NUMBER 5

A CAUSATIVE MATRIX APPROACH TO MOBILITY
  STUDIES

Barry Bye
Division of Disability Studies
Office of Research and Statistics
Social Security Administration

John Hennessey
Department of Mathematics
Loyola College
Baltimore, Maryland

A Causative Matrix Approach to Mobility Studies

by Barry Bye and John Hennessey

I. Introduction

Markov models have been widely used for the analysis and prediction

of shifts in population distribution over time. The point of depar-

ture for most of these analyses has been the finite state, time

stationary Markov chain. The usual Markov chain model has, however,

been shown to be inadequate for most social science applications.

The principle failure of this model is exhibited in the notion of

empirical regularity (see for example Coleman [3], Singer and

Spilerman [16]).

Proposed alternatives to the stationary Markov chain models have

taken various directions depending on the presumed reason for the

empirical regularity. Coleman [3], Lazarsfeld and Henry [9], and

Wiggins [18] have considered the effects of response variability as

the source of the empirical regularity. Population heterogeneity

as a source has been discussed starting with Blumen, Kogan and

McCarthy [1] and furthered by Spilerman [17] and Singer and Spilerman

[14]. A recent publication by Singer and Spilerman [15] has focused

on continuous time stationary models as a possible alternative to

the discrete time Markov chain.

This paper presents a particular kind of discrete time nonstationary

Markov chain. Such chains will be built using a mathematical quantity

called a causative matrix. Causative matrices have been found useful

in the analysis and prediction of certain consumer purchasing

processes [10, 11] and in the analysis of sequences of stochastic

matrices derived from cohort data [4].

Section II of the paper defines the notions of a causative and strongly

causative matrix and presents the constant causative matrix model.

Section III relates the causative matrix model to an empirical

regularity which occurs in many social science data. Section IV

discusses a possible variation of the model. Section V presents

directions for future research.

II. The Causative Matrix

To be best of our knowledge, the notion of a causative matrix

originated with Benjamin Lipstein [10]. Lipstein was interested in

the consumer purchasing process and the ability to detect stability

in the market place after the introduction of a new brand or after

a large advertising campaign. Given a sequence of stochastic matrices

R_1, R_2, ..., R_k ... with states representing brand preference, Lipstein

defined the ith causative matrix,

$$C_i = R_i^{-1} R_{i+1} \quad \text{when } R_i^{-1} \text{ exists.} \qquad \text{(See [11])}$$

Thus, causative matrices are a way of accounting for observed

nonstationarity in the transition probabilities. These matrices

can then be used for prediction purposes.

Following Lipstein's lead, we have chosen to define causative matrices

with respect to nonsingular stochastic matrices as follows:

Definition 1:   Let C be an nxn matrix. C is causative if there

exists a nonsingular, stochastic matrix R such

that RC is stochastic.

It can be shown that the rows of a causative matrix, C, must sum to

one; and, therefore, 1 is an eigenvalue of C. However, the entries

of C need not be greater than or equal to zero, so, although all

stochastic matrices are causative, there are many nonstochastic

matrices which are also causative.

When viewed as a linear transformation of E_n, n dimensional euclidean

space, a causative matrix has specific algebraic properties. The

hyperplanes $x_1 + x_2 + \ldots x_n = k$ of E_n are invariant under C and all

eigenvectors of C corresponding to non-unit eigenvalues must be in

the plane $x_1 + x_2 + \ldots + x_n = 0$. These and other properties concerning

convergence of $\lim_{t \to \infty} C^t$ have been presented by Hennessey and Bye [6].

Necessary and sufficient conditions for a matrix to be causative,

in terms of its entries, have been given by Hennessey and Bye [7].

They can be stated as follows:

Theorem 2: A matrix C is causative if and only if there exists a

positive vector, r, such that rC is positive in all entries except

those corresponding to zero columns of C.

There are ways to test for the existence of such an r vector in

terms of the entries of C alone (see [7]).

It is interesting to consider the following situation: Given a

nonsingular, stochastic matrix R and a matrix C which is causative

with respect to R, consider the sequence $R, RC, RC^2, \ldots, RC^t \ldots$. If

these matrices are all stochastic, one can consider the nonstationary

Markov chain so generated. Thus, it is interesting to consider those

causative matrices for which such a chain can be constructed, and

we have the following definition.

Definition 3: Let C be a causative matrix, then C is <u>strongly</u>

<u>causative</u> if there exists a nonsingular stochastic matrix R such

that RC^t is stochastic for all t.

A full statement of necessary and sufficient conditions (given in [7])

is beyond the scope of this paper. However, to give some feeling for

the nature of the restrictions, we state two of the necessary conditions:

Theorem 4: If C is strongly causative then;

    1. All eigenvalues of C are less than or equal to one

       in absolute value, and

    2. All eigenvalues equal to one in absolute value are

       not linked in the Jordan form of C.

These conditions insure that no entry of C^t and thus of RC^t becomes

indefinitely large in absolute value (which would make entries of

RC^t greater than 1 or less than 0 and thus makes RC^t nonstochastic).

Given a matrix C which is strongly causative it is shown in [7] that

one can construct nonsingular, stochastic matrices, R, such that C

is strongly causative with respect to R. The question then naturally

arises: Given an R and a C; is C strongly causative with respect

to this particular R? This problem is, at present, unsolved and is

the subject of current research.

Given an R and a C which is strongly causative with respect to R,

we consider the nonstationary chain: $\{R, RC, RC^2, \ldots\}$.

We then construct the t step transition matrix, $T_t = P(o,t)$, where

$$T_t = RRCRC^2 \ldots RC^{t-1} = \prod_{k=0}^{t-1} RC^k.$$

Then, given an initial state vector, s, we consider the convergence

properties of $\{sT_t\}$ as $t \to \infty$.

Necessary and sufficient conditions for the convergence of T_t for

two state chains have been given by Harary, Lipstein, and Styan [5].

Their results, however, could not be generalized to chains with

more than two states. Pullman and Styan [13] and Bye and Hennessey [2]

have presented some partial results for n state chains. Necessary

and sufficient conditions for the convergence of T_t are shown in [2]

in the special case where R and C have the same canonical basis and

structure of the Jordan form.

Both papers contain the following result:

Theorem 5: Let C be a convergent strongly causative matrix with respect to R. Let C also be such that there is only one eigenvalue of C equal to 1 and let v be its eigenvector normalized to lie in the plane $x_1 + x_2 + \ldots + x_n = 1$, then $\lim_{t \to \infty} T_t = V$ where V is a matrix with all rows equal to v.

In this case, the process loses memory with all initial distribution vectors converging to v.

IV. An Emperical Regularity

Many authors have noted that social mobility data often yields the following kind of empirical regularity. Given an observed one step transition matrix $P = P(0,1)$ and an observed k step transition matrix $P(0,k)$, the diagonal entries of $P(0,k)$ have the property that they are greater than the diagonal entries of P^k. This means that the probability of being in the same state after k time periods is larger than that predicted by the stationary chain with $P(0,1)$ as its transition matrix. To translate this property into more tractable (though not quite equivalent) form, we define

$\qquad$ trace $P(0,k) \geq$ trace P^k to be the empirical regularily property.

We will be interested in developing certain properties about the trace of k-step transition matrices generated by certain kinds of causative matrices. In order to do this without too much complexity, we will place certain mild restrictions on the stochastic R's and causative C's which we have been discussing in previous sections.

Following the work done by McKenzie [12] we will be interested in stochastic matrices R with "dominant diagonals."

Definition 7: An stochastic matrix $R = (r_{ij})$ has a dominant diagonal (d.d.) if $r_{ii} \geq \sum_{j \neq i} r_{ij}$ for all i.

Such matrices are frequently encountered in social mobility data. They have the following property: (see [12] for details).

Theorem 8: If a stochastic matrix R has a d.d. then

    1. R is non-singular

    2. All eigenvalues of R have positive real parts.

In the following discussion we will restrict ourselves to stochastic matrices, R, which have a d.d. For simplicity, we will also assume that R has a full set of eigenvectors.

First, it should be noted that if R has a d.d. and R and C satisfy the hypothesis of Theorem 5, then

$$tr(\lim_{t \to \infty} R^t) = tr(\lim_{t \to \infty} T_t) = 1.$$
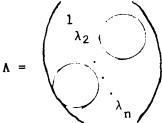
Also, all strongly causative C's, which we have observed empirically have generated chains where trace T_t < trace R^t and, therefore, do not have the empirical regularity property.

However, there is a class of causative (but not strongly causative) matrices which do have interesting properties with respect to empirical regularity. Consider the following situation:

Let R be a stochastic matrix with d.d. where

(9) $R = V^{-1} \Lambda V$

Where V contains the eigenvectors of R in the rows and

$$\Lambda = \begin{pmatrix} 1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

where the eigenvalues λ_i are real and $0 < \lambda_i < 1$ for all i.

Let C be causative with respect to R and

$$C = V^{-1} \Gamma V$$

(10) where $\Gamma = \begin{pmatrix} 1 & & & \\ & \gamma_2 & & \\ & & \ddots & \\ & & & \gamma_n \end{pmatrix}$ γ_i is real and

$\gamma_i > 1$ for all i

(That is, C and R have the same eigenvectors).

Then, we have the following theorem.

<u>Theorem 11</u>. Given an R and C with properties (9) and (10) respectively, the trace of $T_t = RRC...RC^{t-1}$ is greater then trace R^t for all t.

Proof: $T_t = RRC...RC^{t-1} = R^t C^{(t/2)(t-1)} = V^{-1} \Lambda^t \Gamma^{t/2(t-1)} V$.

Therefore, $\text{tr}(T_t - R^t) = \text{tr } \Lambda^t \left((\Gamma^{t/2(t-1)} - I) \right) > 0$ for all t.

QED

Thus, in particular, for all t for which RC^t is stochastic, T_t is stochastic and exhibits the empirical regularity mentioned above.

Although the restriction of C to have the same eigenvectors of R appears to be severe, we have had some success in fitting such C's to observed data. The data in table 1, taken from [8, table 1] show that the one step matrices can be fitted fairly well even with the eigenvectors for C fixed to be the same as those for R_4. The fitting procedure chooses the four non-unit eigenvalues for C which minimize the sums of squared deviations from R_5, subject to the constraints that each estimated eigenvalue be greater than 1.

Thus, it seems reasonable that one can fit C to insure that T_t has this empirical regularity and we have a model that warrants consideration for social mobility processes.

IV. A Variation of the Causative Matrix Model

When considering causative matrices with properties specified in

(10), we note that, by Theorem 4, such causative matrices cannot

be strongly causative. In this case it is useful to consider the

matrix $R^{**} = RC^{t_o}$ where t_o is chosen such that one or more entries

of RC^t first become negative t_o is not necessarily integral). R**

is a limiting value of RC^t in the sense that the process is moving

toward R** as a boundary value. (In table 1 the estimated R**

seems more realistic than the actual R**.)

We also note that such causative matrices will move the rows of

R faster and faster as t gets larger.

A variation of the constant causative matrix model is to consider

a sequence of causative matrices $C_1, \ldots, C_k, \ldots$ with the properties:

1. V is the constant eigenvector matrix for all C_i.

2. The nonunit eigenvalues for all C_i are real and greater
   than one so that

$$T_t = R^t \prod_{j=1}^{t-1} C_j^{t-j-1}$$ has the empirical regularity property.

3. $R, RC_1, RC_1C_2, \ldots \to R^*$, stochastic for some R where
   $$R = V^{-1} \Lambda V$$

We then have:

<u>Theorem 12</u>: $\lim_{t\to\infty} T_t = \lim_{t\to\infty} R^t$

<u>Proof</u>: $T_t = RRC_1RC_1C_2\ldots RC_1\ldots C_{t-1} = V^{-1}\Lambda^t \prod_{j=1}^{t-1} \Gamma_i^{t-j-1} V$

where Γ_i is the diagonal matrix of eigenvalues for C_i,

But $\Lambda^t \prod_{j=1}^{t-1} \Gamma_j^{t-j-1} \to \begin{pmatrix} 1 & 00\ldots 0 \\ 0 & \\ 0 & \\ 0 & \end{pmatrix}$ as $t\to\infty$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ QED

Thus, we have that the long run distributions for both stationary and nonstationary models are the same, with the nonstationary model taking the "high road" with respect to the empirical regularity mentioned in section III.

V. Future Research

The nonstationary chains we have discussed have the property that

$\lim_{t\to\infty} T_t \left(= \lim_{t\to\infty} P(0,t)\right)$ has trace equal to 1. An attempt to construct

models where $\mathrm{tr}\left(\lim_{t\to\infty} P(0,t)\right) > 1$, has led to the following formulation

for $P(0,t)$.

Suppose $P(0,1) = V^{-1}\Lambda V$ let

$P(0,t) = V^{-1}\begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda_1^t\alpha_1(t) & \\ 0 & & \lambda_2^t\alpha_2(t) \end{pmatrix} V$

where $\alpha_i(1) = 1$

$\quad \alpha_i(t) > 1 \quad$ for all i and t

and $\lambda_i^t \alpha_i(t) \to \alpha_i^*$ in such a way that $V^{-1} \begin{pmatrix} 1 & & & \\ & \alpha_1^* & & \\ & & \ddots & \\ & & & \alpha_n^* \end{pmatrix} V$ is stochastic.

P(0,t) then describes a t step transition matrix which guarantees

empirical regularity and in general, $\operatorname{tr} \lim_{t \to \infty} P(0,t) > 1$.

The relationship of this formulation of P(0,t) to certain classes

of mover/stayer and other heterogeneous models by specific para-

meterizations of $\alpha_i(t)$ is currently underway.

TABLE 1

	Actual					Estimated

Actual — Estimated

R4

Transition matrix (Actual):

```
.563  .058  .080  .161  .138
.082  .430  .109  .245  .134
.178  .084  .395  .222  .131
.085  .042  .030  .695  .148
.200  .122  .133  .155  .400
```

Estimated: Same as actual

Eigenvectors (in columns) (Actual):

```
.224  -.406  -.530   .219   .390
.102   .545  -.106   .300   .149
.108   .487  -.217   .472  -.876
.376  -.543   .812  -.223   .188
.190  -.835   .031  -.768   .149
```

Estimated: Same as actual

```
1.00   .394   .507   .271   .312
```

Estimated: Same as actual

R_5 (Actual) — **$\hat{R}_5$** (Estimated)

Transition matrix:

Actual:
```
.596  .053  .074  .147  .130
.078  .455  .105  .236  .176
.172  .072  .419  .215  .122
.066  .033  .024  .735  .142
.167  .093  .111  .129  .500
```

Estimated:
```
.601  .058  .085  .136  .120
.071  .468  .117  .228  .116
.174  .072  .425  .206  .123
.075  .041  .030  .723  .131
.184  .094  .106  .147  .469
```

$C = R_4^{-1} R_5$ (Actual) — $C = R_4^{-1} \hat{R}_5$ (Estimated)

Causative matrix:

Actual:
```
 1.090   .000  -.004  -.024  -.065
  .023  1.080  -.009  -.034  -.060
 -.012   .005  1.090  -.022  -.059
 -.013  -.003   .006  1.080  -.068
 -.126  -.071  -.082  -.007  1.350
```

Estimated:
```
 1.090   .008   .017  -.051  -.065
 -.009  1.110   .018  -.051  -.066
 -.020  -.007  1.090  -.040  -.030
 -.009   .008   .018  1.050  -.066
 -.073  -.079  -.120   .014  1.260
```

Eigenvectors (in columns):

Actual:
```
.242   .353   .791   .822  -.819
.139  -.193  -.561  -.237   .230
.149  -.219   .010  -.377   .518
.313  -.115  -.245  -.425   .083
.157   .881   .530   .004  -.013
```

Estimated: Same as R_4

Eigenvalues:

Actual:
```
1.00   1.41   1.08   1.10   1.09
```

Estimated:
```
1.00   1.10   1.10   1.30   1.10
```

TABLE 1--continued

Actual   Estimated

$$R_4C^2$$

.633	.048	.068	.132	.119		.643	.060	.092	.107	.097
.075	.483	.101	.226	.115		.060	.511	.128	.209	.091
.166	.071	.446	.207	.109		.170	.070	.459	.186	.113
.045	.023	.018	.779	.133		.064	.041	.032	.753	.108
.118	.065	.079	.076	.641		.162	.069	.067	.142	.559

$$R_4C^3$$

.675	.044	.062	.116	.103		.691	.065	.104	.074	.067
.075	.513	.097	.215	.099		.051	.560	.143	.186	.060
.161	.070	.476	.200	.092		.166	.069	.496	.168	.101
.023	.014	.013	.829	.121		.055	.044	.038	.785	.078
.048	.025	.033	.055	.839		.133	.034	.018	.142	.677

$$R_4^{**}$$

.698	.042	.059	.108	.093		.704	.064	.108	.067	.057
.075	.529	.096	.210	.089		.047	.573	.147	.179	.054
.160	.070	.492	.197	.081		.164	.068	.504	.164	.099
.012	.009	.011	.855	.113		.054	.045	.040	.791	.071
.003	.000	.003	.030	.964		.126	.024	.000	.144	.708

Bibliography

1. Blumen, I., Kogen, M., McCarthy, The Industrial Mobility of Labor as a Probability Process, Cornell Studies in Industrial and Labor Relations, 6, (1955) W. F. Humphrey Press Inc., Geneva, New YOrk.

2. Bye, B. and Hennessey, J., "Some Limiting Properties of Finite Nonstationary Markov Chains with Constant Causative Matrix," submitted for publication, obtainable from authors.

3. Coleman, J., Models of Change and Response Uncertainty, Prentice-Hall, 1964.

4. Franklin, P., "Impact of Substantial Gainful Activity on Disabled Beneficiary Work Patterns," Social Security Bulletin, August 1976.

5. Harary, F., Lipstein, B., Styan, G., A Matrix Approach to Nonstationary Markov Chains, Operations Research, Vol. 18 (1970) pp. 1168-1181.

6. Hennessey, J. and Bye, B., "A Linear Algebra Approach to Nonstationary Markov Chains," A working paper, obtainable from the authors.

7. Hennessey, J. and Bye, B., "Necessary and Sufficient Conditions for a Matrix to be Causative in a Nonstationary Markov Chain," submitted for publication; obtainable from authors.

8. Hennessey, J. and Bye, B., "Markov Chain Models of Consumer Behavior-- A New Look at the Causative Matrix" submitted for publication; obtainable from authors.

9. Lazarsfeld, P., and Henry, N., Latent Structure Analysis, 1968, Houghton Mifflin Company, Boston, Massachusetts.

10. Lipstein, B., A Mathematical Model of Consumer Behavior, Journal of Marketing Research, Vol. 2 (August 1965), pp. 259-265.

11. Lipstein, B., Test Marketing: A Perturbation in the Market Place, Management Science, Vol. 14 (8) (April 1968), pp. B437-B448.

12. McKenzie, L., Matrices with Dominant Diagonals and Economic Theory, Mathematical Models in the Social Sciences, 1959, Stanford University Press, 1960, pp. 47-72.

13. Pullman, N., and Styan, G., The Convergence of Markov Chains with Nonstationary Transition Probabilities and Constant Causative Matrix, Stochastic Processes and Their Applications, (1973) pp. 279-285.

14. Singer, B. and Spilerman, S., Social Mobility Models for Heterogeneous Populations, Sociological Methodology (1973-74), San Francisco: Jassey-Bass pp. 356-401.

15. Singer, B. and Spilerman, S., The Representation of Social Processes By Markov Models, Amer. J. Soc., 81, (July).

16. Singer, B., and Spilerman, S., Some Methodological Issues in the Analysis of Longitudinal Surveys, Discussion Paper, Institute for Research on Poverty, University of Wisconsin, Madison, Wisconsin.

17. Spilerman, S. The Analysis of Mobility Processes by the Introduction of Independent Variables into a Markov Chain, Amer. Soc. Rev. 1972, Vol. 37 (June) 277-294.

18. Wiggins, L., Panel Analysis: Latent Probability Models for Attitude and Behavior Processes, Jossey-Bass, Inc., Washington 1973.