Log-linear analysis of contingency tables: An introduction for historians with an application to Thernstrom on the "Floating Proletariat"

J. Morgan Kousser
California Institute of Technology
Gary W. Cox
University of Texas
David W. Galenson
University of Chicago

Suppose a researcher has information on several attributes of a collection of individuals and that the data he has are available only in qualitative (synonyms are categorical, discrete, polytomous, or ordinal- or nominal-level), as opposed to quantitative (continuous or interval-level) form. For instance, imagine that his information is about yes or no votes, occupational classes, or age groups, but none is in the form of, say, the dollar amounts of property held (not broken into categories) or the length of residence, in months or years, at a particular location. Then he might construct tables, such as Table 1, which show how many people have each set of traits; for example, how many young, unskilled, childless men in a sample were found in both the Boston census schedule in 1880 and the city directory in 1890. When there is very little information available, say, data on only two or three variables, commonsensical methods of analysis may suffice. But what should one do when one is confronted by such monsters as the eighty-celled "four-way" Table 1?

The conventional historical answer to this question has been to combine the categories (or, to put it another way, to collapse the table) into what are called "marginal tables," relating two or perhaps three variables, as in the panels in Table 5, below. While this is useful and often informative, the practice may hide information which is available in the full table. Fortunately, in the past fifteen years statisticians have developed new methods for squeezing many more conclusions out of such tables. Historians have made too little use of the new techniques, generally denominated "log-linear contingency table analysis," probably because the initial articles and books explaining them were somewhat obscurely phrased and were not easily accessible to those who lacked fairly advanced statistical training. Now that there are simpler texts on the market and several computer programs available, it is time that many more historians took advantage of them.

This paper is intended to provide both an intuitive and a practical mathematical understanding of the log-linear technique, demonstrate its usefulness by reexamining an important historical topic, and, by making every step in the development and application of the technique explicit, using the notation now common in the literature, encourage and prepare historians to make use of log-linear analysis as well as to be able to go on to more advanced treatments in statistics texts. Those who desire a brief overview of the subject may wish to save Section III, in which we lay out the algebra step by step, for a second reading, while those already familiar with or indifferent to log-linear methods may wish to skip Sections I through V. We will also attempt to show how social scientific theory can help to guide data analysis and shall emphasize a hitherto often overlooked facet of a much-studied historical problem. Thus we hope to blend substantive with methodological points. Written at a fairly elementary level and self-contained, the paper assumes only that one has a speaking acquaintance with

0161-5440/82/1505-0152$1.00

such statistical concepts as Chi-Square tests and regression analysis.

I. Historical Mobility Studies

Before beginning the statistical discussion, let us briefly introduce the substantive problem with which we shall be concerned throughout this article. During the past two decades, many historians have investigated geographic and social mobility in nineteenth century American cities. Aimed primarily at systematically describing those characteristics of individuals which were associated with changes in residence and in occupational or social rank, their works have related both types of mobility to such variables as age, occupation, family social status, property holdings, ethnic origins, and generation of residence in America.[1] Drawing on such previously unexploited sources as federal and state manuscript censuses, city directories, and tax assessment rolls, the "new social historians" have attracted a good deal of attention by taking advantage of the chance these sources offer to study the lives of large numbers of individuals who have previously eluded the view of historians.

Yet these scholars have failed to make use of the available statistical methods and social scientific theory as fully as they have ransacked the sources. More specifically, they have generally related only two or three variables to each other, thereby in effect assuming that the numerous "independent variables" in their mobility analyses were uncorrelated with each other; their implicit statistical models, furthermore, generally assume, without testing, that the relationships they seek are linear. By focusing on the different correlates of mobility only one or two at a time, they have generally settled for mere description, instead of confronting directly the problem of building a cohesive explanation. And their analyses have been less well informed by social scientific theory, particularly economic theory, than they might have been. We shall illustrate how these problems might have been largely obviated by reanalyzing data on geographic mobility gathered by Stephan Thernstrom for his study *The Other Bostonians*, which traces individuals from the 1880 federal manuscript census to the 1890 city directories of Boston and its suburbs.[2]

From Thernstrom's data set, we have chosen three factors which all plausibly bore on the 1880 Bostonian's decision to move or stay: family status, which we will call "*S*" and which we cut into two groups: the first, single or married but childless, and the second, married with children; occupation, or "*O*," which we broke into five classes: high white collar, low white collar, skilled, unskilled, and unemployed; and age, or "*A*," which we cut into four sets: 14-20, 21-30, 31-60, and over 60.[3] The number of people in the sample with each set of traits in both 1880 and 1890 is displayed in the four panels of Table 1. Since it is probable that some of those who were not listed in the 1890 directory were simply overlooked by the canvassers but were still present in the Boston area, we sill hereafter refer to the division of the sample as those who were "found" and "not found" rather than as "stayers" and "movers" or "persisters" and "nonpersisters."[4] To avoid confusion with the conventional notation in the statistical literature, however, we will label the variable "*M*."

Table 1.—Number of males persisting, 1880-90, by age, family status, and occupation

	Occupation					
Age	Hi.W.	Lo.W.	SK.	UNSK.	UNEMP.	TOTAL
Panel A: *Single or Married without Children, Persistent*						
14-20	1	51	21	57	106	236
21-30	13	72	50	54	17	206
31-60	7	12	11	13	1	44
61+	1	1	0	0	0	2
TOTAL	22	136	82	124	124	488
Panel B: *Single or Married without Children, Not Found in 1890*						
14-20	1	28	11	37	46	123
21-30	9	36	37	61	9	152
31-60	2	1	10	11	1	25
61+	0	0	0	0	2	2
TOTAL	12	65	58	109	58	302
Panel C: *Married with Children, Persistent*						
14-2C	0	0	0	0	0	0
21-30	2	16	22	33	1	74
31-60	94	98	168	162	6	528
61+	9	3	1	9	1	23
TOTAL	105	117	191	204	8	625
Panel D: *Married with Children, Not Found in 1890*						
14-20	0	0	0	0	0	0
21-30	2	6	14	17	1	40
31-60	29	43	68	99	1	240
61+	8	3	5	8	5	29
TOTAL	39	52	87	124	7	309

*The occupational categories are high white collar, low white collar, skilled, unskilled, and unemployed. The classification of occupation into categories was done by Thernstrom.

The relationship of migration with occupation plays a large role in Thernstrom's explanation, that with age is stressed in the economics literature, and that with children taps the notion that familial responsibilities constrain. We also initially included measures of home ownership, the number of generations a man had been in America, and ethnicity in order to attempt to measure some of the effects, respectively, of different levels of transactions costs involved in the decision to move, rootedness, and a possible high degree of employment discrimination (especially against the Irish) in Boston. But these three variables were either so closely related to age, occupation, and family status or had so little impact that they did not add much to our explanation. In the interest of simplifying the discussion, therefore, we have left them out of the analysis presented here.

II. Simple Manipulations of Tables

Before beginning the analysis, we need to define a few terms and establish some appropriate conventions. To identify each cell in a table, let us refer to each by a set of subscripts, beginning with "1" or, more generally, with "i" and proceeding by integers or alphabetically as long as we need them. Thus, the cells in Table 1 are identified by four letters or integers. Here and throughout this paper, the variables will be considered in the order M, A, S, O. For instance, the entry in the bottom right-hand corner in Table 1 is referred to as the (2, 4, 2, 5) cell, or that in which the value of M is arbitrarily called 2, the value of A is termed 4, the value of S is 2, and the value of O is 5. Substantively, the cell represents the number of people present in 1880 who were not found in 1890 and who had been aged 61 or older, had children, and were unemployed in 1880. We will refer to the actual cell entries by small f's, and we will subscript them by numbers or by i, j, k, etc. For example, $f_{2425} = 5$. Estimates of the cell entries, obtained by procedures to defined later, will be designated by capital F's. When we sum across all values of a variable (for example, when we add the people in all age categories together but preserve our knowledge of their occupations, family statuses, and whether or not they were found in 1890), we will replace the relevant subscript with a "plus." Summing across age while holding the other variables at levels i, k, and l would thus be noted as f_{i+kl}.

One can often discover a good deal about the relationships in a table by performing quite simple operations on it. Since several of these operations are directly related to the log-linear techniques on which we will focus, a discussion of commonsensical methods will lead naturally into the explanation of these more formal methods. The first step that almost anyone would take after perusing Table 1 would be to form percentages from it, and the first of several possible percentages to calculate would be the percentage "found" within each age (subscripted by j), family (k), and occupational (l) grouping. Using the cell entry notation developed above, this percentage would be:

(1) % found $= f_{1jkl}/f_{+jkl}$.

For instance, the percentage "found" among low white collar childless men aged between fourteen and twenty is 51/(51 + 28) = 65%.

Table 2.—Percent found, 1880 and 1890

Age	Hi.W.	Lo.W.	SK.	UNSK.	UNEMP.	TOTAL
			No Children			
14-20	50	65	66	61	70	66
21-30	59	67	57	47	65	58
31-60	78	92	52	54	50	64
61+	100	100	—	—	0	50
TOTAL	65	68	59	53	68	62
			Children			
14-20	—	—	—	—	—	—
21-30	50	73	61	66	50	65
31-60	76	70	71	62	86	69
61+	53	50	17	53	17	44
TOTAL	73	69	69	62	53	67

Tables of percentages often reveal more striking relationships than Table 2 does. Whereas for the childless, relationships between age and being found are nearly monotonic and are positive for the two higher occupational classes and negative for the three lower classes, the relationships for men with children are much more mixed across class and age. Looking at the "total" or "marginal" rows and columns, it is clear that childless men in their twenties and unskilled men regardless of age were especially likely not to be found in the 1890 city directory, but that the percentage "found" among the high white collar and skilled and unskilled worker classes depended crucially on family status in 1880. Table 2 thus suggests that the three independent variables interacted with each other to produce a pattern too complex to be decoded with simple percentages and linear assumptions.

But of course there are other ways to compute percentages from the raw data, and they may be more revealing. Tables 3 and 4 are calculated by first summing across M and then dividing each entry by the row marginal (total) for Table 3, or the column marginal for Table 4. In the cell entry notation introduced above, the equations are

(2) Table 3 entry $= f_{+jkl}/f_{+jk+}$

and

(3) Table 4 entry $= f_{+jkl}/f_{++kl}$.

Table 3 reveals that 42 percent of males aged fourteen to twenty in 1880 were unemployed, that among those with children, the unskilled made up a strikingly smaller percentage of those above thirty than of those under thirty years old, that the relationship between age and the percentage who were in the highest occupational class was unambiguously monotonic and positive, while that between age and the percentage in the low white collar class was negative, but weak. Table 4 demonstrates that in every class fathers tended to be middle-aged,.and that in all but one class, the single and childless were most likely to be in their twenties. The modal age category for the childless unemployed was the teenage one. The percentages in these tables thus give us

Table 3.—Percent of total sample (found plus not found) summed across rows

Age	Hi.W.	Lo.W.	SK.	UNSK.	UNEMP.	TOTAL
			Panel A: *No Children*			
14-20	0*	22	9	26	42	100
21-30	6	30	24	32	7	100
31-60	13	19	30	35		100
61+	25	25	—	—	50	100
All Ages	4	25	18	29	23	100
			Panel B: *Children*			
14-20	—	—	—	—	—	—
21-30	4	19	32	44	2	100
31-60	16	18	31	34	1	100
61+	33	12	12	33	12	100
All Ages	15	18	30	35	2	100

*0 indicates less than 0.5%; — indicates no cases in cell.

Table 4.—Percent of total sample summed across columns

Age	Hi.W.	Lo.W.	SK.	UNSK.	UNEMP.	TOTAL
			Panel A: *No Children*			
14-20	6	39	23	40	84	45
21-30	65	54	62	49	14	45
31-60	26	6	15	10	1	9
61+	3	0*	—	—	—	1
TOTAL	100	100	100	100	100	100
			Panel B: *Children*			
14-20	—	—	—	—	—	—
21-30	3	13	13	15	13	12
31-60	85	83	85	80	47	82
61+	12	4	2	5	40	6
TOTAL	100	100	100	100	100	100

*0 indicates less than 0.5%; — indicates no cases in cell.

some sense of the interrelationships among the independent variables.

Another way to try to make sense of complex tables is to collapse them into bivariate displays, which is what historians who lack knowledge of multivariate techniques generally do. The panels in Table 5 show the six bivariate tables which can be drawn from Table 1. A convenient way to refer to them is to enclose the symbols for the variables in curled brackets. Thus the shorthand for panel A of Table 5 is $\{MO\}$, for panel B, $\{MA\}$, and so on. For anyone who has had the most elementary statistics course, the immediate reflex action upon confronting such tables is to compute Chi-Squares, and we have done so, finding such high values in each panel that every table contains a significant relationship at the conventional 0.05 level.

In fact, the reflex in this case is quite desirable, for the Chi-Square distribution can be employed to accomplish much more sophisticated purposes than its usual cookbook use suggests. The Chi-Squares computed in Table 5 test whether the two variables in a panel are "independent." Consider panel C. If M and S were independent in a statistical sense, then the value of each cell would be purely a product of the relevant marginals. For instance, the top left cell would be equal to $790 \times 1113/1724 = 510$. The other entries in panel C would be 602, instead of 625; 280, instead of 302; and 331, instead of 309. As applied here, then, the so-called "Pearson Chi-Square" statistic enables one to compare the observed data to a criterion, that is, the Chi-Square distribution, in order to determine whether the model of independence between the two variables fits the data well or not. Its formula suggests its nature quite clearly:

(4) Pearson Chi-Square $= \sum (f_{ij} - F_{ij})^2/F_{ij}$,

where $\sum$ indicates a summation over all the values for two variables, the small f's refer to the cell entries actually observed, and the large F's, to the frequencies expected under the independence model, in this case, the values 510, 602, 280, and 310, respectively. But since independence is not the only possible model, we can substitute for these particular F's predictions generated by any model which we can specify mathematically. This is one of the keys to log-linear analysis.

A final simple but instructive permutation of Table 1 is shown in Table 6. There we have calculated the probability of being found divided by that of not being found for each cell in Table 1. That is, instead of using Equation (1), we have calculated the entries by using the following:

(5) Odds of being found $= f_{1jkl}/f_{2jkl}$.

The "odds," familiar from horse racing, are simply the ratio of the number found to the number not found. Now, to make heads or tails of the probabilities in Table 6, we must compare the cells to each other, and one natural way of doing so is to divide the entry in one cell by that in another, or, to put it another way, to form an "odds ratio." For example, for skilled workers between thirty-one and sixty, the odds of being found rose from 1.10 to 2.47 if they had children in 1880, producing an odds ratio of 2.25; whereas, the analogous odds ratio for high white collar men was $(3.24/3.50) = 0.93$.

To move beyond these commonsensical operations to fully multivariate methods, we need to develop ways of specifying, estimating, and choosing between different models.[5] These models, of which the independence model that forms the basis of the traditional Chi-Square test is the most familiar, will yield various estimates of

Table 5.—Six two-way marginal tables based on Table 1

Panel A: *Occupation and Persistence*

	Hi.W.	Lo.W.	SK.	UNSK.	UNEMP.
Found	127 (71.3)*	253(68.4)	273(65.3)	328(58.5)	132(67.0)
Not Found	51	117	145	233	65
TOTAL	178(100.0)	370	418	561	197

$\chi^2 = 15.66$

Panel B: *Age and Persistence*

	14–20	21–30	31–60	61+
Found	236 (65.7)	280(59.3)	572(68.3)	25(44.6)
Not Found	123	192	265	31
TOTAL	359(100.0)	472	837	56

$\chi^2 = 20.81$

Panel C: *Family Status and Persistence*

	No Children	Children	Total
Found	488 (61.8)	625(66.9)	1113(64.6)
Not Found	302	309	611
TOTAL	790(100.0)	934	1724

$\chi^2 = 4.95$

Panel D: *Occupation and Age*

Age	Hi.W.	Lo.W.	SK.	UNSK.	UNEMP.	TOTAL
14–20	2 (1.1)	79(21.4)	32 (7.7)	94(16.8)	152(77.2)	359
21–30	26(14.6)	130(35.1)	123(29.4)	165(29.5)	28(14.2)	472
31–60	132(74.2)	154(41.6)	257(61.5)	285(50.8)	9 (4.6)	837
61+	18(10.1)	7 (1.9)	6 (1.4)	17 (3.0)	8 (4.1)	56
TOTAL	178(100.0)	370	418	561	197	

$\chi^2 = 559.21$

Panel E: *Age and Family Status*

	14–20	21–30	31–60	61+
No Children	359(100.0)	358(75.8)	69(8.2)	4(7.1)
Children	0	114	768	52
TOTAL	359(100.0)	472	837	56

$\chi^2 = 1105.72$

Panel F: *Occupation and Family Status*

	Hi.W.	Lo.W.	SK.	UNSK.	UNEMP.
No Children	34(19.1)	201(54.3)	140(33.5)	233(41.5)	182(92.4)
Children	144	169	278	328	15
TOTAL	178(100.0)	370	418	561	197

$\chi^2 = 263.77$

*Percentages, summed by column, in parenthesis.

Table 6.—Odds of being found, 1880 and 1890 (X 100)

Age	Hi.W.	Lo.W.	SK.	UNSK.	UNEMP.	TOTAL
			Panel A: *No Children*			
14-20	100	182	191	154	230	192
21-30	144	200	135	89	189	136
31-60	350	1200	110	118	100	176
61 +			—	—		100
TOTAL	183	209	141	114	214	162
			Panel B: *Children*			
14-20	—	—	—	—	—	—
21-30	100	267	157	194	100	185
31-60	324	228	247	164	600	220
61 +	113	100	20	113	20	79
TOTAL	269	225	220	165	114	202

the cell entries. We can then compare different sets of predictions to the observed entries and find one or more which are both sufficiently close to the reality of the table and sufficiently parsimonious to suit our tastes. Because it is the simplest table presented thus far, let us use panel C of Table 5 to outline the techniques.

The analysis of variance, which is often used to examine cross-classification tables, suggested to statisticians that an equation containing a "grand mean effect," separate effects for each variable, and terms for the possible interactions between variables would be a good place to start. Since, as we will show, a multiplicative (but not a linear, additive) equation allows tests for the statistical independence of two or more variables, we will use an equation in multiplicative form to estimate the entries in panel C of Table 5:

$$(6)\ F_{ik} = \eta \tau_i^M \tau_k^S \tau_{ik}^{MS},$$

where the η (eta) is a form of "grand mean effect," the τ_i^M (tau) is the effect of being found or not found, τ_k^S is the effect of familial status, and τ_{ik}^{MS} is the effect of the interaction between M and S.[6] But because, as the subscripts indicate, Equation (6) actually contains nine effects and we have only four cells on which to base our estimates, we have to make additional assumptions in order to be able to estimate anything.[7] More formally, without additional constraints, the model is said to be "underidentified." Fortunately, the necessary assumptions are quite natural. Since we are really interested in the effect of having children, for example, and not in determining separate effects for having and not having children, we assume that:

$$(7)\ \tau_1^M = \frac{1}{\tau_2^M},\ \tau_1^S = \frac{1}{\tau_2^S},\ \tau_{11}^{MS} = \frac{1}{\tau_{12}^{MS}} = \frac{1}{\tau_{21}^{MS}} = \tau_{22}^{MS}.$$

These assumptions reduce the number of parameters to be estimated by five, producing what is called a "saturated" model; that is, one containing as many basic parameters to be estimated as there are cells, and therefore having no "degrees of freedom." Furthermore, it is clear from Equation (6) that the effect of each tau parameter may be measured as a deviation from a value of 1.00, for if any tau equals 1.00 in the multiplicative form, it has no impact on the value of the function. Note that the taus, unlike the Pearsonian correlation coefficient or such familiar coefficients for cross-classification tables as phi or Yule's Q, do not vary only between zero and one or minus one and plus one. In fact, the taus have no upper bounds.

III. The Algebra of Log-linear Models

Equation (6) is directly related to various odds ratios formed from a slightly altered form of panel C of Table 5 (or, more generally, of any table of counts). Table 7 transforms panel C of Table 5 into proportions by dividing each cell entry by the table total. For instance, 488/1724 = 28.3%.

Since it turns out to be more convenient to do computations in logarithmic form, we will take natural logarithms of both sides of Equation (6), producing

$$(8)\ \log p_{ik} = \log \eta + \log \tau_i^M + \log \tau_k^S + \log \tau_{ik}^{MS},$$

where p_{ik} refers to an entry in Table 7.[8] Each of the terms on the right-hand side of Equation (8) can now be expressed in terms of the p_{ik}'s:

$$(9)\ \log \eta = \tfrac{1}{4} (\log p_{11} + \log p_{12} + \log p_{21} + \log p_{22}),$$

$$(10)\ \log \tau_1^M = \tfrac{1}{4} (\log p_{11} + \log p_{12} - \log p_{21} - \log p_{22}),$$

$$(11)\ \log \tau_1^S = \tfrac{1}{4} (\log p_{11} - \log p_{12} + \log p_{21} - \log p_{22}),$$

$$(12)\ \log \tau_{11}^{MS} = \tfrac{1}{4} (\log p_{11} - \log p_{12} - \log p_{21} + \log p_{22}),$$

Because of the constraints imposed in (7) above, the values of the tau parameters for other "levels" (i.e., categories) of the associated variable can be immediately derived, since, for instance (from Equation 7),

$$(13)\ \tau_1^S = \frac{1}{\tau_2^S}.$$

If we take natural logs, we have

$$(14)\ \log \tau_1^S = -\log \tau_2^S,$$

and all we have to do get τ_2^S is to change the signs of the p's in Equation (11).

Notice also that

$$(15)\ \log p_{11} + \log p_{12} - \log p_{21} - \log p_{22} \equiv \log (\frac{p_{11}\, p_{12}}{p_{21}\, p_{22}}) = \log ((\frac{p_{11}}{p_{21}}) (\frac{p_{12}}{p_{22}})).$$

The quantity to the right of the last equals sign is the logarithm of the product of two odds ratios, each of which consists of the proportion of males found divided by the proportion not found, the first fraction in parentheses being for adult males without children, the second, with children.[9] Since both fractions measure part

Table 7.—Panel C of Table 5 expressed as proportions

	No Children	Children	
Found	P_{11} = 28.3	P_{12} = 36.3	P_{1+} = 64.6
Not Found	P_{21} = 17.5	P_{22} = 17.9	P_{2+} = 35.4
	P_{+1} = 45.8	P_{+2} = 54.2	P_{++} = 100

of the "effect" of being found versus not being found for this particular table, the interpretation of log τ_1^M as the "effect" of variable M is quite natural.

Furthermore, the statement that the middle quantity in parenthesis in Equation (15) equals unity, in which case its log is zero and the value of τ_{11}^{MS} is also zero, corresponds to the definition of the statistical independence between the two variables. To see this, not that if any two variables i and j are independent,

$$(16)\ p_{ij} \equiv p_{i+}\, p_{+j}.$$

Substituting these values into the odds ratio in (12), we have

$$(17)\ p_{11}\, p_{22}/p_{12}\, p_{21} = \frac{p_{+1}\, p_{1+}\, p_{+2}\, p_{2+}}{p_{+2}\, p_{1+}\, p_{+1}\, p_{2+}},$$

and since each of the summed terms cancels out algebraically, the odds ratio should be equal to one when the two variables are independent.[10]

It is also easy to show that if we use Equations (9) through (12) to predict cell proportions, the model predictions in a saturated model exactly equal the observed proportions in the original table.[11] From Equations (9) through (12), we know that

$$(18)\ \log p_{11} = \tfrac{1}{4} \log (p_{11}\, p_{12}\, p_{21}\, p_{22}) + \tfrac{1}{4} \log (\frac{p_{11}\, p_{12}}{p_{21}\, p_{22}}) + \tfrac{1}{4} \log (\frac{p_{11}\, p_{21}}{p_{12}\, p_{22}}) + \tfrac{1}{4} \log (\frac{p_{11}\, p_{22}}{p_{12}\, p_{21}}).$$

Rearranging terms and taking antilogs, we have

$$(19)\ p_{11} = (p_{11}\, p_{12}\, p_{21}\, p_{22} (\frac{p_{11}\, p_{12}}{p_{21}\, p_{22}})(\frac{p_{11}\, p_{21}}{p_{12}\, p_{22}}) (\frac{p_{11}\, p_{22}}{p_{12}\, p_{21}}))^{1/4}.$$

Canceling terms algebraically on the right-hand side, we obtain

$$(20)\ p_{11} = (p_{11}^4)^{1/4} \equiv p_{11}.$$

In the same manner, the reader may satisfy himself that the predicted or right-hand side proportions p_{12}, p_{21}, and p_{22} are precisely equal to the observed or left-hand side proportions. And when the two sets of proportions (and therefore the corresponding cell frequencies) are equal, the Pearson Chi-Square statistic given in Equation (4) is zero, for then

$$(21)\ (f - F)^2 = 0.$$

A Chi-Square statistic can therefore be used as a test of the independence between two variables in a log-linear model.

Moreover, if we want to test an "unsaturated" model—that is, one containing fewer parameters than cells, such as Equation (8) with the last term on the right-hand side deleted—we can simply set the term or terms equal to zero in the logarithmic form or one in the multiplicative form and use the remaining log odds ratios from Equations (9) to (11) to estimate the cell proportions. To determine how well the new model fits the original observations, we can compute a Chi-Square statistic with the resulting predicted cell proportions as our F's.

When there are more than two categories for a variable, and/or when there are more than two variables, the model estimation procedure becomes more complex. Many estimates which have the statistically desirable property of being "maximum likelihood" cannot be computed from "closed form" expressions. That is to say, while one can always write out such simple expressions as Equations (9) to (12), in some cases the resulting estimates will not be the "best" which can be obtained. In these cases, fortunately, one can use either of two algorithms, which are called "iterative proportional fitting" or the "Newton-Raphson" procedure, to approximate the F's. Because the principles involved in generating and interpreting the models remain basically the same for larger tables as for 2 × 2 tables, and for numerically approximated as for closed-form estimates, and because computers can so quickly and accurately run through the algorithms that a data analyst need not really understand those parts of the routines in order to interpret the output, we will not prolong the present discussion by explicating these matters.[12]

IV. Hierarchy, Conditional Independence, and Other Models

Log-linear models of the variety with which we are dealing are constructed according to the "hierarchy principle." Consider an equation for a saturated model similar to Equation (8) but involving three, rather than two, variables. For the sake of simplicity, we will hereafter refer to the log of tau terms as λ's (lambdas), drop the subscripts, and temporarily assume that each of the variables is divided into only two classes. Instead of four basic terms, the equation for the three-variable model has eight:

$$(22)\ \log p_{ikl} = \log \eta + \lambda^M + \lambda^S + \lambda^O + \lambda^{MS} + \lambda^{MO} + \lambda^{SO} + \lambda^{MOS}.$$

The hierarchy principle dictates that one cannot consider a model involving only "higher-order" terms without including the corresponding lower-order terms. For example, the following models are forbidden:

(23) $\log p_{ikl} = \lambda^{MOS}$ (leaves out seven lower-order terms);

(24) $\log p_{ikl} = \log \eta + \lambda^{M} + \lambda^{S} + \lambda^{O} + \lambda^{MS} + \lambda^{MO} + \lambda^{MOS}$ (leaves out λ^{SO});

(25) $\log p_{ikl} = \log \eta + \lambda^{M} + \lambda^{S} + \lambda^{SO}$ (leaves out λ^{O}).

It is, however, perfectly proper to propose

(26) $\log p_{ikl} = \log \eta + \lambda^{M} + \lambda^{S} + \lambda^{MS}$,

because all the variables involved in the second-order term λ^{MS} are represented in first-order terms in the equation.

In social scientific applications, the hierarchy principle usually seems quite natural. While it is possible to think of situations in which the idea that one variable is a social catalyst is more than a metaphor, in most instances for which we have enough data to require multivariate methods, the principle of hierarchy will not preclude the testing of any model of interest.

In any case, the hierarchy principle enables us to simplify notation. Hereafter, instead of using equations, we will refer to models using the curled bracket or "fitted marginals" mode of expression already introduced, but we will adopt the shorthand method of explicitly noting only the highest-order terms we want to put in the model, since by the hierarchy principle, all lower-order terms involving those variables must be included.[13] For example, $\{MS\}$ will mean the same thing as Equation (26), and $\{MOS\}$ will be equivalent to Equation (22).

Log-linear models can be constructed so as to give mathematical form to notions other than independence. In "square" tables (that is, tables containing two variables which have the same number of categories) one can test for "symmetry," "quasi-symmetry," and a variety of other interesting concepts, several of which have been applied by sociologists to the study of occupational mobility.[14] If there are "structural zeroes" in a table of any size or shape (for example, if one were trying to determine the importance of legal immigration compared to that of the fertility of the American-born in the growth of various ethnic groups in America, there would be periods when the immigration of natives of some Asian countries or Africa was prohibited by law) one can estimate models of "quasi-independence" which exclude the bothersome cells.

An even more basic idea, useful in analyzing any table with more than two variables, is that of "conditional independence." Put most simply, conditional independence means that if we control for one or more variables, the apparent relationship between two or more other variables disappears, as illustrated for a hypothetical three-variable case in Table 8. If M and S were in fact independent when we took A into account, then we could drop the terms involving the interaction of M and S from the model and still get a good fit between the estimated and observed frequencies.[15] The more general point here is that historians, who nearly always have a rich sense of the interactions between, for example, different social groups in specific historical contexts, may be able to formalize and test a variety of original models using log-linear methods. The flexibility of the technique and the multiplication in the number of possible models as the number of variables grows (for example, there are 113 different ones in any analysis of the relationships between four variables) frees historians to exercise their imagination, rather than being constrained, as they often are with such techniques as multiple regression or factor analysis, to confine themselves to testing concepts laid down by statisticians who had other purposes in mind.[16]

Table 8.—A hypothetical example of conditional independence

Panel A: *Apparent Bivariate Relationship*

	Variable A	
Variable B	Level 1	Level 2
Level 1	35%	10%
Level 2	15%	40%

Panel B: *Variables A and B Conditionally Independent, Given Variable C*

	Variable C, Level 1		Variable C, Level 2	
	Variable A		Variable A	
Variable B	Level 1	Level 2	Level 1	Level 2
Level 1	35%	35%	20%	20%
Level 2	15%	15%	30%	30%

In addition to the flexibility they offer, log-linear techniques have more desirable statistical properties than such linear techniques as regression with dummy variables or multiple classification analysis (MCA), and even more than weighted least-squares. As introductory econometrics texts now conventionally warn us, a regression on either continuous or discrete variables of dependent variables which take on only two (or a small number of) values violates the assumption of "homoscedasticity," or equal variances of the errors. Although the resultant estimates are unbiased, the estimated errors do not have the least possible variance, and the usual significance tests should not be used. Since the log-linear model has no such problems, and its associated significance tests are accurate, it is to be preferred over MCA.[17]

The standard solution to homoscedasticity, weighted least squares—known as GSK or Grizzle-Starmer-Koch techniques in the case of discrete variables—eludes this difficulty, but not another serious one. Unlike log-linear estimates, weighted least squares estimates of probabilities are not constrained to lie between zero and one. Thus it is possible to obtain estimates of, say, the probability of being found for men who have certain traits which are greater than one or less than zero. When all the grouped observations are clustered between about 20 percent and 80 percent on the dependent variable, GSK and log-linear techniques yield quite similar results. But since the log-linear estimates are always at least as good as those from weighted regression, and in the cases of many extreme values, the log-linear predictions are better, we see no reason to employ weighted least squares at all.[18]

One final concept should be clarified before we analyze the Boston mobility data: the difference between the particular genre of log-linear methods which we deal with in this paper and "logit" analysis. To put it simply, logit analysis (which can also be applied where the variables are measured on interval scales) designates one variable as dependent, while what we have been referring to as log-linear analysis treats all variables as jointly dependent.[19] In logit, instead of estimating log (p_{ikl}) we estimate log (p_{1kl}/p_{2kl}) which is the ratio between, say, the proportion found and the proportion not found. The relationship between the two models is obviously very close; the logit coefficients are twice those of the relevant lambdas, and the results using the two methods are generally quite similar.[20]

Choosing one over the other is chiefly a matter of habit, taste, or philosophy, and, in fact, we should note that the data analysis for this paper has changed the authors' preferences somewhat. Before beginning it we had unquestioningly accepted the conventional absolute distinction between "independent" and "dependent" variables. Embodying this black-and-white scheme, logit constrains the analyst to consider only those models in which the "dependent" variable is related to every included "independent" variable. For example, if M is considered the "response" variable, then log-linear models such as $\{ASO\}$, or Models 31 and 34–37 in Table 10, below, would make no sense and would never have been estimated.[21]

Yet, as the analysis in section 5 will show, Models 34–37 are interesting ones which fit this particular data set somewhat better than the logit models do. Furthermore, on reflection, we see no reason to divide the world into two disjoint sets, two black boxes labeled "independent" and "dependent" whose separate contents are never scrutinized. Because we are *more* concerned with predicting geographic mobility than with the relationships between A, S, and O, we will refer to M somewhat loosely as the "dependent" variable. There is simply no available terminology to indicate that we wish to emphasize one set of relationships, while not wholly ignoring other sets. But since we do not wish to ignore possible linkages between A, S, and O, we will employ log-linear, rather than logit analysis.

V. Choosing among Log-linear Hypotheses

After all the preparation, we are ready to return to our substantive example. Using Fay and Goodman's ECTA (Everyman's Contingency Table Analysis) program, we examined the relationships between the variables in Table 1. Available directly from Goodman, ECTA is simple and inexpensive to use.[22] Basically, the analyst provides a table—of raw counts, not proportions—and format information about it, uses the fitted marginals notation to specify models to be estimated, indicates the desired level of closeness of fit for models which have to be approximated, and chooses which statistics and tables are to be printed. All these commands may be stated in as few as four lines, not including the table.[23] Other, similar programs are available from other sources.

Interpretation of ECTA's output should begin with the standardized lambdas estimated from the saturated model, some of which are displayed for the Boston data in Table 9. Standardized lambdas are simply the lambdas of Equation (22), defined for the four-variable case and divided by their estimated standard deviations; they are available as an option of ECTA.[24] Since for large samples the standardized lambdas are distributed approximately as standard normal variables, any absolute (that is, either positive or negative) value over about 1.64 is statistically significantly different from zero at the 0.05 level. Absolute values below 1.64 indicate weaker relationships.

The standardized lambdas are useful in deciding which of the many models to test and in determining whether certain categories of particular variables may be consolidated. Of the 119 lambda effects and the one eta effect calculated, we present only the eta and the 31 lambdas which had standardized values of 1.64 or above.[25] The most striking fact about the table (which was of course implied by the very large Chi-Squares for panels D through E of Table 5) is the strong interactions among the independent variables A, S, O. Few men under thirty were fathers yet; few men over thirty weren't (see rows 11–14 of Table 9). Men in high white collar jobs tended to be middle-aged, while the unemployed tended to be teenagers (see rows 17–29). The single-variable effects (rows 2–9) simply reflect the unequal number people in each age and occupational bracket and the fact that more people were "found" than were not. The dearth of significant three- and four-variable lambdas (3 out of 80) is also important, for it indicates that a fairly simple model containing few terms of very high orders will probably suffice.

Finally, and regrettably, it must be noted that only one of the fifty-nine interaction terms involving M, that measuring the relation between age and being found, was significant. In other words, when all the interactions between variables are taken into account, none of the independent variables, or any combination of them, predicted very well whether an 1880 Boston resident

Table 9.—Standardized lambdas for saturated model for data in Table 1.

Row #	Effect of Variable(s)	Level of Variable	Standardized Lambda
1.	η	—	1.780[1]
2.	*M*	—	1.874[2]
3.	*A*	1(14-20)[3]	−3.435
4.		2(21-30)	8.257
5.		3(31-60)	8.391
6.		4(61+)	−6.656
7.	*O*	1(Hi.W.)	−2.657
8.		4(UNSK.)	3.793
9.		5(UNEMP.)	−3.445
10.	*MA*	3(*A*)[4]	2.308
11.	*AS*	1(*A*)[4]	9.168
12.		2	5.351
13.		3	−10.801
14.		4	−6.134
15.	*SO*	1(*O*)[4]	−2.395
16.		5	2.974
17.	*AO*	1(*A*) 1(*O*)[5]	−2.080
18.		1 5	3.429
19.		2 1	−2.253
20.		2 3	2.334
21.		2 4	1.887
22.		2 5	−2.134
23.		3 1	2.201
24.		3 3	3.028
25.		3 4	1.852
26.		3 5	−4.681
27.		4 1	2.650
28.		4 3	−1.795
29.		4 5	1.724
30.	*ASO*	1(*A*) 1(*O*)[5]	−2.195
31.		2 1	2.055
32.		3 2	−1.828

Notes: 1. There is no standard deviation for η. This is the unstandardized effect.
2. Only standardized lambdas above 1.64 in absolute value are listed in the table.
3. Definitions of age and occupation categories in parenthesis.
4. For 2 variable interactions, the levels listed are for the variable in parenthesis. The other variable in the *MA*, *AS*, and *SO* interactions is at level one (found in the case of *M*, and no children in the case of *S*).
5. In the *AO* and *ASO* interactions the levels are of the variables in parenthesis on their right.

would be found in the area in 1890.[26] It should also be noted that, although Goodman has developed a measure for log-linear analysis which somewhat resembles R^2 for regression, there is no index for logit or log-linear analysis which has nearly so appealing an intuitive interpretation as R^2 does for regression or as any of the Goodman-Kruskal "proportionate error reduction" statistics do for simple cross-classification tables.[27]

The principal method for assessing models involving different sets of independent variables is to compute Chi-Square values comparing the actual cell entries, such as those in Table 1, to entries estimated by using a given model.[28] For this purpose, the "Likelihood Ratio Chi-Square statistic," given by

$$(27)\ \chi_L^2 = 2\sum f_i \log (f_i/F_i)$$

is more useful than the "Pearson Chi-Square" statistic defined in Equation (4) above, because χ_L^2 always gives at least as low an estimate as χ_P^2 does, and because χ_L^2, but not χ_P^2, can be "partitioned," in a sense which will be made clear below.[29] We will therefore use χ_L^2 hereafter.

Table 10 presents a series of χ_L^2 values for a large subset (36 of 113) of the possible log-linear models applied to Table 1. The models, all hierarchical, are identified in the fitted marginals notation. (Substantive discussion of each model will be put off until Section VI.) Whereas an analyst using a Chi-Square test in the traditional manner wishes to find a high value of χ^2, since that indicates that the "null model" of no relationship may be rejected, here we wish to find low values of χ^2, because such values indicate that the postulated model yields estimates close to those in the table of actual cell values.

Note first that Model 1, which contains only the grand mean or eta term and which therefore expresses the notion that all cells have an equal proportion of the total number of individuals, fits extremely poorly. By contrast, Model 18, the saturated model, fits perfectly. This will always be the case. Why not, then, stop with this complete and, in a sense, perfect model, accept the view that everything affects everything else, and be done with it? The reason is that in testing log-linear models, we must simultaneously strive for parsimony and completeness. Indeed, for historians, who have a professional predilection for total or "kitchen sink" explanations (better to include every influence, however small, on some outcome than later to be confronted with the criticism that one neglected some factor), this emphasis on parsimony of explanation is one of the chief heuristic virtues of using multivariate statistical methods. Beyond heuristics, log-linear analysis and other multivariate procedures provide statistical criteria to assist us in deciding just where to compromise between parsimony and completeness.

None of the models numbered from 1 to 17, which represent some of the possible poorly fitting lower-order hypotheses, comes close to any reasonable significance level. Included chiefly for illustrative purposes, they may be largely disregarded. But beginning with Model 18, all the subsequent models fit the data adequately at the 0.05 level of significance. How can one choose between them? There are three ways. First, one or more models might encapsulate more coherent theories, or ones more consistent with basic assumptions or with previous studies than other models do. Yet since several, perhaps all, such models in any particular instance

Table 10.—Chi-square values for models based on Table 1.

Model #	Margins Fit		Degrees of Freedom
	Panel A: *Palpably Unsuitable Models*		
1.	Equiprobability	3134	79
2.	$\{M\}$	2989	78
3.	$\{M\}\{A\}$	2180	75
4.	$\{M\}\{A\}\{S\}$	2168	74
5.	$\{M\}\{A\}\{S\}\{O\}$	1872	70
6.	$\{MA\}\{MS\}$	2142	70
7.	$\{MA\}\{MS\}\{MO\}$	1832	62
8.	$\{MSO\}\{A\}$	1566	57
9.	$\{MAO\}$	1340	40
10.	$\{MAO\}\{S\}$	1329	39
11.	$\{MSO\}\{AO\}$	1070	45
12.	$\{MAO\}\{SO\}$	1049	35
13.	$\{MAS\}$	857	64
14.	$\{MAS\}\{O\}$	561	60
15.	$\{MAS\}\{MO\}$	546	56
16.	$\{MSO\}\{AS\}$	279	54
17.	$\{MAS\}\{MSO\}$	260	48
	Panel B: *Statistically Significant, But Easily Rejected Models*		
18.	$\{MASO\}$	0.0	0
19.	$\{MAS\}\{MAO\}\{ASO\}\{MSO\}$	6.53	12
20.	$\{MAS\}\{MAO\}\{ASO\}$	10.26	16
21.	$\{MAS\}\{MAO\}\{MSO\}$	16.12	24
22.	$\{MAS\}\{AO\}$	64.54	48
23.	$\{MAS\}\{AO\}\{SO\}$	47.59	44
24.	$\{MSO\}\{AS\}\{AO\}$	45.67	42
25.	$\{MAO\}\{AS\}$	41.44	36
26.	$\{MAS\}\{ASO\}$	38.12	32
	Panel C: *Final Contenders*		
27.	$\{MAS\}\{MAO\}$	38.08	32
28.	$\{MAS\}\{MSO\}\{AO\}$	29.08	36
29.	$\{MAO\}\{AS\}\{SO\}$	24.49	32
30.	$\{MAS\}\{ASO\}\{MO\}$	21.69	28
31.	$\{ASO\}\{MA\}\{MO\}$	25.92	32
32.	$\{ASO\}\{MA\}\{MO\}\{MS\}$	23.21	31
33.	$\{AS\}\{AO\}\{SO\}\{MA\}\{MO\}\{MS\}$	32.70	43
34.	$\{AS\}\{AO\}\{SO\}\{MA\}\{MO\}$	35.39	44
35.	$\{AS\}\{AO\}\{SO\}\{MA\}$	50.95	48
36.	$\{AS\}\{AO\}\{SO\}\{MO\}$	57.34	47
37.	$\{AS\}\{AO\}\{MA\}\{MO\}$	52.34	48

might plausibly be related to some theory or theories, this criterion may not be much use. Second, one may adopt the convention that one will prefer one model to another if it has either a lower Chi-Square and the same number of degrees of freedom (e.g., Model 29 would be preferred to Models 26 and 27) or a lower Chi-Square and more degrees of freedom (e.g., Model 28 would win out over Models 26 and 27).

The third and, we think, the best method, decomposing Chi-Square, allows us to choose between certain other models as well. In particular, it enables comparisons of "nested" models; that is, those which contain the same lower-level terms but differ by one or more terms at the same or higher levels. This procedure also allows an assessment of the importance of the linkages between specific variables. An example will clarify the notion. Model 23 contains the same terms as Model 22 and $\{SO\}$, the relation between family and occupational status, as well. Model 22 is therefore said to be "nested" within Model 23 or to be a subset of Model 23, and the important of the term $\{SO\}$ may be gauged by taking the difference in the Chi-Squares for the two models and determining whether that difference, which is also distributed as Chi-Square, is significant.[30] The appropriate number of degrees of freedom for the test is the difference in the degrees of freedom for the two models. In this case, the difference in Chi-Square is 16.95 (64.54 − 47.59 = 16.95) and the difference in degrees of freedom is 4 (48 − 44 = 4). A table of the Chi-Square distribution will show that this is highly significant at the 0.05 level. Model 23 is therefore to be preferred over Model 22, and by this test, at least, the linkage between family status and occupation is judged important.

Table 11 gives the results of a series of similar tests and demonstrates how one chooses between models which are generally acceptable. Test number 1, for example, compares Models 19 and 18 from Table 10. The fact that the difference between the Chi-Square for the two models is not significant with 12 degrees of freedom means that the models yield about equally good predictions of the internal cell entries. We therefore choose Model 19 over Model 18 for reasons of parsimony and conclude that the four-variable interaction term is not a necessary part of a satisfactory explanation. Similarly, in the tests numbered 2–9 we reject as unnecessarily complex Models 19, 20, 21, 26, and 27. For tests 10–18 we cannot reject the model containing more terms, since the differences between the Chi-Squares are all significant at the 0.05 level for the appropriate degrees of freedom, but these tests allow us to reject Models 20 and 22 through 26. (Duplication of tests for some models is not strictly necessary, but such tests guard against arithmetic and transcription errors and increase our confidence in the stability of the results.) Panel A of Table 11 thus leaves three models, none of which is nested in either of the others or has the same number of degrees of freedom as the others, still unrejected.[31]

In panel B of Table 11, we first compare the models which survived the panel A comparisons with the models nested in them numbered 31, 33, and 34. Since each of the pairs of models in tests 19–22 is statistically indistinguishable at the 0.05 level, we reject the models containing more higher-order terms and fewer degrees of freedom. Second, we compare Models 31 and 32 with their nest mates 34 and 33 and reject 31 and 32.[32] Third, we compare 33 and 34, again find them statistically similar, and therefore choose the model containing fewer terms, that is, 34. Finally, we present three ways of simplifying Model 34, find that all of them fit the

Table 11.—Assessing the effect of terms in models from Table 10.

			Difference in		
Test #	Model #'s (From Table 10)	Terms Assessed	χ^2_L	Degrees of Freedom	Preferred Model
	Panel A: *Narrowing Down Acceptable Models*				
1.	18,19	{*MASO*}	6.53	12	19
2.	19,20	{*MSO*}	3.73	4	20
3.	20,30	{*MAO*}	11.43	12	30
4.	20,29	{*MAS*}{*ASO*}	18.82	20	29
5.	21,28	{*MAO*}	12.96	12	28
6.	21,27	{*MSO*}	21.96	8	27
7.	26,23	{*ASO*}	9.47	12	23
8.	20,29	{*MAS*}{*ASO*}	14.23	16	29
9.	27,25	{*MAS*}	3.36	4	25
10.	20,26	{*MAO*}{*MO*}	27.86*	16	20
11.	29,25	{*SO*}	16.96*	4	29
12.	28,24	{*MAS*}	16.59*	6	28
13.	23,22	{*SO*}	16.95*	4	23
14.	28,23	{*MSO*}	18.51*	8	28
15.	30,26	{*MO*}	16.43*	4	30
16.	30,22	{*ASO*}{MO}	42.85*	20	30
17.	28,17	{*AO*}	231.0*	12	28
18.	29,12	{*AS*}	1025.0*	3	29
	Panel B: *Choosing The Best Model*				
19.	28,33	{*MAS*}{*MSO*}	3.62	7	33
20.	30,33	{*MAS*}{*ASO*}	11.01	15	33
21.	30,31	{*MAS*}	4.23	4	31
22.	29,34	{*MAO*}	10.90	12	34
23.	31,34	{*ASO*}	9.47	12	34
24.	32,33	{*ASO*}	9.49	12	33
25.	33,34	{*MS*}	2.69	1	34
26.	34,35	{*MO*}	15.56*	4	34
27.	34,36	{*MA*}	21.95*	3	34
28.	34,37	{*SO*}	16.95*	4	34

Note: *indicates a significant difference in the Chi-Squares at the 0.05 level.

data significantly worse than 34, and conclude that 34 is the best we can do.[33]

Table 11 also facilitates the assessment of particular interactions, but a set of tests may lead to ambiguous results. For instance {*MAS*} is evaluated by Test 9 and found unnecessary, but by Test 12 the same term comes out to be significant. Likewise, {*MSO*} is rated unimportant in Tests 2 and 6 but important in Test 14. Tests for some of the interactions, fortunately, are less equivocal. {*SO*}, {*MO*}, {*MA*}, and particularly {*AS*} and {*AO*} are undoubtedly crucial parts of each model.

Another way of ascertaining the importance of various terms and of determining the extent of superiority of one model over another is to divide the differences in Chi-Square between nested models, such as those given in Table 11, by the Chi-Square in the simpler of the two models. Thus the second line in Table 12 shows that when we move from Model 25 to Model 29, which amounts to adding to Model 21 the term {*SO*}, we reduce the Chi-Square by 40.9 percent (16.96/41.44 = 0.409). Although such percentages may be considered counterparts of the coefficients of "multiple determination" and "partial correlation" in multiple correlation and regression analysis, they are not really directly analogous, for they cannot be interpreted as measuring reductions in the percentage of variance explained. These percentages do measure the increase in one's ability to reproduce the original cell entries, but those entries can always be predicted exactly by (saturated) models which may or may not capture causal relationships between the independent and dependent variables at all. We therefore prefer to refer to these measures simply as percentage reductions in Chi-Square.

Table 12 demonstrates that the {*AO*} and {*AS*} interactions are particularly important, by this measure,

Table 12.—Percentage reduction in χ_L^2 due to particular terms

Test # (From Table 11)	Model #'s (From Table 10)	Terms Assessed	% Reduction In χ_L^2
10.	20,26	$\{MAO\}\{MO\}$	73.1
11.	29,25	$\{SO\}$	40.9
12.	28,24	$\{MAS\}$	36.3
13.	23,22	$\{SO\}$	26.3
14.	28,23	$\{MSO\}$	38.9
15.	30,26	$\{MO\}$	43.1
16.	30,22	$\{ASO\}\{MO\}$	66.4
17.	28,17	$\{AO\}$	88.8
5.	21,28	$\{MAO\}$	44.6
7.	26,23	$\{ASO\}$	19.9
9.	27,25	$\{MAS\}$	8.1
18.	29,12	$\{AS\}$	97.7
21.	30,31	$\{MAS\}$	16.3
22.	29,34	$\{MAO\}$	30.7
25.	33,34	$\{MS\}$	7.6
26.	34,35	$\{MO\}$	30.5
27.	34,36	$\{MA\}$	38.3
28.	34,37	$\{SO\}$	32.4

while those of $\{SO\}$, $\{MO\}$, $\{MA\}$, and each of the three-variable interactions are of less consequence, and that including terms such as $\{ASO\}$ or $\{MS\}$ reduces the Chi-Square hardly at all. Depending on which models are compared to each other, the assessments of the importance of particular terms may differ, but these differences are usually small. For instance, including $\{SO\}$ reduces the Chi-Square by 40.9 percent by Test 11, but by only 26.3 percent by Test 13. As this example shows, it is possible to rank the interactions somewhat differently in terms of the percentage reduction in Chi-Square, depending on which model comparison is used. According to Test 11, $\{SO\}$ reduces the Chi-Square by more than $\{MAS\}$ does by Test 12, but by Test 13, $\{SO\}$ reduces Chi-Square less than $\{MAS\}$ does by Test 12. By Test 9, $\{MAS\}$ reduces Chi-Square less than $\{SO\}$ does in either Tests 11 or 12. This observation provides another reason not to rely heavily on the percentage reduction in Chi-Square in evaluating models.

VI. Substantive Conclusions

If the proof of the methodological pudding is in the substantive pie, where do all these numbers get us? In *The Other Bostonians*, Thernstrom found that men in the higher occupational groups were more likely to persist from 1880 to 1890 in Boston than were men on the lower rungs of the ladder. (He did control for age, to some extent, by presenting statistics in the relevant table only for men between nineteen and forty years old in 1880.)[34] He then speculated that nineteenth-century America contained ". . . a permament floating proletariat made up of men ever on the move spatially but rarely winning economic gains as a result of spatial mobility" and suggested that this transiency made it difficult to mobilize the urban masses socially and politically and facilitated social control by the prosperous.[35] These were among the most striking insights in Thernstrom's stimulating and influential book.

The multivariate analysis which we have presented suggests a more complex picture with somewhat different implications than Thernstrom drew.[36] Rather than a floating proletariat, our analysis suggests that the outstanding features of the landscape were youthful mobility, comparatively settled middle age, and the accumulation of human and physical capital over the life cycle. Rather than an irrationally or at least unsuccessfully gyrating whirlpool of movement, our models are consistent with—though they do not, of course, prove that there were—patterned searches for opportunity by rational individuals.

Suppose job chances in each occupation differed somewhat from place to place and that one was trying to decide whether to move or stay put. Consider two men who could, by moving to a particular city, each raise their salaries by the same amount, whose costs of moving were fairly substantial and roughly equal, but who were different ages. Then the younger of the two would be more likely to move, for he could expect to realize the higher salary for a longer time (such things as age-specific health being assumed equal between the two men). With luck, the younger man would much more than make up for the costs of relocating, while a much older man might barely cover those costs before disability or death overtook him.

To make the example a bit more realistic, suppose that we do not assume that each man has perfect information about the present discounted value of the wages in the two places. Then if they were each about equally certain of the geographical wage differential, the younger man would still be more likely to move than the older. For if they guessed either too low or correctly on the wage differential, and both moved, the younger man would enjoy the higher wage in the new area for a longer time; if they overestimated the wage differential, and both moved, the younger man, by moving again, could rectify his mistake, while the older man, pushing into the age when health and other concerns make starting out once more increasingly difficult, would be more likely to be stuck with his bad decision. In short, the young can better afford to take risks because they have more to gain.

They also, on average, have less to lose, both in economic and in social terms. Less likely to own homes and other fixed property, they have fewer transactions costs to bear if they vacate. Likely to have invested less heavily in building up good will with employers, employees, or customers, they can move on without discarding so much of this "capital." Since they have had fewer years to make friendships and, especially if they are single, have a lower probability of belonging to a family unit which has large numbers of friends and relatives in an area, there are also fewer social ties to keep young men in a particular place. Although they were in general less able to make independent economic decisions than men in the nineteenth century, the same considerations would, naturally, apply to women.

A person's stage in the life cycle, furthermore, can be expected to affect his social, as well as his geographical, mobility. Young people almost invariably make some investment in human and/or physical capital. Later in life, they may enjoy the returns from their earlier investments in the form of immediate measurable social mobility by buying shops or by moving from unskilled to skilled or low white collar to high white collar jobs, as well as in the form of consumption by purchasing homes or of intergenerational social mobility by increasing their children's life chances. In cross-sectional data, therefore, we should expect to find disproportionate numbers of youths in lower occupational strata and disproportionate numbers of the middle-aged in higher strata.[37]

These theoretical considerations suggested an analysis of geographic mobility that ought to include age and family status among the independent variables.[38] Our multivariate analyses generally diminish—but do not entirely eliminate—occupational class differentials in persistence, and they underline the importance of age. The preferred model, that numbered 34 in Table 10, includes direct links between age and persistence as well as between occupation and persistence; Tables 11 and 12 especially highlight the interactions between age and occupation and between age and marital status. No model of persistence in Boston which is at all satisfactory, as least among those containing the independent variables we examined, can disregard age as a direct influence, nor can it neglect the interrelationships between age and other independent variables. For example, Model 7 in Table 10, which contains only $\{MA\}$, $\{MS\}$, $\{MO\}$, and lower-order terms, shows a Chi-Square of 1832, an extremely poor fit. And Model 8, which includes all the interactions between mobility, status, and occupation but excludes the direct and indirect effects of age on mobility, also does very badly.

The lambda or effects coefficients for Model 34 from Table 10, displayed in Table 13, further demonstrate these points. Unlike regression coefficients, which measure the impact on a dependent variable of a given change in an independent variable, the lambdas have no simple intuitive interpretation. The essential reason for this is that the relationships in log-linear (or logit or probit) analysis are assumed to be nonlinear and conditional, as opposed to the linear and unconditional effects of the linear regression model. For instance, change due to age in the probability of being found in 1890 varies for each age category, and, within age groups, for each occupational class and family condition. Thus the effect on mobility of being in one's twen-

Table 13.—Lambda or effects coefficients for Model 34 of Table 10.

Variable	Level	Lambda	Standard Error
Panel A: *Single-Variable or Unequal Marginal Effects*			
Eta	—[1]	1.765	—[1]
M	Found	0.225*	0.126
A	(14–20)	−0.840*	0.338
	(21–30)	0.909*	0.140
	(31–60)	0.981*	0.143
	(61 +)	−1.050*	0.189
S	No Kids	0.192	0.126
O	(Hi.W.)	−0.687*	0.410
	(Lo.W.)	0.307)	0.187
	(Sk.)	0.190	0.211
	(Unsk.)	0.779*	0.165
	(Unemp.)	−0.588*	0.205
Panel B: *Interactions of Independent Variables With Dependent Variable*			
MA	(14–20)	0.110	0.338
	(21–30)	0.004	0.140
	(31–60)	0.197**	0.143
	(61 +)	−0.311**	0.189
MO	Hi.W.	0.106	0.410
	Lo.W.	0.056	0.187
	Sk.	−0.036	0.211
	Unsk.	−0.164	0.165
	Unemp.	0.038	0.205
Panel C: *Interactions Among Independent Variables*			
AS	No Kids, 14–20	1.894*	0.338
	No Kids, 21–30	0.465*	0.140
	No Kids, 31–60	−1.244*	0.143
	No Kids, 61 +	−1.115*	0.189
SO	No Kids, Hi.W.	−0.141	0.410
	No Kids, Lo.W.	0.037	0.187
	No Kids, Sk.	−0.187	0.211
	No Kids, Unsk.	−0.197	0.165
	No Kids, Unemp.	0.488*	0.205
AO	14–20, Hi.W.	−1.606**	1.192
	21–30, Hi.W.	−0.224	0.433
	31–60, Hi.W.	0.731*	0.427
	61 +, Hi.W.	1.098*	0.477
	14–20, Lo.W.	0.255	0.459
	21–30, Lo.W.	0.251	0.207
	31–60, Lo.W.	0.055	0.214
	61 +, Lo.W.	−0.561**	0.350
	14–20, Sk.	−0.248	0.505
	21–30, Sk.	0.453*	0.226
	31–60, Sk.	0.535*	0.232
	61 +, Sk.	−0.739*	0.415
	14–20, Unsk.	0.240	0.403
	21–30, Unsk.	0.174	0.182
	31–60, Unsk.	0.080	0.189
	61 +, Unsk.	−0.494**	0.309
	14–20, Unemp.	1.358*	0.488
	21–30, Unemp.	−0.654*	0.280
	31–60, Unemp.	−1.401*	−4.614
	61 +, Unemp.	0.696*	0.313

Notes: 1. The constant or grand mean effect has neither categories nor a standard error associated with it.

*Designates lambdas which are statistically significant at the 0.05 level.

**Designates lambdas which are statistically significant at the 0.10 level.

ties in 1880 differs for the unskilled and high white collar workers, and within the unskilled category, for those with and without children, and each such effect differs for each age group. The lambdas in essence "average out" all the conditional effects for each category of each variable. Rather than transform them and increase both the number of coefficients and perhaps, the reader's confusion, we concentrate on the lambdas alone, considering them simply as measures, without any precise natural meaning, of the relevant effects.

The single-variable or "main" effects in panel A of Table 13 show only that there were, for instance, fewer people in the fourteen to twenty age group than in other age categories, fewer high than low white collar men in 1880, and so on. Included for reasons of completeness only, they may be largely disregarded.

While none of the lambdas in panel B is significant at the 0.05 level, two of the effects for age—but none of those for occupation—are significant at the 0.10 level. Men in their twenties were about as likely to move as to stay in Boston during the 1880s, if their other traits are statistically controlled, while those between thirty and sixty were much more likely to stay than to move, even taking into account their different class and familial situations. The old apparently died.[39] 1880 professionals were somewhat more likely to be found in the Hub City than low white collar workers were ten years later, and men in blue collar occupations in 1880 were even less likely than clerks were to appear in the 1890 city directory. The effects of occupation on mobility, however, are in each case smaller than their associated standard errors, and none even comes close to statistical significance. Coltrolling for age and family status, therefore, almost entirely "washes out" the relationship which Thernstrom stressed between occupation and persistence. Furthermore, if death rates were higher for lower-class than for upper-class men among the older group, and if directory enumerators were more prone to skip lower-class than upper-class men (because poor neighborhoods were more dangerous, because businessmen often advertised in directories, or simply because it was probably harder to locate every individual in a densely populated tenement than along a tree-lined street of single-family houses, then the "true" effects of occupational class on persistence might vanish entirely.

Panel C displays the relationships among the age, family, and occupation variables. The coefficients for $\{AS\}$ and $\{SO\}$ show, not surprisingly, that few of the young or the unemployed had children yet, while those over thirty were generally fathers. It is the twenty $\{AO\}$ coefficients, one for each combination of age and occupation, which draw the most interest. Nearly half of them are statistically significant at the 0.05 level. High white collar men were very likely to be middle-aged or older, while low white collar men tended to be below the age thirty. Skilled workers were overwhelmingly middle-aged, rather than either below twenty or over sixty, while the numbers in the unskilled group varied monotonically and inversely with age. The unemployed were usually either teenaged or elderly. All these cross-sectional relationships suggest that many Bostonians who began in lower strata could expect to move up an occupational notch as their human and physical capital matured, and they reinforce Thernstrom's more general picture of the late-nineteenth-century American labor market as one of limited but real opportunities.[40]

For although he found no impermeable division between classes, Thernstrom did, in effect, describe late-nineteenth-century society as separated into two basic sorts of men. On the one side were striving workers, clerks, and professionals who with a little luck and perserverance could reasonably expect to climb at least a small way up the social ladder or to increase their initial wealth somewhat. On the other side were the evanescent wandering workers, disappearing from the sample, assumed never to find a settled place in society, unions, or politics—a conjectured reserve labor army on the move. By taking age into account in a multivariate analysis of spatial and, to a very limited extent, of social mobility, we have largely eliminated the necessity for postulating the existence of that second class, at least to the extent that it emerged out of the Boston data. Of course there were many mobile Americans; some undoubtedly never found a comfortable niche, and disproportionate numbers of them were probably relatively unskilled and poor. But if the class differences in geographic mobility were principally a product of age differences, if age also correlated strongly with a man's place in the occupational strata in 1880 (both of which we have tried to show) and if many workers, both blue and white collar, progressed upward during their working lives, which Thernstrom showed, then Thernstrom's second class merges with his first, both apparently engaged in rational searches for job opportunities and a good many enjoying some success at it.

Many of these conclusions would have been missed—in fact, were missed—by historians who ignored the interrelationships between independent variables and who were content to use the available data merely to describe bivariate relationships instead of combining appropriate theory with multivariate techniques, such as log-linear ones, to build and test more comprehensive explanations. Having rendered the technique more accessible to historians, we invite them to use it to divise more sophisticated approaches to this and other similar problems in social history.[41]

An earlier version of this paper was presented at the Social Science History Association convention in 1981. We wish to thank this journal's anonymous reader and Stanley Engerman, Douglas Hibbs, Philip Hoffman, Colin Loftin, Douglas Rivers, Stephan Thernstrom, Quang Vuong, Sally Ward, and especially Robert McCaa for comments on various iterations of the piece. Kousser's research was partially supported by grant RO–20225–82 *from the National Endowment for the Humanities. We take responsibility for all remaining errors.*

NOTES

1. Since Stephan Thernstrom launched this area of study with his *Progress and Poverty: Social Mobility in a Nineteenth Century City* (Cambridge, Mass.: Harvard University Press, 1964), the bibliography has become much too long to list here. Some of the leading recent works include Thernstrom's *The Other Bostonians: Poverty and Progress in the American Metropolis, 1880–1970* (Cambridge, Mass.: Harvard University Press, 1973); Michael B. Katz, *The People of Hamilton: Family and Class in a Mid-Nineteenth Century City* (Cambridge, Mass.: Harvard University Press, 1975); Clyde and Sally Griffen, *Natives and Newcomers: The Ordering of Opportunity of Mid-Nineteenth Century Poughkeepsie* (Cambridge, Mass.: Harvard University Press, 1977).

2. If the example which Thernstrom set by depositing his data at the Historical Data Archives of the Inter-University Consortium for Political and Social Research at the University of Michigan could be as widely followed as his path-breaking studies of historical mobility were, the profession would benefit greatly. As we hope to demonstrate, secondary data analyses, too seldom performed by historians, may uncover new facets of the data. We obtained Data Set ICPSR # 7550 from the consortium. Naturally, neither Thernstrom nor the ICPSR bears any responsibility for the analyses we performed.

3. The occupational classifications are Thernstrom's. Age, of course, could be treated as an interval level variable. We cut it into categories only in order to illustrate this particular form of log-linear analysis. The number of cases in the sample was cut from 3,362 to 1,724 by our decision to exclude the thirty-five Negroes, white men for whom any data was missing, and, most importantly, all males under fourteen years of age in 1880. Of our exclusions, 84 percent (1,362 of the 1,628) were because of age. One indication that eliminating cases for which there was missing data did not seriously distort our findings was that the proportion "found" in the smaller sample differed from that in the larger sample by less than 1 percent. The age and status variables were collapsed into four and two categories, repsectively, to simplify the presentation. Log-linear runs on a 100-cell table with status broken into three categories and age into five produced results very similar to those presented below.

4. On this point, see Richard J. Jensen, "Found: Fifty Million Missing Americans," paper delivered at the Social Science History Association Convention, 8 November 1980.

5. We make no claim to originality in our discussion of log-linear analysis. We have merely combined the discussions of other scholars in a way which clarifies the subject, at least to us. We have relied chiefly upon Yvonne M. Bishop, Stephan E. Fienberg, and Paul W. Holland, *Discrete Multivariate Analysis: Theory and Practice* (Cambridge, Mass.: The MIT Press, 1975); Stephen E. Fienberg, *The Analysis of Cross-Classified Categorical Data* (Cambridge, Mass. and London: The MIT Press, 1977); Leo A. Goodman, *Analyzing Qualitative/Categorical Data: Log-Linear Models and Latent Structure Analysis* (Cambridge, Mass.: Abt Books, 1978); H. T. Reynolds, *The Analysis of Cross-Classifications* (New York: The Free Press, 1977); and David Knoke and Peter J. Burke, *Log-Linear Models* (Beverly Hills, Calif.: Sage Publications, Inc., 1980). We shall hereafter cite these books at only a few points, but we acknowledge our general dependence on them.

6. Note that the superscripts do *not* mean that, e.g., τ_i is raised to the *M*th power. We show below that in the multiplicative form, a statement that $\tau_{ik}^{MS} = 1$ is equivalent to saying that the two variables are statistically independent. For a proof that that is not the case for a linear, additive form, such as $F_{ik} = n + \tau_k^S + \tau_{ik}^{MS}$, see Bishop, Fienberg, and Holland, *Discrete Multivariate Analysis*, pp. 23–24. There is a good brief critique of one such technique, AID, in ibid., p. 360.

7. For example, τ_i^M can be divided into τ_1^M, the effect of being found, and τ_2^M, the effect of not being found. The total number of effects is therefore one for the eta, two for the τ^M, two for the τ^S, and four the the τ^{MS} (i.e., τ_{11}^{MS}, τ_{12}^{MS}, τ_{21}^{MS}, and τ_{22}^{MS},), for a total of nine.

8. A change from F_{ik} to p_{ik} does not change the taus, but does require a redefinition of the eta. (Note that $F_{ik} = N_{pik}$, where N is the number of observations.) For details on the redefinition, see Bishop, Fienberg, and Holland, *Discrete Multivariate Analysis*, p. 19.

9. The three parallel lines indicate that the quantities are equal by definition.

10. Note that if we make the same substitutions into Equations (10) or (11), the resulting quantities do not equal unity. Try it.

11. All Equations (18) through (20) really do is illustrate the notion of "closed form" estimates and show that substituting equations (9) to (12) in Equation (8) and simplifying give us the proper identities. In the special case of saturated models, expressions like (9) to (12) are equivalent to the "maximum likelihood" estimates, which may be generated either through a more complex procedure or just through inserting the relevant observed proportions in (9) to (12), or, for cases with larger numbers of variables, analogues of these equations. On the methods used to generate maximum likelihood estimates for cross-classification tables, see Bishop, Fienberg, and Holland, *Discrete Multivariate Analysis*, chapter 3.

12. The interested reader may consult Bishop, Fienberg, and Holland, *Discrete Multivariate Analysis*, chapters 2 and 3.

13. The term "fitted marginals" refers to the fact that for hierarchical models, the maximum likelihood estimates insure that the estimated marginals are equal to the observed marginals.

14. See Otis Dudley Duncan, "How Destination Depends on Origin in the Occupation Mobility Table," *American Journal of Sociology* 84 (1979): 793–803; Leo A. Goodman, "Multiplicative Models for the Analysis of Occupational Mobility Tables and Other Kinds of Cross-Classification Tables," ibid., pp. 804–19.

15. In the actual analysis, below, of Thernstrom's data, only one of the two-variable interactions can be eliminated. Table 8 is *purely* hypothetical.

16. The reader should be able to satisfy himself just by permuting combinations of all four letters that the number of possible models is much larger than the 37 given in Table 10, below. The fact that the total number is 113 for the 4-variable case comes from Bishop, Fienberg, and Holland, *Discrete Multivariate Analysis*, p. 77.

17. On heteroscedasticity, see, e.g., Eric A. Hanushek and John E. Jackson, *Statistical Methods for Social Scientists* (New York: Academic Press, 1977), pp. 141–46. For an application of MCA to nineteenth-century geographical mobility data, see Michael B. Katz, Michael J. Doucet, and Mark J. Stern, "Population Persistence and Early Industrialization in a Canadian City: Hamilton, Ontario, 1851-1871," *Social Science History* 2 (1978): 208-29. From our textual discussion, it is obvious that we disagree with Richard J. Jensen's statement in "New Presses for Old Grapes: I: Multiple Classifica-

tion Analysis," *Historical Methods* 11 (1978): 175-76, that" MCA may now be considered the 'technique of choice' for most problems in Quantitative social history."

18. Compare Michael Swafford, "Three Parametric Techniques for Contingency Table Analysis: A Nontechnical Commentary," *American Sociological Review* 45 (1980): 664-90, with Takeshi Amemiya, "Qualitative Response Models: A Survey," *Journal of Economic Literature* 19 (1981): 1486-87.

19. Logit (as well as probit and Tobit) analysis can be applied to left- and/or right-hand side variables measured on nominal, ordinal, or interval scales, as illustrated, for example, in J. Morgan Kousser, "Making Separate Equal: Integration of Black and White School Funds in Kentucky," *Journal of Interdisciplinary History* 10 (1980): 399–428.

20. Fienberg, *Analysis of Cross-Classified Categorical Data*, chapter 6, contains a good treatment of the relationship between logit and the more general log-linear model. Fienberg refers to the logit model in the text as a "linear logistic response model" reserving the term "logit" for models which predict ratios of raw numbers in the cell entries, instead of proportions in the cell entries. Formally, however, the two models are the same, so we make no distinction here. For some empirical results comparing the two, see Swafford, "Three Parametric Techniques," pp. 664–90.

To see the relation between the logit and the log-linear coefficients, start with Equation (22) in the text. For p_{i11}, the proportion "found,"

$$(22.1)\ \log p_{1kl} = \log \eta + \lambda_1^M + \lambda^S + \lambda^O + \lambda_{1k}^{MS} + \lambda_{1l}^{MO} + \lambda^{SO} + \lambda_{1kl}^{MOS},$$

where the subscripts on every term involving "*M*" indicate that the equation models the effects of the first level of *M* only. Likewise, for the proportion not found,

$$(22.2)\ \log p_{2kl} = \log \eta + \lambda_2^M + \lambda^S + \lambda^O + \lambda_{2k}^{M1S} + \lambda_{21}^{MO} + \lambda^{SO} + \lambda_{2kl}^{MOS}.$$

Since it is a mathematical fact that

$$\log(\frac{a}{b}) = \log a - \log b,$$

and we define the logit of variables *i*, *j*, and *k*, considering the first variable as dependent, as

$$\frac{p_{1jk}}{p_{2jk}},$$

we obtain the equation for that logit by subtracting Equation (22.2) from Equation (22.1). All terms not involving *M* drop out, and we have

$$(22.3)\ \log p_{1jk} - \log p_{2jk} = (\lambda_1^M - \lambda_2^M) + (\lambda_1^{MS} - \lambda_2^{MS}) + (\lambda_1^{MO} - \lambda_2^{MO}) + (\lambda_1^{MOS} - \lambda_2^{MOS}),$$

where the parentheses are inserted for convenience. But by the assumption stated in equation (7) in the test, transformed into logarithms,

$$(22.4)\ \begin{aligned} \lambda_1^M &= -\lambda_2^M \\ \lambda_1^{MS} &= -\lambda_2^{MS} \\ \lambda_1^{MO} &= -\lambda_2^{MO} \\ \lambda_1^{MOS} &= -\lambda_2^{MOS} \end{aligned}$$

Therefore,

$$(22.5)\ \log p_{1jk} - \log p_{2jk} = 2\lambda^M + 2\lambda^{MS} + 2\lambda^{MO} + 2\lambda^{MOS}.$$

And if we rename the logit coefficients, and drop the *M* superscripts and the constant 2's because they appear for every variable on the right-hand side of the equation, we have the logit equation

$$(22.6)\ \log \frac{p_{1jk}}{p_{2jk}} = W + W^S + W^O + W^{OS},$$

and the *W* or logit coefficients are equal to twice the relevant log-linear coefficients. Note also that for a two-category dependent variable expressed in proportions,

$$(22.7)\ p_{2jk} = 1 - p_{1jk}.$$

Therefore, the left-hand side of Equation (22.6) can be expressed as

$$\log \frac{p_{1jk}}{1 - p_{1jk}},$$

and, dropping subscripts, we have an equation which looks very similar to a conventional multiple regression equation:

$$(22.8)\ \log \frac{p}{1-p} = W + W^S + W^O + W^{OS}.$$

21. $\{ASO\}$ does not contain the "dependent" variable *M* at all. Models 31 and 34-37 do not include an $\{MS\}$ terms, although *S* is part of each of the models.

22. Department of Statistics, University of Chicago, Chicago, IL 60637.

23. For tables which have zeroes in any cell, many statisticians advise the user to add some small number, such as 0.50, to each cell. This makes it possible to estimate many models which cannot otherwise be estimated because they contain zeroes in the marginals and also makes convergence go much faster in tables with zeroes in the cells. Furthermore, it reduces the "asymptotic bias" and the "mean squared error" for estimates of the lambdas, which, translated into English, means that if one estimated the coefficients over and over again from similar data or once from an extremely large sample, the results would be that the estimated lambdas were on average closer to the values for the population as a whole if one added a small value to each cell than if one didn't. See Goodman, *Analyzing Qualitative/Categorical Data*, p. 114. We should note that some statisticians do not approve of this procedure and that theoretical and simulation work on it is needed.

24. For the two-variable case, if lambda is written as

$$\sum_{ij} a_{ij} \log f_{ij},$$

where a_{ij} is a constant depending on the number of levels for each variable (if $i = j = 2$, as in Equations (9) to (12) in the text, $a_{ij} = 1/4$), then Goodman has shown that the standard deviation of lambda is the square root of

$$\sum_{ij} a_{ij}^2/f_{ij}$$

for the saturated case, and that this quantity is a lower bound of the standard deviation of each lambda for unsaturated models. See Goodman, *Analyzing Qualitative/Categorical Data*, p. 114, and citations given there.

25. There are separate lambda effects—not all independent of each other—calculated for each level of a multi-category variable. Thus, for instance, there are twenty separate effects (four categories times five categories) for the interactions of age and occupation.

26. Table 5 showed weak but significant Chi-Squares between *M* and *A*, *M* and *S*, and *M* and *O*, taken two at a time. But since the independent variables were related to each other, Table 5 distorts the actual nature of the causal relationships. Table 10, by controlling each of the bivariate relations for the effects of the other variables, gives a more accurate picture of the actual effects on *M* of *S*, *A*, and *O*.

27. No summary statistic based on the idea of "proportionate reduction in errors" (PRE) can be calculated for log-linear models, for all estimation algorithms always fit the marginal relating to the dependent variable exactly. Summary measures based on the PRE concept calculate the number of errors one would make in putting subjects into each class of the dependent variable. For example, if 1,113 people in the Boston subsample were found in 1880 and 611 were not found in 1890, then the best way for an analyst who knew no more about the

people to guess which group each was in would be to put everyone into the "found" category. He would therefore guess wrong 611/1724 = 35.4 percent of the time. PRE measures are based on gauging how much better the analyst would do if he had information about, say, the subjects' ages, occupations, and so on. But since the internal cell estimates in log-linear models are obtained by using information about the marginal cells—for instance, the estimate of the percentage of teenagers found is based on knowing the percentage of all age groups found—the marginals will always be fit as closely as one desires. As a consequence, PRE measures cannot be defined for log-linear models.

28. It is also possible to calculate standardized cell residuals—i.e., to subtract the actual cell entries in Table 1 from those estimated using a particular model and then to divide them by some measure of their variance—in order to assess which cells fit particularly poorly, or, to put it in more substantive terms, which combinations of the independent variables do not predict the dependent variable well. Space limitations prevent us from describing the procedure more fully, but see, e.g., Bishop, Fienberg, and Holland, *Discrete Multivariate Analysis*, pp. 136-55.

29. See ibid., pp. 124-30. χ^2_L is asymptotically—i.e., for very large samples—distributed as χ^2. The partitioning procedure is based on the handy and well-known fact that the sum of two independent χ^2 variates also follows the χ^2 distribution.

30. This method of testing is actually a likelihood ratio test. By contrasting the fit of a given model with the observed frequencies or proportions, χ^2_L in effect compares the fit of some given model with that of the saturated model, since the saturated model fits the observed data exactly. For this and a series of other tests of goodness of fit in what are often referred to in the economics literature as "quantal choice" methods. See Takeshi Amemiya, "Qualitative Response Models," pp. 1502-7.

31. Since every comparison of an acceptable with an unacceptable model—i.e., between any of Models 18-30 with any of Models 1-17 which are nested in them—will show a significant difference, we offer only two of them (Tests 17 and 18). Those tests are included in order to allow us to assess the importance of the terms $\{AO\}$ and $\{AS\}$.

32. In logit, as in the usual multiple regression, all possible relationships between independent variables must be allowed for, but analysts do not usually pay much attention to them. In this case, $\{ASO\}$ would appear in every logit model. That the analyst is not so constrained in log-linear analysis seems to us an advantage.

33. Actually, we tried all five of the models formed by eliminating one two-variable term at a time from Model 34, as well as all ten of the models formed by eliminating two terms at a time. All fail the tests against Model 34 by larger margins than do 35-37.

34. Thernstrom, *Other Bostonians*, Table 3.3, p. 40.

35. Ibid., pp. 42, 231-2.

36. In his "Found: Fifty Million Missing Americans," Jensen has suggested that low-measured rates of persistence were to a large degree artifacts of the sloppiness of census takers, the employees of city directory companies, and transcribers.

37. Thernstrom does not specifically treat the connection between age and social mobility, although he does "control for" age, in a fashion, by confining some of his tables to men between the ages of twenty and thirty-nine. See, e.g., *Other Bostonians*, Table 4.3, p. 53. Some of his tables—e.g., Tables 4.6 and 4.7 on pp. 60 and 61—imply that people did move up the occupational ladder over time from 1880 to 1900.

38. For a review of the "human capital" and other economic approaches to the topic of geographic mobility, see Michael J. Greenwood, "Research on Internal Migration in the United States: A Survey," *Journal of Economic Literature* 13 (1975): 397-433, especially pp. 406-8. On human capital, the (neo)classic starting point is, of course, Gary S. Becker, *Human Capital*, 2nd ed. (Chicago: University of Chicago Press, 1975).

39. Using national figures on age-specific mortality, we ran analyses in every respect parallel to those presented herein, eliminating the estimated proportion of men of each age who could be expected to have died in the decade. Unfortunately, we know of no occupation-specific estimates, but if they were available, they could hardly help strengthening the general points made in the text. The parallel analyses lead to the same model choice and generally track the argument so as to make their presentation needless.

40. *Other Bostonians*, p. 258.

41. There are a great many illustrations of applications of log-linear and related methods in the economics literature listed and cited in Amemiya, "Qualitative Response Models," pp. 1483-84. For an interesting historical example of the use of log-linear techniques to evaluate the validity of mortality statistics, using a capture-recapture model, see Sheryl B. Dow, "The Mortality Rate in Norfolk, Virginia, in 1870: A New Approach to the Application of the Capture-Recapture Method" (unpublished paper, Harvard University, 1981). The problem of the degree to which occupational mobility is merely a product of a shift in the mix of jobs might also be approached by concentrating attention on the relevant single-letter terms in the sorts of models discussed in our paper. In fact, the range of applications seems limited more by analysts' imaginations than by data or computer availability.