 & \text{if } i\neq i' \\ -\dfrac{\beta_{ij}\beta_{ij'}}{K_i^2(K_i+1)} & \text{if } i = i',\ j\neq j'. \end{cases} \tag{4.4}$$

We define Σ_i as covariance matrix of g_i. It will be convenient to let $E\{\text{terms of order } \varepsilon^2\} = 0$.

We first consider the moments of the random vector $\lambda(\varepsilon) = (1,\lambda_2(\varepsilon),\ldots,\lambda_s(\varepsilon))^T$. Since $\lambda(\varepsilon) \in \mathbb{C}^s$, we express $\lambda(\varepsilon)$ as $\alpha(\varepsilon) + i\beta(\varepsilon)$, where $\alpha(\varepsilon)$ and $\beta(\varepsilon)$ are real s-dimensional vectors, and define

$$E\{\lambda(\varepsilon)\} = u_\lambda(\varepsilon) = E\{\alpha(\varepsilon)\} + i\ E\{\beta(\varepsilon)\}$$

$$\text{Cov}\{\lambda_i(\varepsilon),\lambda_j(\varepsilon)\} = \sigma_{ij}(\varepsilon) = E\{(\lambda_i(\varepsilon)-u_{\lambda_i}(\varepsilon))\ \overline{(\lambda_j(\varepsilon)-u_{\lambda_j}(\varepsilon))}\ \}$$

where $\lambda_i(\varepsilon)$ and $\lambda_j(\varepsilon)$ are elements of $\lambda(\varepsilon)$. Note that $\sigma_{ij}(\varepsilon) \neq \sigma_{ji}(\varepsilon)$ (they are complex conjugates, however) so that in general, $\Sigma(\varepsilon) = (\sigma_{ij}(\varepsilon))$ is an sxs Hermitian matrix (see Goodman, 1963).

Theorem 1

$$E\{\lambda_i(\varepsilon)\} = \lambda_i + \varepsilon \sum_k \sum_\ell h^*_{ik}\ \beta^*_{k\ell}\ h_{\ell i}$$

$$= \lambda_i + \varepsilon\ h^*_i\ \underset{\sim}{\beta}^*\ h_i \qquad (i \neq 1) \tag{4.5}$$

where h^*_{ik} is the k^{th} element of the i^{th} left eigenvector, and $h_{\ell i}$, the ℓ^{th} element of the i^{th} right eigenvector, corresponding to λ_i, and $\underset{\sim}{\beta}^* = (\beta^*_{k\ell}) = (\beta_{k\ell}/K_k)$.

The proof of the theorem follows directly from (4.2) and the fact that since the first right eigenvector $h_1 \propto \underset{\sim}{1}$, Σh^*_{ik} must be zero (for $i \neq 1$), so that $\sum_k \sum_\ell h^*_{ik} h_{\ell i} = 0$, $i \neq 1$.

Thus the "bias" of $\lambda_i(\varepsilon)$ is $\varepsilon\, h_i^* \underset{\sim}{\beta}^* h_i$. This leads to

Corollary 1:

If $\beta_{k\ell} \equiv \beta_k$ or $\beta_{k\ell} \equiv \beta_\ell$ (same parameters for each g_i, or a constant parameter for every g_i), then the bias of $\lambda_i(\varepsilon)$ is zero.

Consequently, if errors are equally likely to occur in the s cells of g_i, or if the average error for the j^{th} cell is the same for every g_i, there is no bias.

Corollary 2:

If $\underset{\sim}{\beta}^* = \alpha \underset{\sim}{I} + \beta \underset{\sim}{J}$, where J is a sxs matrix of ones, then $h_i^* \underset{\sim}{\beta}^* h_i = \alpha (i \neq 1)$ and $E\{\lambda_i(\varepsilon)\} = \lambda_i + \varepsilon\alpha$.

The proof is based on the fact that $h_i^* J h_i = 0$ $(i \neq 1)$. Thus, if $E\{g_{ii}\} - E\{g_{ij}\}$ is constant, the bias is constant for all roots. This situation is likely to occur with the addition of stayers as sampling error.

Theorem 2:

Considering $\ddagger$, the variance-covariance matrix of $\lambda(\varepsilon)$, note that $\sigma_{1j}(\varepsilon) = \sigma_{j1}(\varepsilon) = 0$, for all j, since $\lambda_1 = \lambda_1(\varepsilon) \equiv 1$. The remaining $(s-1)^2$ elements are

$$\sigma_{ii}(\varepsilon) = \varepsilon^2 \sum_{k=1}^{s} h^*_{ik}\, \overline{h^*_{ik}}\, h_i^T \Sigma_k\, \overline{h_i}$$

(4.6)

$$\sigma_{ij}(\varepsilon) = \varepsilon^2 \sum_{k=1}^{s} h^*_{ik}\, \overline{h^*_{jk}}\, h_i^T \Sigma_k\, \overline{h_j} \; .$$

As before, the "bar" indicates complex conjugate, and Σ_k is defined as the variance-covariance matrix of g_k, with elements given by (4.3) and (4.4).

Proof

First consider $\sigma_{ii}(\varepsilon) = \text{Var}\,(\varepsilon h_i^* G h_i)$. Since the rows of G are independent,

$$\text{Var}(\varepsilon h_i^* G h_i) = \varepsilon^2 \sum_{k=1}^{s} h_{ik}^* \overline{h_{ik}^*}\, \text{Var}(G h_i)$$

$$= \varepsilon^2 \sum_{k=1}^{s} h_{ik}^* \overline{h_{ik}^*}\, h_i^T \Sigma_k \overline{h_i}$$

The expression for $\sigma_{ij}(\varepsilon)$ follows, using similar reasoning

To study the affect of varying the Dirichlet parameters on Σ, we can write

$$h_i^T \Sigma_k h_j = \sum_{\ell=1}^{s} h_{\ell i}\, \overline{h_{\ell j}} [K_k^2(K_k+1)]^{-1} \beta_{k\ell}(K_k - \beta_{k\ell})$$

$$- \sum_{\substack{\ell \\ \ell \neq m}} \sum_{m} h_{\ell i}\, \overline{h_{mj}}\, [K_k^2(K_k-1)] \beta_{k\ell}\beta_{km}$$

Corollary 3:

If $\beta_{k\ell} \equiv \beta_k$, then $\Sigma_k = s\alpha_k \underset{\sim}{I} - \alpha_k \underset{\sim}{J}$ where $\alpha_k = [s^2(s\beta_k+1)]^{-1}$. Unfortunately, the expressions for $\sigma_{ij}(\varepsilon)$ do not simplify.

The distribution of $\lambda_i(\varepsilon)$ is quite complicated. The standard result on symmetric matrix roots (Muirhead, 1978) do not apply. If we just consider the quantities $h_i^* G h_i = \sum_{j=1}^{s} h_{ij}^* (g_j^T h_i)$, $g_j + 1/s \sim$ Dirichlet (β), for real λ_i, we see that we have a linear combination of translated Dirichlet random variables, which unfortunately, does not follow a standard distribution, with closed form density function. Consequently, we simply construct concentration ellipsoids

$$(4.7) \quad \{\lambda^* : (\lambda^* - u_\lambda(\varepsilon))^T \overline{\Sigma^-} (\overline{\lambda^* - u_\lambda(\varepsilon)}) \leq c\}$$

for some $c \geq 0$, where Σ^- is a generalized inverse of Σ, in order to study the dispersion of $\lambda(\varepsilon)$ in more detail. We are not able to assign a probability content to the ellipsoids, however.

Example 4

Consider the array, given by Singer and Spilerman (1976a),

$$\hat{P} = \begin{pmatrix} 0.600 & 0.330 & 0.070 \\ 0.302 & 0.560 & 0.138 \\ 0.380 & 0.040 & 0.580 \end{pmatrix}$$

which has eigenvalues $\underset{\sim}{\lambda} = (1, 0.370+0.01095i, 0.370-0.01095i)^T$ and eigenvectors, the columns of

$$H = \begin{pmatrix} -0.6998 & 11.176-1.337i & 11.176+1.337i \\ -0.6998 & -3.587+1.049i & -3.587-1.049i \\ -0.6998 & -19.602+1.1977i & -19.602-1.1977i \end{pmatrix} .$$

This array has logarithm

$$\hat{Q} = \begin{pmatrix} -0.692 & 0.639 & 0.053 \\ 0.496 & -0.733 & 0.257 \\ 0.707 & -0.144 & -0.563 \end{pmatrix}$$

so that it is not embeddable, since $\hat{q}_{23} < 0$. However, this $\hat{P}$ is within "error-distance" of the array

$$\tilde{P} = \begin{pmatrix} 0.598 & 0.334 & 0.068 \\ 0.298 & 0.568 & 0.134 \\ 0.349 & 0.104 & 0.547 \end{pmatrix}$$

which is embeddable:

$$\tilde{Q} = \log \tilde{P} = \begin{pmatrix} -0.692 & 0.639 & 0.053 \\ 0.496 & -0.733 & 0.237 \\ 0.635 & 0 & -0.635 \end{pmatrix}$$

In fact, $\tilde{P}$ is the "closest" array, in the sense of minimizing $||\hat{P}-\tilde{P}||^2 = \sum_{ij} (\hat{p}_{ij}-\tilde{p}_{ij})^2$. To characterize other arrays that are as close to $\hat{P}$ as $\tilde{P}$, we consider $\tilde{P}$ as a perturbation of $\hat{P}$:

$$\tilde{P} = \hat{P} + EG + O(||E||^2)$$

where $E = \text{diag}(\varepsilon_i)$. To find a suitable $\underset{\sim}{\beta}$, we match $E\{\varepsilon_i g_{ij}\}$ to the difference $\tilde{p}_{ij} - \hat{p}_{ij}$. If we choose $\varepsilon_1 = 0.01$, $\varepsilon_2 = 0.05$, $\varepsilon_3 = 0.40$, then we obtain

$$\underset{\sim}{\beta}^* = \begin{pmatrix} 1/7 & 5/7 & 1/7 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/2 & 1/4 \end{pmatrix}$$

and

$$E\{\tilde{P}-\hat{P}\} = \begin{pmatrix} -0.0019 & 0.0038 & -0.0019 \\ -0.0042 & 0.0084 & -0.0084 \\ -0.033 & 0.066 & -0.033 \end{pmatrix} \simeq \tilde{P}-\hat{P}.$$

Since ε now depends on the row of $\hat{P}$, equations (4.5) and (4.6) must be altered slightly.

We find that $u_\lambda(E) = (1, 0.3566+0.01564i, 0.3566-0.01564i)^T$ resulting in a bias for λ_2 or λ_3 of $-0.0134 \pm 0.00469i$. We note that $u_{\lambda_2}(E)$ is quite close to the second root of $\tilde{P}$, $\tilde{\lambda}_2 = 0.3565+0.01318i$, so that our "mean-matching" has been successful.

Further calculations using this $\underset{\sim}{\beta}^*$ matrix yield

$$\ddagger = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0.1761 & -0.1730-0.0259i \\ 0 & -0.1730+0.0259i & 0.1761 \end{pmatrix}$$

and we obtain the concentration ellipse (in the second and third dimension of $\underset{\sim}{\lambda}(E)$)

$$\{(\alpha,\beta): 1767.48[\alpha-0.3566]^2 - 262.72[\alpha-0.3566][\beta-0.01564] + 15.480[\beta-0.01564]^2 \leq c\}$$

where $\lambda = \alpha \pm i\beta$ is a possible second or third root. This ellipse, with $c = 1$, is given in Figure 1. Note how thin it is, indicating a very narrow range for α, but a large range for the imaginary part β. Note that the real

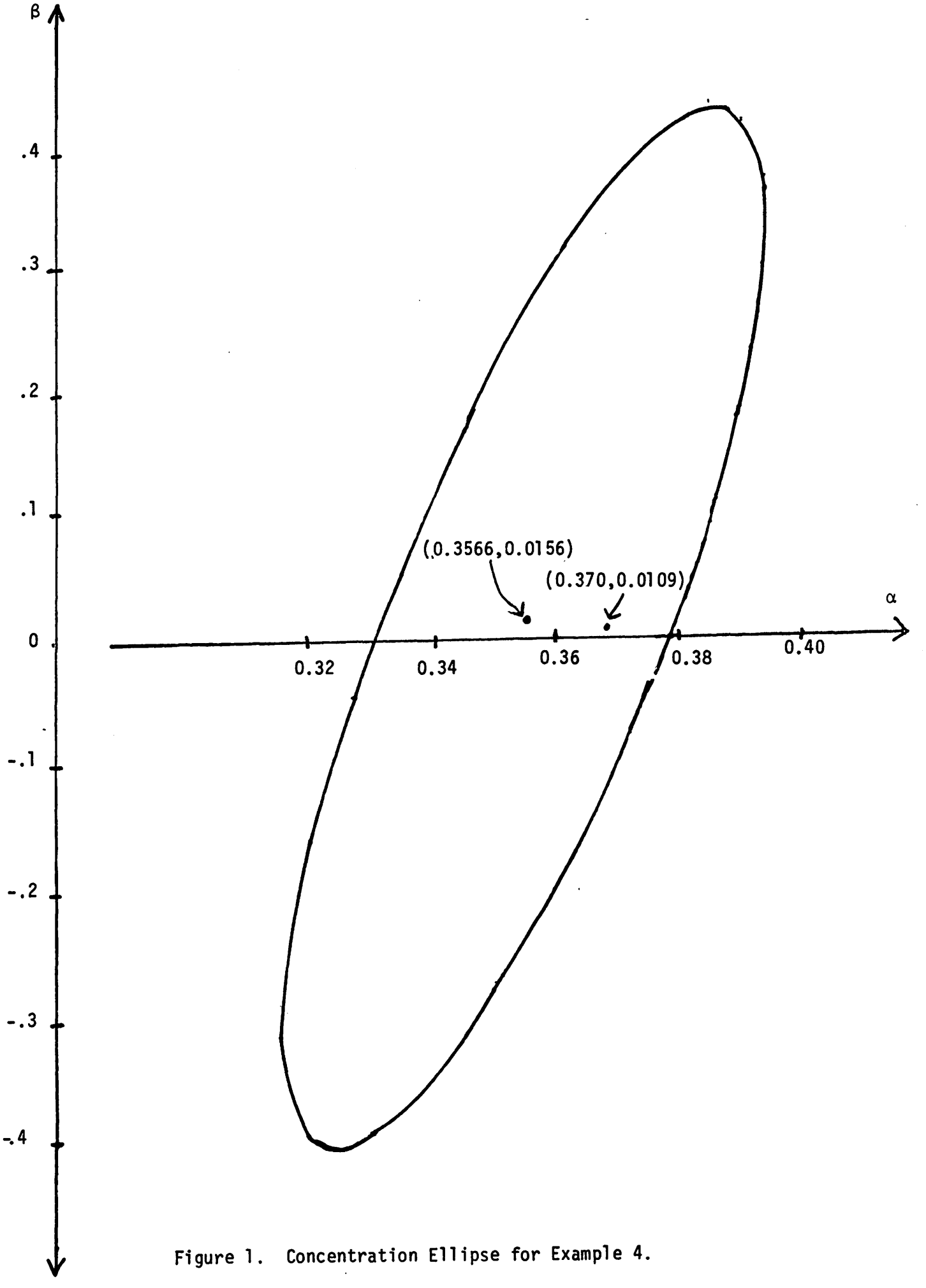

Figure 1. Concentration Ellipse for Example 4.

numbers 0.3318 and 0.3785 lie on the perimeter so that any root in the real interval [0.3318,0.3785] is possible. If so, we would have 2 equal real eigenvalues. Apparently, there is quite a range for transition arrays that are of distance $||\hat{P}-\tilde{P}||^2$ from $\hat{P}$.

Example 5

Reconsider the array studied in example 3, which has eigenvalues $\lambda = (1,0.5,0.675,0.824)^T$. If we stay within the space R^4, and let $\varepsilon = 1$ and $\beta_{ij} = 1$ for all (i,j), then $u_\lambda(1) = (1,0.5,0.675,0.825)^T$, and

$$\Sigma = \begin{pmatrix} 0 & 0 & 0 & 0 \\ & 0.06275 & -0.0128 & 0.0000 \\ & & 0.06425 & -0.0015 \\ & & & 0.0514 \end{pmatrix} .$$

The concentration ellipsoid in R^3 (ignoring the non-varying first dimension) has an orientation given by the last 3 eigenvectors of Σ

$$e_2 = (0.684,-0.728,0.0437)^T$$
$$e_3 = (0.446,0.370,-0.815)^T$$
$$e_4 = (0.577,0.577,0.578)^T$$

which are represented in Figure 2. The associated eigenvalues are 0.0764, 0.0520, and 0.05, indicating that e_2 is relatively more important than e_3 and e_4, which are roughly of equal importance.

Cross-sectional views of a unity (c=1) concentration ellipsoid taken with $\lambda_4 = 0.824,0.90,0.95$, and 0.99, are shown in Figure 3 as projections onto the (λ_2,λ_3) plane. With c=1, λ_4 varies from 0.597 to 1.00, and the constancy of the e_4 vector indicates that the ellipsoid is parallel to the λ_4 axis.

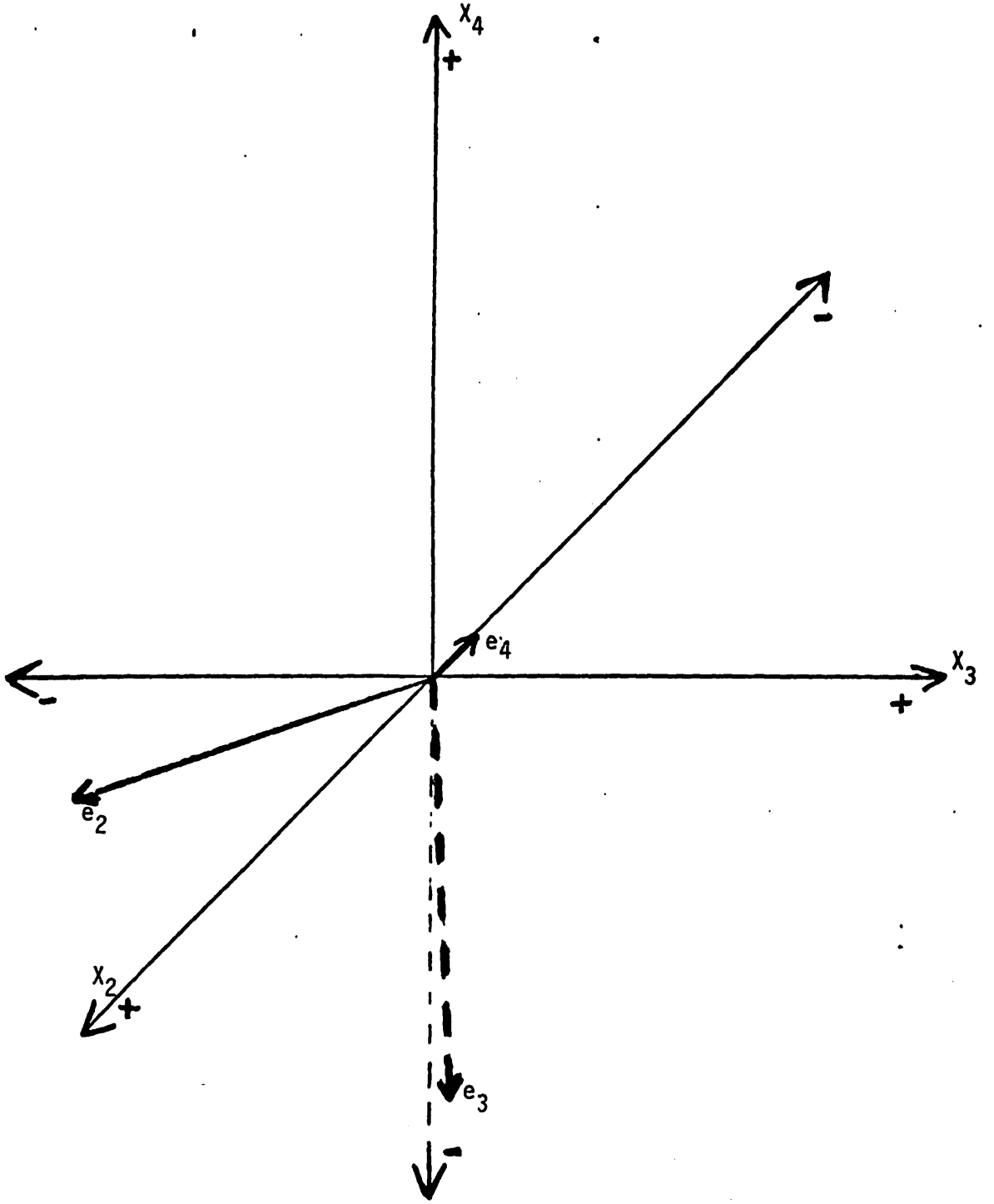

Figure 2. Axes of concentration ellipsoid for Example 5.

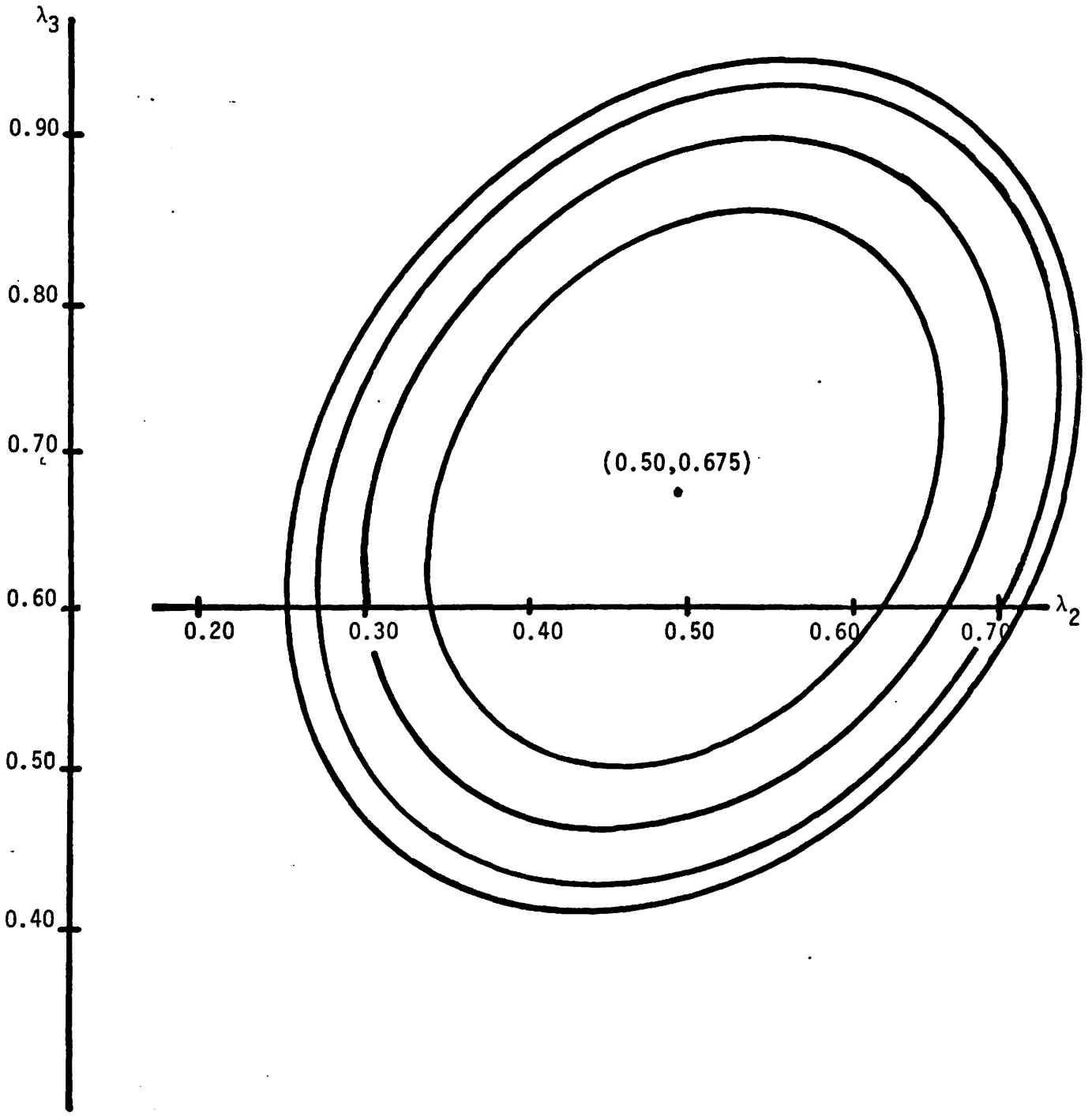

Figure 3. Concentration ellipsoid for Example 5, with c=1. Cross-sectional slices, from center outward, for λ_4 = 0.99,0.95,0.90,0.824.

5. Discussion: Embeddability and Identification

Our purpose here has been to investigate the sensitivity of solutions to (1.1) to perturbations or slight changes in P(t). To best study this, we should find an expression for $\log \hat{P}(\varepsilon,G)$ using the spectral decomposition (2.4). We have

$$\log \hat{P}(\varepsilon,G) = \sum_{i=1}^{S} \log(\lambda_i+\varepsilon h_i^* G h_i) h_i h_i^* + \varepsilon \sum_{i=1}^{S} \sum_{\substack{j=2 \\ i\neq j}}^{S} (h_i^* G h_j) h_i h_j^* + O(\varepsilon^2).$$

Using the approximation $\log(1+\delta) = \delta - \frac{1}{2}\delta^2 + O(\delta^3)$, we can expand

$$\log(\lambda_i+\varepsilon h_i^* G h_i) = \log \lambda_i + \log(1+\lambda_i^{-1}\varepsilon h_i^* G h_i)$$

$$= \log \lambda_i + \lambda_i^{-1}\varepsilon h_i^* G h_i + O(\varepsilon^2)$$

so that

$$\log \hat{P}(\varepsilon,G) \approx \sum_{i=2}^{S} (\log \lambda_i) h_i h_i^* + \varepsilon \sum_{i=2}^{S} \lambda_i^{-1} h_i^* G h_i h_i h_i^* + \varepsilon \sum_{i=1}^{S} \sum_{\substack{j=2 \\ i\neq j}}^{S} h_i^* G h_j h_i h_j^* . \tag{5.1}$$

Table 2 gives a brief view of the accuracy of the approximation (5.1), which apparently can be quite good when the ratio of λ to $\varepsilon h^* G h$ is large.

We noted in section 3 that with the Hadamard multiplication error model, $\log \hat{P}(E) \neq \log \hat{P} + \log E$. However, only the eigenvectors of $\hat{P}$ are needed to calculate $\hat{P}(E)$ so that

$$\log \hat{P}(E) \approx \sum_{i=2}^{S} (\log \lambda_i) h_i h_i^* + \{- \sum_{i=2}^{S} h_i h_i^* + \sum_{i=1}^{S} \sum_{j=2}^{S} [(\frac{1}{\lambda_i} - 1)\delta_{ij}+1] h_i^* \hat{P} * E h_j h_i h_j^*\}. \tag{5.2}$$

The term in brackets gives the effect of the addition of stayers on $\hat{Q}$.

Throughout the previous sections, we have assumed that $\hat{P}(t)$ had distinct eigenvalues, so that if a specific $\hat{P}(t)$ is embeddable (i.e., set of all solutions to (1.1) $\subset \mathcal{Q}$), there is at worst, a finite number of solutions, and at best, an unique $\hat{Q}$. Given the sets of roots of $\hat{P}$ and its perturbations, one

Table 2

Accuracy of logarithmic eigenvalue approximation

λ_i	.8	.8	.5	.5	.2	.2	.05	.05
$\epsilon h_i^* G h_i$	.05	.01	.05	.01	.05	.01	.05	.01
$\log(\lambda_i+\epsilon h_i^* G h_i)$	-.1625	-.2107	-.5978	-.6733	-1.386	-1.561	-2.303	-2.813
$\log \lambda_i^{-1}\epsilon h_i^* G h_i$	-.1606	-.2106	-.5931	-.6731	-1.359	-1.559	-1.996	-2.796

can identify all possible solutions. The relationship between the roots and the identification problem is nicely summarized in Table 3, reproduced from Singer and Spilerman (1976a). We hope that the approximations given here (especially 5.1 and 5.2) can be used in conjugation with the necessary conditions for embeddability (in particular those of Runnenberg given in Singer and Spilerman, 1976, page 11) so that further insights into the identification problem are gained.

Table 3

Eigenvalues of $P(t)$ and the number of matrices $Q \in \mathcal{Q}$

Eigenvalue Characteristics	Embeddable?	How Many Q's?
1. Positive, distinct	Possibly	One
2. Positive, repeated, distinct elementary divisors	Possibly	One
3. Positive, repeated, nondistinct elementary divisors	Possibly	One or continuum
4. Negative, distinct	Never	...
5. Negative, repeated, odd multiplicity ...	Never	...
6. Negative, repeated, even multiplicity ..	Possibly	Continuum
7. Complex, distinct, member of a conjugate pair	Possibly	One or multiple
8. Complex conjugate, repeated	Possibly	One, multiple, or continuum
9. Mixture of the types above	Possibly	The most extreme form of nonuniqueness present in any component of the mixture

6. References

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

Blumen, I., M. Kogan, and P.J. McCarthy (1955). The Industrial Mobility of Labor as a Probability Process. Ithaca, N.Y.: Cornell University Press.

Frydman, H. (1980). "The embedding problem for Markov chains with three states," Math. Proceedings of the Cambridge Philosophical Soc., 87:285-294.

Frydman, H. and B. Singer (1979). "Total positivity and the embedding problem for Markov chains," Math. Proceedings of the Cambridge Philosophical Soc., 86:339-344.

Goodman, G.S. (1970). "An intrinsic time for non-stationary finite Markov chains," Z. Wahrscheinlich keitstheorie und Verw. Gebiete, 16:165-180.

Goodman, N.R. (1963). "Statistical analysis based on a certain multivariate complex Gaussian distribution," Annals of Mathematical Statistics, 34:152-177.

Johansen, S. (1973a). "A central limit theorem for finite semigroups and its application to the imbedding problem for finite state Markov chains," Z. Wahrscheinlich keitstheorie und Verw. Gebiete, 26:171-190.

Johansen, S. (1973b). "The Bang-Bang problem for stochastic matrices," Z. Wahrscheinlich eitstheorie und Verw. Gebiete, 26:191-195.

Johansen, S. and F. L. Ramsey (1979). "A Bang-Bang representation for 3x3 imbeddable stochastic matrices," Z. Wahrscheinlich keitstheorie und Verw. Gebiete, 47:107-118.

Kingman, J.F.C. (1962). "The imbedding problem for finite Markov chains," Z. Wahrscheinlich keitstheorie und Verw. Gebiete, 1:14-24.

Muirhead, R.J. (1978). "Latent roots and matrix variates: a review of some asymptotic results," Annals of Statistics, 6:5-33.

Singer, B. and J.E. Cohen (1980). "Estimating malaria incidence and recovery rates from panel surveys," Mathematical Biosciences, 49:273-305.

Singer, B. and S. Spilerman (1974). "Social mobility models for heterogeneous populations," Sociological Methodology 1973-1974, edited by H. L. Costner, San Francisco: Jossey-Bass, 356-401.

Singer, B. and S. Spilerman (1976a). "The representation of social processes by Markov models," American Journal of Sociology, 82:1-54.

Singer B. and S. Spilerman (1976b). "Some methodological issues in the analysis of logitudinal surveys," Annals of Economic and Social Measurement, 5:447-474.

Singer, B. and S. Spilerman (1977). "Trace inequalities for mixtures of Markov chains," Advances in Applied Probability, 9:747-764.

Singer, B. and S. Spilerman (1978). "Clustering on the main diagonal in mobility matrices," Sociological Methodology 1979, edited by K. F. Schuessler, San Francisco: Jossey-Bass, 172-208.

Stewart, G.W. (1973). Introduction to Matrix Computations. New York: Academic Press.

Wilkinson, J.H. (1965). The Algebraic Eigenvalue Problem, London: Oxford University Press.