last revised: 2 February 2009

2 Social Mobility

The previous chapter introduced the topic of social mobility with a hypothetical example. In this chapter, we'll examine some actual data on social mobility, applying Markov chain analysis and testing some of its assumptions. We'll also consider the mover-stayer model, developed to account for persistence within social classes not well explained by simpler Markov chain models.

2.1 A cross-national comparison

Table 2.1 reports recent estimates of father-to-son mobility between income quintiles within the United States and Sweden. For each matrix in this table, element (i, j) thus indicates the probability that the son's income falls within quintile j given that

Table 2.1 Intergenerational mobility between income quintiles

United States

		Son's income quintile				
		1	**2**	**3**	**4**	**5**
	1	0.422	0.245	0.153	0.102	0.079
Father's	**2**	0.194	0.283	0.208	0.174	0.140
income	**3**	0.194	0.186	0.256	0.202	0.162
quintile	**4**	0.125	0.182	0.198	0.252	0.243
	5	0.095	0.122	0.189	0.234	0.360

Sweden

		Son's income quintile				
		1	**2**	**3**	**4**	**5**
	1	0.258	0.243	0.215	0.176	0.109
Father's	**2**	0.209	0.225	0.237	0.195	0.133
income	**3**	0.183	0.211	0.219	0.223	0.164
quintile	**4**	0.175	0.177	0.196	0.218	0.234
	5	0.163	0.140	0.134	0.193	0.371

Source: Jäntti, et al (2006) "American Exceptionalism in a New Light: Comparison of Intergenerational Earnings Mobility in the Nordic countries, the United Kingdom, and the United States," IZA Discussion paper 1938.

his father's income fell within quintile i. Following the usual convention, social classes are indexed so that class 1 is the bottom quintile (i.e., the 0 to 20th percentile) of the income distribution, while class 5 is the top quintile (i.e., 80th to 100th percentile) of the distribution.

Do these matrices indicate that social mobility is "high" or "low" in each country? One possible benchmark is the case of "perfect mobility" in which the father's class has no effect on the son's life chances. Because we've defined social classes as income quintiles, this would require every son to have a 20% chance of each occupying each class regardless of his father's class. Another benchmark is "zero mobility" which implies that no son ever leaves his father's class. In this case, the transition matrix would appear as an identity matrix (with 1's along the main diagonal and 0's elsewhere). Comparing actual mobility to these extreme cases, it appears that social mobility is closer to the perfect mobility case in both countries.

Nevertheless, comparison of the two matrices reveal some interesting differences. Perhaps the most striking is intergenerational persistence in the lowest quintile (i.e, the probability of transition from class 1 to class 1). In the US, given a father in the lowest quintile, the son has a greater than 42% chance of remaining in that quintile. In Sweden, the corresponding probability is less than 26%. While the differences are less striking, very large "jumps" either upward (from class 1 to 5) or downward (from class 5 to 1) also appear slightly more likely Sweden than in the US. On the other hand, intergenerational persistence in the highest quintile (from class 5 to class 5) is actually slightly higher in Sweden.

To facilitate cross-national comparisons, researchers have developed a variety of mobility indices. Perhaps the simplest is the *trace index* of mobility, defined as

$$m_T \;=\; \frac{k - \mathit{trace}(P)}{k-1}$$

where P is the transition matrix and k is the number of classes. Recalling that the trace of a (square) matrix is the sum of its diagonal elements, note that zero mobility would imply $m_T = 0$ while perfect mobility would imply $m_T = 1$. Social mobility might also be measured by *Bartholomew's index*, which may be defined in the present context as

$$m_B \;=\; (1/k) \sum_i \sum_j P(i,j)|i-j|.$$

Thus, Bartholomew's index weights transitions by the number of categories traversed.

Using Matlab to compute these indices, we find that Sweden scores higher than the US on both measures of social mobility.

```
>> U; % mobility matrix for US (upper panel of Table 2.1)
>> S; % mobility matrix for Sweden (lower panel of Table 2.1)

>> (5-trace(U))/(5-1)   % trace index for US
```

```
ans =    0.8568

>> (5-trace(S))/(5-1)   % trace index for Sweden

ans =    0.9273

>> % computing Bartholomew's index of mobility

>> X = [0 1 2 3 4; 1 0 1 2 3; 2 1 0 1 2; 3 2 1 0 1; 4 3 2 1 0]

X =
     0     1     2     3     4
     1     0     1     2     3
     2     1     0     1     2
     3     2     1     0     1
     4     3     2     1     0

>> U.*X

ans =
         0    0.2450    0.3040    0.3060    0.3160
    0.1940         0    0.2090    0.3480    0.4200
    0.3880    0.1860         0    0.2020    0.3240
    0.3750    0.3640    0.1980         0    0.2430
    0.3800    0.3660    0.3780    0.2340         0

>> (1/5)*sum(sum(U.*X))   % Bartholomew's index for US

ans =    1.1960

>> (1/5)*sum(sum(S.*X))   % Bartholomew's index for Sweden

ans =    1.3678
```

Researchers have proposed many other mobility indices, and there are many more countries that might be compared. But these findings are consistent with other recent comparisons of the US and Sweden.

To push our analysis further, we might suppose that intergenerational social mobility is a Markov chain process, and examine how a father's social class affects the life chances of future descendents. But before jumping immediately to the matrix computations, it is instructive to look more closely at the transition matrices in Table 2.1. By construction, each row of these matrices is a probability vector (allowing for rounding error). Because the social classes are income quintiles – so that each class necessarily contains 20% of the population – we might naïvely expect that the *columns* of these tables must also be probability vectors.[1] However, as you can

[1]Because 20% of fathers belong to each class, the proportion of sons in class j is equal to $\sum_i (0.2)P(i,j)$. If 20% of sons belong to each class, we obtain $\sum_i (0.2)P(i,j) = 0.2$ and hence the requirement that $\sum_i P(i,j) = 1$.

check, some of the column sums are a bit smaller or larger than 1. While this could result partly from sampling variation (since the probabilities are estimates based on samples of the population), it might also reflect a variety of population processes (e.g., differential fertility or immigration) ignored in the simple Markov-chain model developed in the last chapter. Thus, while there is some insight to be gained from Markov chain analysis, we shouldn't ignore these hints that real-world social mobility processes may be more complicated.

To proceed, I have slightly modified the mobility matrices, rounding the entries so that both the rows and the columns of each matrix are probability vectors. Essentially, for analytical purposes, I have created two hypothetical social systems in which social mobility follows a Markov chain process. The first has a "US-like" mobility pattern while the second has a "Sweden-like" mobility pattern. Following our analysis in Chapter 1, we may use matrix computations to determine how an individual's class will affect the life chances of his descendents for the next three generations.

```
>> U % elements rounded so that both rows and COLUMNS are probability vectors

U =
    0.4200    0.2400    0.1500    0.1100    0.0800
    0.1900    0.2800    0.2100    0.1800    0.1400
    0.1900    0.1800    0.2500    0.2100    0.1700
    0.1100    0.1800    0.2000    0.2600    0.2500
    0.0900    0.1200    0.1900    0.2400    0.3600

>> U^2

ans =
    0.2698    0.2244    0.1881    0.1687    0.1490
    0.2053    0.2110    0.2024    0.1958    0.1855
    0.1999    0.1992    0.2031    0.2012    0.1966
    0.1695    0.1896    0.2038    0.2141    0.2230
    0.1555    0.1758    0.2026    0.2202    0.2459

>> U^3

ans =
    0.2237    0.2097    0.1967    0.1892    0.1808
    0.2030    0.2023    0.2001    0.1985    0.1961
    0.2002    0.2001    0.2002    0.2000    0.1995
    0.1896    0.1958    0.2014    0.2048    0.2086
    0.1836    0.1922    0.2017    0.2076    0.2151

>> S % elements rounded so that both rows and COLUMNS are probability vectors

S =
    0.2600    0.2400    0.2100    0.1800    0.1100
    0.2100    0.2300    0.2400    0.1900    0.1300
```

```
    0.1900    0.2100    0.2200    0.2200    0.1600
    0.1700    0.1800    0.2000    0.2200    0.2300
    0.1700    0.1400    0.1300    0.1900    0.3700

>> S^2

ans =
    0.2072    0.2095    0.2087    0.1991    0.1755
    0.2029    0.2061    0.2070    0.2008    0.1832
    0.1999    0.2021    0.2035    0.2013    0.1932
    0.1965    0.1960    0.1968    0.2009    0.2098
    0.1935    0.1863    0.1840    0.1979    0.2383

>> S^3

ans =
    0.2012    0.2021    0.2023    0.2002    0.1941
    0.2006    0.2014    0.2016    0.2002    0.1962
    0.2001    0.2005    0.2006    0.2001    0.1986
    0.1995    0.1991    0.1991    0.2000    0.2024
    0.1985    0.1969    0.1964    0.1995    0.2086
```

Perhaps the most obvious result is that, in either system, there is little effect of the initial father's class after two or three generations. That is, both U^t and S^t converge rapidly to the "perfect mobility" matrix as t increases. Nevertheless, to the extent that there are lingering effects of the father's initial class, these effects are most evident at the corners of these matrices. For instance, focusing on a father initially in the lowest quintile, suppose that intergenerational mobility is governed by the U matrix. His child has a 42% chance of occupying the lowest quintile, his grandchildchild has a 27% chance of occupying the lowest quintile, and his greatgrandchild has a 22% chance of occupying this quintile. Because social mobility is somewhat higher given the S matrix, the corresponding percentages are 26% for children, 21% for grandchildren, and 20% for greatgrandchildren. Thus, effects of lower class status are completely "erased" within three generations in the "Sweden-like" system, but linger slightly longer in the "US-like" system.

2.2 Social mobility over three generations

The preceding results *assumed* that intergenerational social mobility is a Markov chain process. But we already have some cause to believe that real-world social mobility processes are more complicated. To begin to *test* the assumptions implicit in Markov models, it would be useful to have data on the social classes of not only fathers and sons, but also previous (or subsequent) generations within these families.[2]

[2] As a practical matter, such data is sometimes obtained by asking survey respondents about the social class of their fathers and grandfathers. Alternatively, respondents might be questioned about the social class of their fathers and sons.

Table 2.2 reports the data from a Canadian study of social mobility across multiple generations. In this study, social classes were defined by the broad occupational categories of white collar (1), blue collar (2), and farm (3). Because the top table reports frequency counts, the sum of all elements (= 697) indicates the number of respondents. In the bottom table, each row is normalized to be a probability vector, so that we obtain the probability of the respondent's class conditional on the class of his father and paternal grandfather.

To test some of the assumptions of the Markov model, we first need to use this data to construct three transition matrices: a grandfather-to-father matrix A, a father-to-respondent matrix B, and a grandfather-to-respondent matrix C. While this may be accomplished in Matlab in several different ways, I've included my intermediate computations to show you one approach.

```
>> M   % frequency counts from upper panel of Table 2.2
M =
    67    11     0
    19    19     1
     3     2     6
    45    18     1
    55    80     2
     5     8     1
    37    18     3
    56    47     4
    54    89    46

>> % useful to reorganize data into a 3 x 3 x 3 array

>> N(1,:,:) = M(1:3,:);   % first submatrix from M (grandfather's class is 1)
>> N(2,:,:) = M(4:6,:);   % second submatrix from M (grandfather's class is 2)
>> N(3,:,:) = M(7:9,:);   % third submatrix from M (grandfather's class is 3)

>> % N(i,j,k) = number respondents with grandfather in i, father in j, self in k

> NA = sum(N,3)   % grandfathers by fathers

NA =
    78    39    11
    64   137    14
    58   107   189

>> A = diag(1 ./ sum(NA')) * NA   % normalize rows

A =
    0.6094    0.3047    0.0859
    0.2977    0.6372    0.0651
    0.1638    0.3023    0.5339

>> NB = squeeze(sum(N,1))   % fathers by respondents
```

Table 2.2 Occupational mobility over three generations

Frequency counts of paternal grandfather's occupation and father's occupation by respondent's occupation

Grandfather's Occupation	*Father's Occupation*	*Respondent's occupation* White collar	Blue collar	Farm
White collar	White collar	67	11	0
White collar	Blue collar	19	19	1
White collar	Farm	3	2	6
Blue collar	White collar	45	18	1
Blue collar	Blue collar	55	80	2
Blue collar	Farm	5	8	1
Farm	White collar	37	18	3
Farm	Blue collar	56	47	4
Farm	Farm	54	89	46

Transition probabilities

Grandfather's Occupation	*Father's Occupation*	*Respondent's occupation* White collar	Blue collar	Farm
White collar	White collar	0.8590	0.1410	0
White collar	Blue collar	0.4872	0.4872	0.0256
White collar	Farm	0.2727	0.1818	0.5455
Blue collar	White collar	0.7031	0.2813	0.0156
Blue collar	Blue collar	0.4015	0.5839	0.0146
Blue collar	Farm	0.3571	0.5714	0.0714
Farm	White collar	0.6379	0.3103	0.0517
Farm	Blue collar	0.5234	0.4393	0.0374
Farm	Farm	0.2857	0.4709	0.2434

Source: John C. Goyder and James E. Curtis (1977) "Occupational mobility in Canada over four generations," *Canadian Review of Sociology and Anthropology* 14(3):303-319. See Table IV, p 312.

```
NB =
   149    47     4
   130   146     7
    62    99    53

>> B = diag(1 ./ sum(NB')) * NB   % normalize rows

B =
    0.7450    0.2350    0.0200
    0.4594    0.5159    0.0247
    0.2897    0.4626    0.2477

>> NC = squeeze(sum(N,2))   % grandfathers by respondents

NC =
    89    32     7
   105   106     4
   147   154    53

>> C = diag(1 ./ sum(NC')) * NC  % normalize rows

C =
    0.6953    0.2500    0.0547
    0.4884    0.4930    0.0186
    0.4153    0.4350    0.1497
```

One key assumption made in Markov-chain models is that the transition matrix does not change over time. Having constructed the grandfather-to-father matrix A and the father-to-son matrix B, the question is whether $A = B$. It is possible to reject this hypothesis using a formal statistical test.[3] But even casual inspection reveals that these matrices are quite different. To faciliate comparison, let's divide every element of the B matrix by the corresponding element of the A matrix.

```
>> B ./ A

ans =
    1.2226    0.7713    0.2327
    1.5432    0.8096    0.3799
    1.7683    1.5305    0.4639
```

The elements in the first column of this matrix are all greater than 1, indicating that transitions into white collar jobs have become more likely over time, regardless of the class of origin. Conversely, the elements in the second column are all smaller than 1, indicating that transitions into farm jobs have become less likely over time, regardless of class of origin.

[3] See Goodman (1962) for details.

Perhaps these changes in the transition matrix could have easily been predicted given some background knowledge of the dramatic shifts which occurred in the occupational structure over the 20th century. Indeed, using this data, we can see how white-collar employment has risen over time while farm employment has fallen.

```
>> sum(NA')/697   % occupational distribution for grandfathers

ans =
    0.1836    0.3085    0.5079

>> sum(NB')/697   % occupational distribution for fathers

ans =
    0.2869    0.4060    0.3070

>> sum(NB)/697   % occupational distribution for respondents

ans =
    0.4892    0.4189    0.0918
```

Such changes in the occupational structure complicate the analysis of social mobility. While further discussion is beyond our present scope, sociologists often address this issue by decomposing total mobility into two components: "structural" mobility due to shifts in the occupational structure, and "exchange" mobility that would occur within a fixed occupational structure.

Even if the transition matrix is changing over time, there are other assumptions of the Markov chain model that might still be tested. In particular, do the transition probabilities for respondents depend only on father's class? Or, conditioning on the father's class, does the grandfather's class have some influence on the life chances of the respondent? A Markov chain model assumes no "history dependence" beyond the father. If so, the probability that a grandfather in class i has a grandson in class k is given by

$$\sum_j A(i,j)B(j,k)$$

and hence the "expected" grandfather-to-respondent transition matrix is AB. Comparing this expected matrix AB to the actual two-generation transition matrix C computed above, we can thus assess history (in)dependence by examining whether $AB = C$.

```
>> A*B   % "expected" two-generation transition matrix

ans =
    0.6188    0.3401    0.0410
    0.5333    0.4288    0.0378
    0.4156    0.4414    0.1430
```

```
>> C ./ (A*B)   % comparing "actual" to "expected"

ans =
    1.1236    0.7350    1.3336
    0.9157    1.1497    0.4916
    0.9992    0.9855    1.0471
```

To facilitate comparison, I've divided each element of C by the corresponding element of AB. It is intriguing to note that the elements of this matrix are greater than 1 along the main diagonal, indicating more persistence in the grandfather's class than would have been expected if social mobility was history independent.

Returning to Table 2.2, we could also test the history-independence assumption more directly by comparing some of the rows in the second panel. Suppose that the respondent's father held a white-collar job. In the absence of history dependence, the probability vectors given in row 1 (grandfather held white-collar job) and row 4 (grandfather held blue-collar job) and row 7 (grandfather held farm job) should be similar. Similarly, we should see similarity between rows 2 and 5 and 8 (for fathers in blue-collar jobs) and between rows 3 and 6 and 9 (for fathers in farm jobs). While a formal statistical test might be somewhat less conclusive (given the small numbers of respondents in some cells of Table 2.2), some skepticism about the history-independence assumption seems warranted.

Especially given the changes in occupational structure, any attempt to impose a Markov model on this data seems rather perilous. Nevertheless, for conceptual purposes, it may still be useful to see how some degree of history-dependence can be built into Markov chain processes through a clever respecification of the states of the system. To illustrate, suppose that an individual's social class is given by $s \in \{W, B, F\}$. Further suppose that an individual's transition probabilities depend on both his father's and grandfather's class. The states of the Markov chain process might now be specified as

$$S = \{WW, WB, WF, BW, BB, BF, FW, FB, FF\}$$

where each element $ij \in S$ denotes the grandfather's class i and the father's class j. In this way, history dependence is captured by moving from a 3-state model to a 9-state model. Indeed, we could use the data from Table 2.2 to specify the transition matrix shown in Table 2.3. Given this specification of the Markov chain, the father's class in period t becomes the grandfather's class in period $t+1$. Thus, as illustrated, a transition from state ij to state $j'k$ is possible only if $j = j'$.[4]

[4] To incorporate history dependence in a slightly different way, we might retain the simpler specification of the set of states, $S = \{W, B, F\}$, but move from a "first-order" Markov chain process to a "second-order" process which assumes that

$$prob(s_t = k \mid s_{t-1} = j, s_{t-2} = i) = P(i, j, k)$$

where $P(i, j, k)$ is a parameter fixed for every triple (i, j, k). In contrast, by elaborating the set of

Table 2.3 Transition matrix for 9-state Markov chain process

	WW	WB	WF	BW	BB	BF	FW	FB	FF
WW	0.8590	0.1410	0	0	0	0	0	0	0
WB	0	0	0	0.4872	0.4872	0.0256	0	0	0
WF	0	0	0	0	0	0	0.2727	0.1818	0.5455
BW	0.7031	0.2813	0.0156	0	0	0	0	0	0
BB	0	0	0	0.4015	0.5839	0.0146	0	0	0
BF	0	0	0	0	0	0	0.3571	0.5714	0.0714
FW	0.6379	0.3103	0.0517	0	0	0	0	0	0
FB	0	0	0	0.5234	0.4393	0.0374	0	0	0
FF	0	0	0	0	0	0	0.2857	0.4709	0.2434

2.3 The mover-stayer model

Having focused so far on *inter*generational mobility, we now turn to a classic study of *intra*generational mobility conducted by Blumen, Kogan, and McCarthy (1955). This study examined the transitions of male workers between 11 classes of jobs.[5] These workers were observed at quarterly (i.e., three-month) intervals over the course of 8 quarters (i.e., two years). The top panel of Table 2.3 reports transition probabilities after one quarter. Denoting this matrix as P, element $P(i, j)$ is the probability that a worker initially in state i occupied state j at the end of quarter 1. The bottom panel reports transition probabilities after 8 quarters. Denoting this matrix as A, element $A(i, j)$ is the probability that a worker initially in state i occupied state j at the end of quarter 8.

If intragenerational mobility is a Markov chain process (so that the quarterly transition matrix P remains constant over time), mobility at the end of quarter t can be determined by raising the P matrix to the power t. In particular, at the end of 8 quarters, the "expected" transition matrix P^8 should be approximately equal to the "actual" transition matrix A. Thus, a test of the Markov chain assumption is given by the Matlab computations below. As in the preceding section, I have divided each element of the "actual" matrix by the corresponding element of the "expected" matrix to facilitate comparison.

states, we retain a first-order model which assumes that

$$prob(s_t = jk \mid s_{t-1} = ij) = P(ij, jk)$$

where $P(ij, jk)$ is a parameter fixed for every pair (ij, jk), and $P(ij, j'k) = 0$ if $j \neq j'$.

[5]For our present purposes, the precise definition of these classes (labeled **A** through **U** in Table 2.4) is not important. They correspond to the worker's industry (rather than occupation or income) and are not necessarily rankable.

Table 2.4 Intragenerational Mobility

Transitions after one quarter

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**J**	**K**	**U**
A	0.8260	0.0280	0.0180	0.0090	0.0090	0.0180	0.0180	0.0090	0	0	0.0640
B	0.0010	0.8240	0.0110	0.0060	0.0080	0.0080	0.0230	0.0040	0.0100	0.0030	0.1040
C	0	0.0130	0.8850	0.0040	0.0060	0.0060	0.0160	0.0010	0.0050	0	0.0630
D	0	0.0070	0.0030	0.9210	0.0070	0.0040	0.0160	0.0020	0.0020	0	0.0380
E	0	0.0060	0.0050	0.0030	0.9280	0.0040	0.0100	0.0020	0.0050	0.0010	0.0360
F	0	0.0140	0.0060	0.0020	0.0070	0.9010	0.0090	0.0010	0.0060	0.0010	0.0530
G	0	0.0120	0.0100	0.0050	0.0080	0.0040	0.8790	0.0020	0.0100	0.0010	0.0690
H	0.0010	0.0140	0.0020	0.0030	0.0080	0.0010	0.0110	0.8960	0.0090	0.0020	0.0540
J	0	0.0200	0.0090	0.0040	0.0120	0.0090	0.0370	0.0030	0.8220	0	0.0850
K	0	0.1210	0.0400	0.0240	0.0640	0.0320	0.0400	0	0.0080	0.4840	0.1860
U	0.0010	0.0470	0.0320	0.0090	0.0210	0.0150	0.0560	0.0080	0.0260	0.0030	0.7820

Transitions after eight quarters

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**J**	**K**	**U**
A	0.5000	0.0750	0	0	0.0250	0.0250	0.1250	0	0	0	0.2500
B	0.0010	0.6490	0.0220	0.0060	0.0310	0.0210	0.0410	0.0060	0.0090	0.0030	0.2100
C	0.0010	0.0260	0.6810	0.0300	0.0200	0.0250	0.0430	0.0050	0.0090	0.0010	0.1590
D	0	0.0170	0.0130	0.7490	0.0150	0.0200	0.0640	0.0070	0.0100	0.0010	0.1030
E	0	0.0190	0.0260	0.0130	0.7490	0.0210	0.0400	0.0050	0.0090	0.0040	0.1130
F	0	0.0250	0.0090	0.0040	0.0230	0.7560	0.0290	0.0020	0.0070	0.0020	0.1430
G	0	0.0290	0.0300	0.0140	0.0290	0.0130	0.6930	0.0060	0.0280	0.0020	0.1560
H	0	0.0280	0.0070	0	0.0140	0.0020	0.0420	0.7450	0.0350	0	0.1260
J	0.0010	0.0300	0.0230	0.0160	0.0500	0.0180	0.0690	0.0070	0.5730	0	0.2140
K	0	0.1430	0.1040	0.1170	0.1560	0.1690	0.0130	0.0130	0.1170	0.0130	0.1560
U	0.0020	0.0900	0.0630	0.0240	0.0470	0.0330	0.1420	0.0260	0.0430	0.0040	0.5260

Source: Isadore Blumen, Marvin Kogan, and Philip J. McCarthy (1955) *The Industrial Mobility of Labor as a Probability Process,* Cornell University Press. See p 60 and p 63.

```
>> P;   % observed transitions after one quarter (top panel of Table 2.4)

>> A:   % observed transitions after eight quarters (bottom panel of Table 2.4)

>> P^8   % expected transitions after eight quarters

ans =
    0.2179    0.1083    0.0873    0.0445    0.0629    0.0742    0.1177    0.0362    0.0303    0.0022    0.2148
    0.0033    0.2775    0.0778    0.0374    0.0665    0.0497    0.1448    0.0240    0.0554    0.0041    0.2689
    0.0010    0.0790    0.4117    0.0290    0.0526    0.0407    0.1140    0.0130    0.0392    0.0019    0.2132
    0.0007    0.0535    0.0366    0.5274    0.0521    0.0300    0.1028    0.0148    0.0250    0.0014    0.1563
    0.0006    0.0505    0.0435    0.0229    0.5700    0.0301    0.0810    0.0146    0.0333    0.0025    0.1514
    0.0009    0.0800    0.0530    0.0206    0.0555    0.4521    0.0883    0.0123    0.0400    0.0027    0.1953
    0.0010    0.0791    0.0694    0.0331    0.0617    0.0350    0.4252    0.0169    0.0522    0.0028    0.2243
    0.0037    0.0813    0.0402    0.0251    0.0606    0.0226    0.0969    0.4244    0.0486    0.0037    0.1992
    0.0012    0.0973    0.0692    0.0304    0.0754    0.0502    0.1712    0.0205    0.2426    0.0024    0.2448
    0.0019    0.1509    0.1018    0.0556    0.1330    0.0712    0.1535    0.0185    0.0524    0.0062    0.2551
    0.0026    0.1237    0.1058    0.0409    0.0904    0.0585    0.1847    0.0287    0.0699    0.0034    0.2927

>> A ./ P^8   % ratio of actual to expected

ans =
    2.2948    0.6924         0         0    0.3974    0.3371    1.0618         0         0         0    1.1639
    0.3037    2.3391    0.2828    0.1606    0.4658    0.4228    0.2832    0.2499    0.1624    0.7282    0.7810
    1.0295    0.3291    1.6540    1.0351    0.3801    0.6141    0.3772    0.3834    0.2296    0.5378    0.7457
         0    0.3176    0.3556    1.4202    0.2877    0.6674    0.6228    0.4744    0.3999    0.7338    0.6591
         0    0.3760    0.5978    0.5675    1.3141    0.6975    0.4938    0.3426    0.2702    1.6275    0.7466
         0    0.3126    0.1697    0.1946    0.4146    1.6722    0.3286    0.1626    0.1748    0.7395    0.7321
         0    0.3668    0.4325    0.4234    0.4696    0.3716    1.6296    0.3555    0.5367    0.7261    0.6954
         0    0.3446    0.1740         0    0.2309    0.0887    0.4335    1.7555    0.7199         0    0.6324
    0.8165    0.3083    0.3325    0.5263    0.6632    0.3584    0.4031    0.3416    2.3620         0    0.8740
         0    0.9476    1.0213    2.1045    1.1728    2.3731    0.0847    0.7045    2.2328    2.0953    0.6115
    0.7548    0.7275    0.5956    0.5862    0.5201    0.5642    0.7687    0.9071    0.6151    1.1732    1.7970

>> diag(ans)'

ans =
    2.2948    2.3391    1.6540    1.4202    1.3141    1.6722    1.6296    1.7555    2.3620    2.0953    1.7970
```

Although it may be difficult to quickly make sense of these results (given the size of the matrices), the key observation is that the elements along the main diagonal of this last matrix are all greater than 1 (while most of the off-diagonal elements are less than 1). Thus, there is more persistence within job classes (and less movement between job classes) than would be expected if mobility was a Markov chain process. (Recall that we saw a similar result in the last section with regard to intergenerational mobility.)

What sort of process could be generating these results? Blumer, Kogan, and McCarthy hypothesized that the population contains two different types of workers: "movers" whose transitions between job classes follow a simple Markov chain process, and "stayers" who never transition between classes. Intuitively, the stayers will "add weight" to the main diagonal of the transition matrix, and could thus account for the discrepancy between actual and expected outcomes.

To formalize this "mover-stayer" model, let

$$s(i) \;=\; \frac{\text{number of stayers in class } i \text{ in period } 0}{\text{number of individuals in class } i \text{ in period } 0}$$

and hence

$$1 - s(i) \;=\; \frac{\text{number of movers in class } i \text{ in period } 0}{\text{number of individuals in class } i \text{ in period } 0} \;.$$

Let M denote the movers' transition matrix. Finally, let A_t denote the transition matrix for the combined population of movers and stayers after t periods have elapsed. Thus, $A_t(i,j)$ is the probability that an individual in class i in period 0 will occupy class j in period t. Given this notation, the top panel of Table 2.4 becomes A_1 while the bottom panel becomes A_8.

Given these assumptions, we can now determine the elements of the A_t matrix. Intuitively, for individuals initially in class i, proportion $s(i)$ are stayers who will never leave class i, and proportion $1 - s(i)$ are movers who will occupy class i in period t with probability $M^t(i,i)$. Thus,

$$A_t(i,i) \;=\; s(i) + (1 - s(i))M^t(i,i).$$

Because transitions between classes are made only by movers, we obtain

$$A_t(i,j) \;=\; (1 - s(i))M^t(i,j) \quad \text{for all } j \neq i.$$

To adopt more elegant matrix notation, let

$$S \;=\; \begin{bmatrix} s(1) & 0 & \dots & 0 \\ 0 & s(2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & s(n) \end{bmatrix}.$$

The mover-stayer model can then be summarized by the equation

$$A_t \;=\; S + (I - S)M^t$$

where I is the identity matrix.

To illustrate, consider a simple three-class example. If we naïvely supposed that mobility follows a simple three-state Markov chain process, we would expect A_8 to be equal to $(A_1)^8$. But as shown, the A_8 matrix displays higher persistence within classes.

```
>> S   % proportion of stayers in each occupation in period 0

S =
```

```
    0.1000         0         0
         0    0.2000         0
         0         0    0.3000

>> M   % transition matrix for movers

M =
    0.7500    0.2000    0.0500
    0.3000    0.5000    0.2000
    0.1000    0.2000    0.7000

>> I = eye(3);

>> A1 = S + (I-S)*M

A1 =
    0.7750    0.1800    0.0450
    0.2400    0.6000    0.1600
    0.0700    0.1400    0.7900

>> A8 = S + (I-S)*(M^8)

A8 =
    0.5152    0.2571    0.2277
    0.3581    0.4286    0.2133
    0.3006    0.2000    0.4994

>> A1^8   % a naive prediction

ans =
    0.4348    0.2933    0.2719
    0.4066    0.2906    0.3029
    0.3608    0.2825    0.3567

>> A8 ./ (A1^8)   % ratio of actual to expected

ans =
    1.1849    0.8768    0.8373
    0.8808    1.4751    0.7042
    0.8333    0.7079    1.4000
```

Given that it was originally developed as an alternative, it may be surprising to learn that the mover-stayer model can itself be formulated as a Markov chain process. Assuming n classes, this process has $2n$ states, which might be arranged as

$$1\text{S}, 2\text{S}, \ldots, n\text{S}, 1\text{M}, 2\text{M}, \ldots, n\text{M}$$

where iS denotes a stayer in class i, and iM denotes a mover in class i. The $2n \times 2n$

transition matrix may be written as

$$P = \begin{bmatrix} I & 0 \\ 0 & M \end{bmatrix}$$

where I denotes an $n \times n$ identity matrix, 0 denotes an $n \times n$ matrix of zeros, and M is the $n \times n$ transition matrix for movers. Thus, for our 3-class example, the transition matrix is given by

```
>> P = [eye(3) zeros(3); zeros(3) M]  % mover-stayer transition matrix

P =
    1.0000         0         0         0         0         0
         0    1.0000         0         0         0         0
         0         0    1.0000         0         0         0
         0         0         0    0.7500    0.2000    0.0500
         0         0         0    0.3000    0.5000    0.2000
         0         0         0    0.1000    0.2000    0.7000
```

Unlike other examples we have encountered so far, the limiting distribution for this Markov chain process *does* depend on the initial condition. To illustrate, consider the limiting distribution for two different initial distributions of workers across states.

```
>> x = [1/3 1/3 1/3]*[S, I-S]   % one initial distribution

x =
    0.0333    0.0667    0.1000    0.3000    0.2667    0.2333

>> x*P^100   % limiting distribution

ans =
    0.0333    0.0667    0.1000    0.3592    0.2286    0.2122

>> x = [1/4 1/4 1/2]*[S, I-S]   % a different initial distribution

x =
    0.0250    0.0500    0.1500    0.2250    0.2000    0.3500

>> x*P^100   % limiting distribution

ans =
    0.0250    0.0500    0.1500    0.3480    0.2214    0.2056
```

According to the Theorem stated in Chapter 1, primitivity of the transition matrix implies that a Markov chain process reaches a unique limiting distribution regardless of the initial condition. Thus, from our computations, we know that the transition matrix P must not be primitive. In fact, it is easy to see that

$$P^t = \begin{bmatrix} I & 0 \\ 0 & M^t \end{bmatrix}$$

and hence some elements of this matrix will be 0 for all t. For our present example, the primitivity of M ensures that every row of M^t converges to its limiting distribution, but this not true for every row of the full P matrix.

```
>> P^100   % long-run outcome

ans =
    1.0000         0         0         0         0         0
         0    1.0000         0         0         0         0
         0         0    1.0000         0         0         0
         0         0         0    0.4490    0.2857    0.2653
         0         0         0    0.4490    0.2857    0.2653
         0         0         0    0.4490    0.2857    0.2653
```

The mover-stayer model is the first example we have encountered of a Markov chain process with *absorbing states*. We will study this important class of Markov chain models in Chapter 4.

2.4 Further reading

Sociologists have long been interested in intergenerational social mobility. The books by Blau and Duncan (1967) and Featherman and Hauser (1978) are classics. Recent surveys include Breen and Jonsson (Ann Rev Soc 2005), and Beller and Hout (The Future of Children 2006). See Boudon (1973) for review and development of matrix methods for mobility research.

While sociologists usually define social classes based on occupation, economists tend to define classes by income. See Solon (J Econ Persp 2002) for a survey of economic research on income mobility. Jäntti et al (2006) compare mobility across 6 countries using 4 different mobility indices. For a recent attempt to bridge economic and sociological approaches to social mobility, see the volume edited by Morgan, Grusky, and Fields (2006).

My analysis of mobility over three generations draws on Hodge (Demography 1966) and also Goodman's (AJS 1962) discussion of statistical tests of Markov chain assumptions. While Hodge (1966) and Goyder and Curtis (1977) find evidence of history dependence beyond the father's generation, a recent study by Warren and Hauser (ASR 1997) reaches the opposite conclusion.

The mover-stayer model was originally developed by Blumen, et al (1955). My presentation draws heavily on Leik and Meeker (1975, Chap 9). See also Bradley and Meeks (1986, Chap 7).