Caversham Working Paper, 1997-4

Log-Linear Models of Caversham Occupational Mobility

Hamish James

The Analysis of Categorical Data

Much data of interest in the social sciences is categorical, each case is represented by a series of 'belongs in this category'/'does not belong in this category' distinctions for one or more variables. Often used examples of categorical variables include religion, ethnicity, and political preference; each of these variables can in practice only be collected by assigning individuals to one of a number of exhaustive and mutually exclusive categories (these conditions ensure that each case is assigned to one, but only one, category), Catholic/Protestant; German/Irish/English; Conservative/Liberal. Typically we view and think about this type of data as tables, also called cross-tabulations or cross-classifications. Any number of categorical variables can be cross-classified with one another, the *dimensonality* of the resulting table refers to the number of variables being cross-classified. A variable corresponds to a single dimension of the table and the individual categories that variable is divided into are often referred to as levels. Table 1 is an example of a simple two dimensional table with variable A on the rows of the table and variable B on the columns of the table. Variable A has I levels and variable B has J levels, hence the table is termed an IxJ table. By convention rows are placed before columns in this notation.

Table 1: Table Notation

		B					
		j=1	j=2	j=3	...	j=J	
A	i=1	n11	n12	n13	...	n1J	n1+
	i=2	n21	n22	n23			n2+
	i=3	n31	n32	n33			n3+
	...	...			...		n4+
	i = I	nI1				nIJ	nI+
		n+1	n+2	n+3	n+4	n+J	n++

A three dimensional table of variables A, B, and C is an easy extension of this notation being an IxJxK table where K is the number of levels of variable C, often referred to as the layers of the table. Specific combinations of levels of each variable in a cross-classification are called *cells* in the table and are identified by subscripts, so n_{ijk} refers to the number of observations at the i'th level of variable A, the j'th level of variable B and the k'th level of variable C where i is between 1 to I, j between 1 to J and so on. For example, in a two dimensional table, n_{23} would refer to the second row and third column. The *margins*, of a table are calculated by summing across the levels of one variable while holding the level of all other variables constant. A common notation is to replace the subscript of the variable summed across with a '+' sign, so, in table 1, summing across all the levels of variable B for the second levels of A would give the marginal total n_{2+}. The grand total for a table is calculated by summing across the categories of all variables in the cross-classification, in table 1 this is n_{++} as there are only two variables.

Simple Analysis of Mobility Tables

Data in the social sciences often takes the form of a cross-classification of the same phenomenon for two different times, places, groups of people, or other changes in circumstance. This situation leads to a *square* two dimensional table where variables A and B represent the same phenomenon with variable B being displaced in time, or space, or people affected, or some other factor, from variable A. The *main diagonal* (from top left to bottom right, see Table 1) of a square table contains the series of cells in the table where i = j, that is where the level of variable A is the same as the level of variable B. Cells along the main diagonal contain observations where the category of variable A is the same as the category of variable B, in other words, where there was no change across the change in circumstance between variable A and variable B. Cells off the diagonal represent observations where some change in category was observed.

A mobility table is a square two dimensional cross-classification of the same variable purporting to measure social mobility at two different times. Social mobility is often measured through the use of occupational position. Mobility tables either cross-classify the occupation of fathers at one point in time against the occupations of sons at a later point in time, giving a inter-generational mobility table, or they cross-classify the occupations of the same individual at two points in their career, giving a intra-generational mobility table. In both types of table the variable displayed on the rows can be called the origin and the variable displayed on the columns can be called the destination.

At this point the intra-generation mobility data for Caversham data can be usefully introduced. The data is presented in two tables and relates to the occupational mobility of males within the Caversham Project 9 group occupational classification scheme. Table 2 shows the 498 men found on the 1902 and 1911 electoral rolls covering the suburb of Caversham. Table 3 show the 599 men found on the 1911 and 1922 electoral rolls covering the suburb of Caversham.

Table 2: Male Occupational Mobility in Caversham 1902-1911: Numbers

Count	1911									
1902	1	2	3	4	5	6	7	8	9	Total
1	15			6	2					23
2		13		1			1			15
3			5			1				6
4	2	1		50	1	2	7	2	1	66
5					4		2			6
6	1	5	1	5		60	2	1	3	78
7	3	4	1	13	1	1	140	2	4	169
8		1		1			2	22	2	28
9	2		2	5	1	2	12	3	80	107
Total	23	24	9	81	9	66	166	30	90	498

Table 3: Male Occupational Mobility in Caversham 1911-1922: Numbers

Count	1922									
1911	1	2	3	4	5	6	7	8	9	Total
1	17			1	1					19
2	1	26			1		2		1	31
3			8							8
4	9	1		57		2	8	1	1	79
5					7		2			9
6	2	1		4	2	67			1	77
7	2	4		7	2	5	158	2	9	189
8				1			3	31	9	44
9				4	1	4	6	4	124	143
Total	31	32	8	74	14	78	179	38	145	599

In both tables the majority of individuals are found in cells along the main diagonal, indicating that they remained in the same occupational group across the period. For 1902-1911 78 percent of the cases are on the main diagonal and this increases to 83 percent for 1911-1922. These percentages are calculated by taking n_{ij}/n_{++} for cells where i=j. A crude measure of 'upwards' and 'downwards' (terms which must be used cautiously as the Caversham scheme is not a strictly ordered scale of occupational groupings) is given by calculating n_{ij}/n_{++} for $i > j$ (upward mobility, the cells below the main diagonal) and $i<j$ (downward mobility, the cells above the main diagonal). The results of these calculations are:

Occupational Mobility	1902-1911	1911-1922
Upward	14%	7%
Stable	78%	83%
Downward	8%	11%

Comparing the two periods suggests that Caversham was becoming more stable, with the proportion of downward movers increasing as opportunities for upward mobility decreased.

Percentages can be of further use in a preliminary examination of the Caversham data. By calculating row percentages (n_{ij}/n_{i+}) and column percentages (n_{ij}/n_{+j}) the movement of individuals between occupational groups can be viewed. Column percentages are also called *inflow percentages* as they show the proportion that each origin occupational group contributes to a particular destination occupational group. Row percentages, or *outflow percentages* , show the proportion of a particular origin occupational group which goes to each of the destination occupational groups.

Beginning at the top left of tables 4 and 5 it is interesting to note, in light of the apparent decrease in opportunities for upward mobility, that occupational group 1 was growing more strongly in the second period (from table 5, while 89 percent of those who were in occupational group 1 in 1911 remained in 1922, by this time they made up only 55 percent of the group total). In contrast, occupational groups 2 and 3 display growth during 1902-1911 which is not continued in the later period. In the earlier period new recruits to the elite occupations enter occupational groups 2 and 3, while in the later period they enter occupational group 1. There is virtually no movement between these three occupational groups suggesting that while they might be viewed as the elite of Caversham they do not form an internally homogenous group and a variety of career pathways may be involved. Those upwardly mobile into this elite occupational strata come from occupational groups 4, 6, and 7. Those downwardly mobile depart from occupational group 1 for occupational groups 4 and 6, and from occupational group 2 to groups 4, 6, and 7.

Table 4: Male Occupational Mobility in Caversham 1902-1911: Inflow and Outflow Percentages

Percentage	1911									
1902	1	2	3	4	5	6	7	8	9	Total
Inflow Percent										
1	65			7	22					5
2		54		1			1			3
3			56			2				1
4	9	4		62	11	3	4	7	1	13
5					44		1			1
6	4	21	11	6		91	1	3	3	16
7	13	17	11	16	11	2	84	7	4	34
8		4		1			1	73	2	6
9	9		22	6	11	3	7	10	89	21
Total	100	100	100	100	100	100	100	100	100	100
Outflow Percent										
1	65			26	9					100
2		87		7			7			100
3			83			17				100
4	3	2		76	2	3	11	3	2	100
5					67		33			100
6	1	6	1	6		77	3	1	4	100
7	2	2	1	8	1	1	83	1	2	100
8		4		4			7	79	7	100
9	2		2	5	1	2	11	3	75	100
Total	5	5	2	16	2	13	33	6	18	100

Occupational groups 4, 6, and 7 have the widest spread of movement, recruiting from and departing to most other occupational groups. This suggests that these groups may be central to the occupational pathways within the tables, acting as the 'reservoir' for members of the employer and professional classes. One kink in this pattern is the non-symmetrical relationship between employers and the skilled. Perhaps those who rise from ranks of skilled workers to become employers do not fail dramatically but rather slide back into the small-employer category. This pattern is investigated further below, using the more sophisticated techniques of log-linear modelling. It is also worth noting that the small size of the semi-professional and petty-official groups prevents much analysis of these groups.

The manual groups, the skilled, semi-skilled, and unskilled form a block with relatively free internal movement. In the earlier period the semi-skilled and unskilled depart to, between them, all the other occupational groups, but this upward mobility has been reduced by the second period with no members of these groups rising further than occupational group 4, perhaps reflecting an increasingly static occupational structure. Given their apparent links to the non-manual occupational groups the skilled could be proposed as a bridging class between the employer/small employer and professional groups and the manual block. This idea will also be investigate using log-linear modelling, a technique which is discussed in the next section.

Table 5: Male Occupational Mobility in Caversham 1911-1922: Inflow and Outflow Percentages

Percentage	1922									
1911	1	2	3	4	5	6	7	8	9	Total
Inflow Percent										
1	55			1	7					3
2	3	81			7		1		1	5
3			100							1
4	29	3		77		3	4	3	1	13
5					50		1			2
6	6	3		5	14	86			1	13
7	6	13		9	14	6	88	5	6	32
8				1			2	82	6	7
9				5	7	5	3	11	86	24
Total	100	100	100	100	100	100	100	100	100	100
Outflow Percent										
1	89			5	5					100
2	3	84			3		6		3	100
3			100							100
4	11	1		72		3	10	1	1	100
5					78		22			100
6	3	1		5	3	87			1	100
7	1	2		4	1	3	84	1	5	100
8				2			7	70	20	100
9				3	1	3	4	3	87	100
Total	5	5	1	12	2	13	30	6	24	100

Log-Linear Modelling

Cross-classifications are constructed to investigate the nature and strength of relationships between variables - does variable A influence variable B and if so, how strongly does variable A influence variable B? The statistical approach taken to answer such questions is to develop models which describe the processes the investigator believes underlie the observed data. These models are used to calculate *expected frequencies*, the frequencies that would be expected to occur in each cell if the model was an accurate explanation of the observed data. Once expected frequencies have been calculated a comparison can be made between the observed frequencies and the expected frequencies to gauge the model's *goodness of fit* (how closely the expected frequencies match the observed frequencies). If the expected and observed frequencies are the same then the fit is perfect, the more they differ the poorer the fit is. A poor fit suggests that the model's explanation of the underlying processes is not adequate because the data we would expect to see if it were correct is different to the data we have collected. A statistical test, a formal way of defining the *significance* of the comparison between the expected and observed frequencies is used to ensure that judgements about the adequacy of the fit of different models are consistent. By calculating the value of a test statistic that in some way captures the comparison between the expected and observed frequencies and comparing this against values of this statistic given by a theoretical probability distribution (which, in essence, tells us how likely a specific value of the test statistic is) we can decide if the difference between the model and the observed data is significant (unlikely to occur by chance alone), indicating a model which does not explain the observed frequencies well, or insignificant, differences that can reasonably be ascribed to chance variation[1].

The statistical techniques and procedures referred to here as log-linear modelling are one modelling approach suitable for categorical data. Log-linear modelling was developed in the 1960s and 1970s and adopted widely by sociologists interested in social mobility during the 1970s and 1980s. Log-linear modelling provides a powerful multi-variate analytical technique for categorical data:

> Besides aiding the search for meaningful relationships, log-linear models are in many respects similar to well-established statistical procedures such as analysis of variance and regression analysis. the new technique seems to put the investigation of nominal [categorical] data on a par with study of interval-level variables.[2]

The timing of these developments along with the sociological focus of the work meant that the quantitative social history of the 1970s did not take advantage of these new techniques. With the subsequent decline of interest in historical social mobility there has been little historical work done using log-linear modelling, despite the fact that log-linear modelling offers a way to solve some of the statistical short-comings which contributed to the decline (Grusky and Fukumoto, 1989).

A log-linear model consists of a number of *parameters* which are combined together to generate an estimated value for each cell in the cross-classification under examination. How the parameters are combined and what the estimated value represents depend on the way in which the model has been expressed. Usually social scientists express log-linear models in terms of expected frequencies, in which case parameter are multiplied together, or in terms of natural logarithms of the expected frequencies, in which case parameters are added together. Log-linear modelling is so called because the equation of a model in this form calculates the logarithm of the expected values from a linear combination of parameters.

Written in the logged form the elementary model for an IxJ table is written

$$L_{ij} = a_0 \qquad [1]$$

where L_{ij} is the natural logarithm of the expect frequency for cell (i,j) and a_0 is the natural logarithm of the *grand mean* . The grand mean for a log-linear model is a geometric mean (the *nth* root of the product of *n* frequencies) rather than the everyday arithmetic mean (the sum of *n* cell frequencies divided by *n*.). Alternatively this model could be written as:

$$F_{ij} = A_0 \qquad [2]$$

where F is the expected frequency for cell (i,j) and A_0 is the grand mean. These two ways of expressing the same model are related by $L_{ij} = \ln(F_{ij})$, or the reverse, $F_{ij} = e^{L_{ij}}$, where ln is the natural logarithm and *e* is the base of the natural logarithm. Equations [1] and [2] are equivalent ways of expressing the same model. The model states that all expected frequencies are equal to a constant value, given by the value of the grand mean. There are no underlying processes occurring so with all things being equal we would expect the same value to occur in all cells in the cross-classification.

---

[1] A good explanation of the conceptual basis of statistical modelling is provided by Gilbert (1981).

Obviously this is not a particularly useful model. To build models that might explain the processes behind the observed data more parameters must be added to the model. These additional parameters are deviations from the grand mean, adding to or subtracting from the base value given by the grand mean. For each variable in the cross-classification there is a *main effect* parameter. Including a main effect parameter in a model allows for the distribution of the marginal frequencies across a variable to be skewed or uneven, that is, the expected value in each cell is not the same. Using the grand mean plus main effects for a two dimensional table gives us the familiar model of statistical independence, the model which underlies the chi-squared test for independence:

$$F_{ij} = A_0 A_{1i} A_{2j} \qquad [3]$$

or

$$L_{ij} = a_0 + a_{1i} + a_{2j} \qquad [4]$$

The expected value is based on the grand mean and main effects for each variable, variable 1 with i categories, a_{1i} and variable 2 with j categories, a_{2j}. The main effects allow for different numbers of observations to appear in each category of a variable. The model can be interpreted as saying that we do not think the two variables are inter-related, but we do believe that specific levels within each variable may have different likelihoods of occurring. What the model does not so far include are *interaction effects* which allow for the expected values to be influenced by specific combinations of levels on each variable in the cross-classification. An interaction effect is a way of saying that we think that the expected value for cell (i,j) may be different to that of cell (i-1,j-1), or cell (i-3,i-4), or any other combination of categories from variable 1 and 2. If we include interaction effects for our two variable example the new model can be written:

$$L_{ij} = a_0 + a_{1i} + a_{2j} + b_{ij} \qquad [5]$$

The interaction effect, b_{ij}, means that each combination of level i on variable 1 and category j on variable 2 now contributes to L_{ij}. Because there are only two variables in this example, [5] is a *saturated model* ; one which includes all the possible effects, in this case we have the grand mean effect, a_0, plus separate allowance for the levels of each variable, a_{1i} and a_{2j}, and an interaction effect which allows for the different combinations of levels on variable 1 and variable 2. If we had three variables then the saturated model would be

$$L_{ijk} = a_0 + a_{1i} + a_{2j} + a_{3k} + b_{ij} + b_{ik} + b_{jk} + b_{ijk} \qquad [6]$$

which has three main effects, one for each variable, three two way interactions and a single three way interaction. There is a main effect for each variable in the cross-classification and an interaction effect for each possible combination of a variable with one or more of the other variables.

A saturated model will fit the observed data perfectly, the expected and observed values will be the same. A saturated model is not, however, a good explanatory model as to achieve this perfect fit every possible effect is deemed relevant. In doing this the relative importance of different effects is obscured and statements cannot be made about which effects are important and which are unimportant. The objective of the log-linear modelling process is not just to find the best fitting model, as by definition this is the saturated model, but to find a model which is both simple, includes only a few of the possible effects, and has a good fit. If a good fit can be maintained while the number of effects is reduced then the effects which remain in the model must become more significant as individually they explain a greater proportion of the pattern of the observed data.

Based on this concept the methodology for using log-linear models in research can be described as:

a) Use topic-specific knowledge and theory to propose models to test against the observed data.

b) Calculate expected values for these models and compare them to the observed values using a goodness of fit statistic, discarding models which fit poorly.

c) Compare those models with reasonable fits by testing the differences in their goodness of fit for statistical significance. Using these tests of significance, in conjunction with expert knowledge, determine which gives the best explanation of the observed data, seeking a simpler model over a more complex one.

This methodological outline is similar to that common to many forms of statistical model building.[3] The statistical techniques should be used to order and control the analysis, but models which fit well should be discarded if they do not have a sensible explanation, equally models that do not fit that well may be retained if they have a strong explanation.

Technical Background to the Log-linear Model[4]

A log-linear model can be described using the notation introduced in the previous section, but to be useful the description must be interpreted in some way to produce expected values for the model to compare against the observed data. Expected frequencies for log-linear models are calculated using the methods of *maximum likelihood estimation*. Maximum likelihood estimates have the useful characteristic that they can be calculated solely from marginal totals. For many statistical models the expected values are calculated using a formula, such as the one used in the chi-squared test for independence, $F_{ij} = n_{i+}n_{+j}/n_{++}$, which is also a maximum likelihood estimate. Similar formulas exist for some, but not all, log-linear models. Expected frequencies for [3] can be calculated using $F_{ij} = n_{i+}n_{+j}/n_{++}$ as it is the same model as that used in the chi-squared test. Where a formula does not exist the useful ability to calculate maximum likelihood estimates using only marginal totals enables us to calculate the expected frequencies by an *indirect*, as opposed to *direct* formula based method. This is done using a procedure called *iterative proportional scaling*[5] that Gilbert likens to mixing plaster:

> One way to mix the kind of plaster which is used for filling cracks in walls is to obey the instructions on the packet that specify how much waters is to be added to how much powder [the instructions provide a formula for mixing the plaster]. But the way most people do it is to add a little water to some powder, mix them together and test to see whether the result has the right consistency. If it is too dry, some more water is added, or if it is too wet, some more powder Then the mixture is stirred

[3] Kousser, Cox and Galeson (1982) demonstrate the methodology used to compare different log-linear models in a historical research context. Van Leeuwen & Maas (1997) is a recent example of the use of log-linear models in historical research.
[4] For alternative simple introductions see: Knoke & Burke, 1980; Gilbert, 1981; Reynolds, 1977a; Jones & Davis, 1986. For full statistical introductions see: Reynolds 1977b; Bishop, Fienberg, and Holland, 1975.
[5] Another, more generally applicable method, called the *Newton-Raphson algorithm* can also be used,

and tested again. This goes on until the balance of powder to water seems right.[6]

A set of expected frequencies calculated using maximum likelihood estimation will have the same marginal totals, for those effects included in the model, as the observed data. Using iterative proportional scaling, we begin by assigning *starting values* to each cells expected frequency. The starting values are an arbitrary guess at the expected frequency. Next one set of marginal totals calculated from these initial expected values are compared to the corresponding marginal totals for the observed data. If the observed and expected marginal totals are different then the expected values are adjusted proportionately to the observed values, so that the expected and observed marginal totals will match. This procedures continues for each different set of marginal totals in the model until they all match the corresponding observed marginal totals. Where direct and indirect methods are available for a model both will arrive at the same set of expected frequencies. It is worth looking at a simple example of direct estimation in order to gain some insight into what the parameters of a model mean.

There are three elements to a log-linear model as written in the previous section: a grand mean; none, one, or more main effects and; none, one, or more interaction effects. Consider a simple log-linear model, the saturated model for a 2x2 cross-classification, expressed in terms of expected frequencies:

$$F_{ij} = A_0 A_{1i} A_{2j} B_{ij} \quad [7]$$

where B_{ij} is the interaction effect expressed in terms of expected frequencies. As discussed earlier the grand mean, A_0, is the geometric mean.:

$$A_0 = (n_{11} n_{12} n_{21} n_{22})^{1/4} \quad [8]$$

in terms of natural logarithms [8] can be expressed as,

$$a_0 = \ln(A_0) = 1/4(\ln(n_{11}) + \ln(n_{12}) + \ln(n_{21}) + \ln(n_{22})) \quad [9]$$

which is the arithmetic mean of the natural logarithms of the cell frequencies.

The main effects for the 2x2 saturated model are a geometric mean of the values of one category compared to another, as with [9] when the geometric mean is converted into logarithmic form it becomes an arithmetic mean, for example, the main effect of variable 1, for level 1 is:

$$A_{11} = (n_{11} n_{12} / n_{21} n_{22})^{1/4} \quad [10]$$

and

$$a_{11} = 1/4(\ln(n_{11}) + \ln(n_{12}) - \ln(n_{21}) - \ln(_{22}) \quad [11]$$

The other three main effects, a_{12}, a_{21}, and a_{22} can be calculated in the same manner by adjusting the order of the cells in [11]. Do not confuse the fact that there is only one main effect per variable with the fact that each level of a variable may have a different value of the main effect for the variable.

Both the main and interaction effects are functions of the *odds-ratio*. Before explaining how the interaction effects are calculated it is necessary to examine some of the properties of the odds-ratio, which is a measure of association for 2x2 tables that can be

6

expanded for IxJ tables by calculating the odds-ratios of 2x2 sub-tables within an IxJ table. An illustrative example of the odds-ratio is provided by collapsing Table 2 into Table 6, a 2x2 table of non-manual and manual occupations.

Table 6: Manual and Non-manual workers in Caversham

		Occupation in 1911	
		non-manual	manual
Occupation	non-manual	175	19
in 1902	manul	37	267

The odds of being in cell n_{11} as opposed to cell n_{21} are 175/37, or 4.73, this can be read as; a non-manual worker in 1911 is 4.73 times more likely to have been a non-manual, rather than a manual, worker in 1902. Similarly, the odds of being in cell n_{12} as opposed to n_{22} are 19/267 which equals 0.07, the odds of a 1911 manual worker having been a non-manual worker in 1902. The ratio of these odds,

$$\frac{n_{11}/n_{21}}{n_{12}/n_{22}} \qquad [12]$$

which can be rearranged as

$$\frac{n_{11}\, n_{22}}{n_{12}\, n_{21}} \qquad [13]$$

is the odds ratio, …. For this example … is 66.47. The odds-ratio varies from 0 to positive infinity, with 1 indicating no association (as both odds must have been the same for the division to be equal to 1). For nominal (unordered categories) data values greater than or less than one both measure association, but on different scales. The interval 0 to 1 represents the same range of strength of association as the interval 1 - positive infinity does. To see this, start by working out the two odds above in terms 1902 workers instead of 1911 workers. The odds are now n_{21}/n_{11}, 0.21, and n_{22}/n_{12}, 14.05, and the ratio of these odds is:

$$\frac{n_{21}/n_{11}}{n_{22}/n_{12}} \qquad [14]$$

which can be rearranged as

$$\frac{n_{21}\, n_{12}}{n_{11}\, n_{22}} \qquad [15]$$

Equation [15] is the reciprocal of [14], which is equal to 1/66.47, or 0.015. So, the odds-ratios 66.47 expresses the same strength of association as the odds-ratio 0.015. This situation can be improved somewhat by taking the natural logarithm of the odds-ratio, called the *log-odds* , with the symbol …*, which ranges from positive to negative infinity, with zero indicating no-association and the strength of association being the same for the same absolute values of the log-odds.

The odds-ratio is based on a 2x2 table, so there are four ways of calculating the odds-ratio, each one starting with a different cell in the table. To differentiate between these four odds-ratios we use subscripts. The first odds-ratio we calculated was $\ldots_{11}$, as cell (1,1) is the first cell in that odds-ratio, similar the second was $\ldots_{21}$. The four ways of calculating the odds-ratio are related:

$$\ldots = \ldots_{11} = \ldots_{22} = 1/\ldots_{12} = 1/\ldots_{21} \qquad [16]$$

$$\ldots^{*} = \ldots^{*}_{11} = \ldots^{*}_{22} = -\ldots^{*}_{12} = -\ldots^{*}_{21} \qquad [17]$$

We can compare the non-manual/manual association for 1902-1911 with 1911-1922. For 1902-1911 $\ldots = 66.47$ and $\ldots^{*} = 4.20$. For 1911-1922 $\ldots = 149.21$ and $\ldots^{*} = 5.01$. The strength of association is greater in the later period, not surprising as we already know that mobility declined. Unfortunately, the odds-ratio lacks a clear interpretation so we cannot say by how much the association has increased. Nevertheless the odds-ratio does have several useful characteristics. As we have just seen, the ordering of categories does not affect the result, and also, changing the margins of a table by multiplying the rows and columns does not affect the odds-ratio, unlike other measures of association[7] This second characteristic is regarded as very useful in the social sciences.

Comparing the form of [10] with that of [13] it should be apparent that main effects are a function of the odds-ratio. The interaction effects are also a function of the odds-ratio. To see this we return to the 2x2 saturated model. There are four cells in a 2x2 table, so if we replace the subscripts i and j with specific cells we obtain four equations for the expected frequencies:

$$\begin{aligned} F_{11} &= A_0 A_{11} A_{21} B_{11} \\ F_{12} &= A_0 A_{11} A_{22} B_{12} \\ F_{21} &= A_0 A_{12} A_{21} B_{21} \\ F_{22} &= A_0 A_{12} A_{22} B_{22} \end{aligned} \qquad [18]$$

When the log-linear model is written this way it is clear that the equations for the four expected frequencies could be substituted into [11], the equation of the odds-ratio:

$$\ldots_{11} = A_0 A_{11} A_{21} B_{11} A_0 A_{12} A_{22} B_{22} \,/\, A_0 A_{11} A_{22} B_{12} A_0 A_{12} A_{21} B_{21} \qquad [19]$$

The terms in [19] simplify leaving:

$$\ldots_{11} = B_{11} B_{22} \,/\, B_{12} B_{21} \qquad [20]$$

Equation [20] shows clearly that the interaction effects are linked to the odds-ratio. With further manipulation[8] it turns out that the interaction effect for cell (1,1) is

$$B_{11} = (\ldots_{11})^{1/4} \qquad [21]$$

The interaction effects are the fourth root of the odds-ratio, in logarithmic form (refer to [17]) they are,

$$b_{11} = 1/4(\ldots^{*})$$
$$b_{22} = 1/4(\ldots^{*})$$

---

[7] See Reynolds (1977) ch. 2.

$$b_{12} = 1/4(-\ldots^*)$$
$$b_{21} = 1/4(-\ldots^*) \qquad [22]$$

It may be apparent by now that the parameters in a log linear model have no intuitive interpretation. It is a mistake to assume simple interpretations of the actual parameters of the model and if unsure, they are best left alone.

Goodness of Fit and Selecting Models

Once we have calculated expect frequencies the next step, as discussed above, is to compare them with the observed frequencies, using a statistical test to aid our judgement of whether or not the model fits the observed data well. The commonly known Pearson chi-squared statistic (X^2) can be used to judge the goodness of fit of log-linear models, but the likelihood ratio chi-squared statistic (L^2) is preferred to X^2 because it can be partitioned[9], a term explained in the next section. Either statistic provides a measure of the difference between the expected and observed frequencies, and the probability of a particular value of either statistic occurring should be approximately the same as the probability of that value occurring in the theoretical chi-squared probability distribution.

To determine the probability from the chi-squared distribution a second value, the *degrees of freedom* (df) is needed for a model. Degrees of freedom are a measure of how flexible a model can be in calculating expected frequencies. Fewer degrees of freedom mean that there are more constraints on the values that the expected frequencies can take. A complex model will impose more constraints and therefore have fewer degrees of freedom. Because we are seeking simple models this leads to the general principle that a model with more degrees of freedom is preferable to one with fewer degrees of freedom, if both fit the data.

The effects for each cell are interrelated with each other. Because both main and interaction effects are deviations from the grand mean they will, in logarithmic form, sum to zero across the levels of each variable (in terms of expected frequencies the product should be equal to one):

$$\sum_{i=1}^{I} a_{1i} = 0 \qquad [23]$$

which means that the sum of the values a_{1i}, where i is a value starting at 1 and increasing in steps of 1 to the value I, is equal to zero. Similarly,

$$\sum_{j=1}^{J} a_{2j} = 0 \qquad [24]$$

and for the interaction effects,

$$\sum_{i=1}^{I} \sum_{j=1}^{J} b_{ij} = 0 \qquad [25]$$

The constraints, [23], [24], and [25] indicate that not all the parameters in a log-linear model are *independent.* For example, in the saturated model for a 2x2 table, if we know

9

b_{11}, b_{12}, and b_{21} then we also know b_{22} as $b_{11} + b_{12} + b_{21} + b_{22}$ must equal zero. We do not need to calculate effects for every cell in order to completely calculate the model. The basic number of degrees of freedom for a log linear model is the number of independent effects.

The L^2 in the theoretical chi-squared probability distribution for the models degrees of freedom is found and the probability of a value larger than that value occurring is then used to judge the fit of the model, a large probability indicates a better fit.

It is possible for many different models to fit the data well, in which case we test the difference between the L^2 values from two models to see if it is significant or not. Remembering that a simpler model is preferable to a more complex one, if the difference is not significant we discard the more complex of the two models as it does not improve the goodness of fit. The explanatory power of the model should always be kept in mind when choosing between alternative models.

Log-Linear Modelling for Mobility Tables

Usually, log-linear modelling is applied to the problem of seeking a model, in terms of main and interaction effects, which is simple but also fits a complex cross-classification of many variables. When mobility data is cross-classified against other variables such as age, ethnicity and education, this approach to log-linear modelling can be used in social mobility research, but when mobility data consists of a simple two dimensional table it is of little use. With only two main effects and one interaction effect there are very few models available to investigate, and as it is to be expected that the variables will be related the only model likely to fit well is the saturated model, [5]. A more informative way to apply log-linear models to mobility tables is to construct a more elaborate interaction effect to replace that used in [5].

The starting point for these mobility log-linear models are the log-linear models [4] and [5]. When applied to mobility data [4], the model of statistical independence is called the perfect mobility model because the lack of an interaction effect implies that the occupational group of origin has no effect on the occupational group of destination, or visa-versa. If perfect mobility holds there will be no barriers to individuals changing between occupational groups at will. Main effects can be interpreted as measuring the impact of changes in the underlying demographic, economic and social structure of the population, it is the interaction effect which measures relative mobility between occupations.

Unsurprisingly, the model of perfect mobility does not fit the Caversham data for either period.

1902-1911			
L^2	989.08234	DF = 64	P = .000
1911-1922			
L^2	1395.20568	DF = 64	P = .000

The nature of occupational mobility virtually rules out [4] fitting observed data well. On the other hand [5], the saturated model, fits the observed data exactly as it includes all the possible parameters. Remember that the aim of log-linear model building is to find a simple model that fits the data well; the first model is too simple and the second is too complex. The useful models of mobility lie between these two extremes. An obvious way to understand these in-between models is as mixtures of the characteristics of the two extreme cases [4] and [5]. The mixing of the characteristics of [4] and [5] is achieved by *partitioning* the mobility table. Main effects which measure underlying factors are applied to all the cells in the table, but the table is partitioned into a number subsets of cells, each of which is given a different interaction effect. The model of mobility is described by what type of interaction we assign to each partition and how we divide the cells of the table into different partitions.

Quasi-perfect mobility (QPM) is a simple model that illustrates these ideas. QPM proposes that a certain proportion of individuals will remain in their original occupations and only the remainder of individuals will experience perfect mobility between occupations. In model terms we divide the mobility table into two partitions, one containing the cells off the main diagonal, the other containing the cells on the main diagonal. The first partition contains all cells representing mobility from one occupational group to another, the model of perfect mobility is applied to this partition, so no interaction effect is assigned. The second partition contains cells representing immobility, people who remain in the same occupational group. An interaction effect is added for this partition implying that, unlike the off diagonal cells, your occupational group of origin will influence your occupational group of destination.

$$L_{ij} = a_0 + a_{1i} + a_{2j} \qquad \text{where } i \neq j \text{ (off diagonal cells)} \qquad [26]$$

$$L_{ij} = a_0 + a_{1i} + a_{2j} + b_{ij} \qquad \text{where } i = j \text{ (main diagonal cells)} \qquad [27]$$

Equations [26] and [27] show that the QPM model is built by applying the model of perfect mobility to one part of the table and the saturated model to the other part. Procedurally, QPM is tested by ignoring the cells along the main diagonal while fitting the model of perfect mobility to the rest of the table. This is done by defining the diagonal cells as *structural zeros*. A structural zero is a logically inconsistent cell, one for which a frequency of greater than zero is impossible. No expected frequency needs to be calculated for cells defined as structural zeros. Structural zero's differ from *sampling zeros* which are merely cells that have an observed frequency of zero due to the size of the sample.

Results from the model of QPM for the Caversham data are:

1902-1911						
	L^2	60.34077	DF	55	P	0.29
1911-1922						
	L^2	79.51103	DF	55	P	0.02

Note the large reduction in the size of the L^2 statistic, indicating that although QPM does not fit well, it does fit considerably better than perfect mobility. Most of the individuals in the Caversham data are immobile, so QPM's hypothesis of occupational inheritance for the immobile is a reasonable one. This suggests that further models designed to improve the fit should seek to incorporate QPM's separation of cells on the main diagonal from cells off the diagonal. A straight forward extension of QPM is the 'up down' model (there seems to be no commonly used name for this model). The up down model breaks the table into three partitions; immobility (i=j), downwardly mobile ($i < j$), and upwardly mobile ($i > j$). Results for this model are:

1902-1911						
	L^2	35.61245	DF	21	P	0.02
1911-1922						
	L^2	35.68515	DF	21	P	0.02

The results suggest the up down model does not add to the explanatory power of QPM. We can formally compare the two models to judge if the difference in fit between them is significant by testing the difference in L^2 values with degrees of freedom equal to the difference between each models degrees of freedom. For 1902-1911 this gives L^2 = 24.72832 with df = 34, which is not significant. For 1911-1922 the difference between models is significance at the 5 percent level. This suggests that there may be some non-symmetrical mobility in the table or that the amount of 'upward' mobility differs from the amount of 'downward' mobility.

At this point we turn to the structural (also called topological) model introduced by Hauser (1978). The structural model allows a more flexible partitioning of a table by assigning each cell to one of a number of mutually exclusive and exhaustive subsets of the mobility table. Each subset is called an interaction level, and all cells in the same level have a common interaction parameter. The model can be understood as one of statistical independence for three variables; origins, destinations and interaction levels. Each interaction level represents a different density, or amount, of mobility. In this way the table is divided up into 'regions of (net) association, or interaction, as between origins and destinations categories'.[10]

Before continuing, it is important to sound a warning that,

[10]

difficulties are likely to occur in interpreting, and in choosing between, topological models in so far as the allocation of cells to different interactions levels is determined *ad hoc*, with the aim simply of fitting the data, rather than being guided by some theoretical rationale.[11]

In our efforts to build a good-fitting model we may end up responding to the observed data, introducing interaction levels and assigning cells to particular levels on the basis of the goodness of fit statistics rather than *a priori* on the basis of theory about the mobility pattern under examination. Letting the data guide the construction of theory is also problematic because more than one structural model, each with differing interpretations, can generate exactly the same expected values.[12] Because of the exploratory nature of this analysis these dangers are particularly acute here as the preliminary analysis of raw numbers and percentages above is being used to guide the formulation of the structural models below.

A further complication which deserves mention is similar to the problem encountered in inter-generational mobility tables where it is difficult to be sure that all fathers are at one, and all sons at another, stage in their careers. If this is not so then the aggregation of father-son pairs into a mobility table is not comparing like with like. The Caversham data used here is intra-generational and is divided into two decade periods, each of which is assumed to be independent of the other. This assumption is strained by the 224 men, constituting roughly a third of the cases in each table, who remain in Caversham throughout the 1902 to 1922 period. These men are included in both tables, so, for a large proportion of the cases the 1911-1922 table is measuring the occupational mobility of men known to be at a later stage in their career than that measured by the 1902-1911 table. As occupational mobility is generally treated as more likely amongst younger age groups this offers one explanation for the fall in mobility in the second period.

The design matrix for model I is shown in Table 7. The cells in the table are grouped into 5 levels, the fifth level is regarded as a 'default' level and is not labelled to improve the readability of the table.

Table 7: Design Matrix for Model I

	1	2	3	4	5	6	7	8	9
1	1			4		4			
2		1		4		4	4		
3			1	4		4	4		
4	4	4	4	3					
5					3				
6	4	4	4			3			
7	4	4	4				2	4	4
8							4	2	4
9							4	4	2

The design matrix for model I proposes that immobility is divided into three bands corresponding to the elite, non-manual, and manual occupational groups, All significant

[11] Erikson & Goldthorpe, 1993:122

[12] See Hout, 198:46. Replying to Macdonald (1981), Hauser had this to say on the problem:

> I have worried along with Macdonald about the possibility of specifying equivalent models, but the fact is that I do not share his concern abouth the matter. I do not believe that a model consists only of a set of expected values, but it also (and mainly) consists of the structure or story that we use to interpret and explain those expected values.

mobility is grouped in level 4, which proposes symmetrical movement 'up' and 'down' the occupational structure with one exception. Movement is proposed within the manual groups as well as between the elite and the small employers, and white collar workers and skilled. There is also movement from the skilled to the employers but not the reverse. All this movement is assigned to the same level implying that there is only one density of significant mobility (and three of immobility) in the table. Model I fits the 1902-1911 period well, but does not fit the 1911-1922 period:

1902-1911
L^2 44.7336 DF 60 P 0.93
1911-1922
L^2 84.882 DF 60 P 0.02

Model II modifies model I by introducing a second density of mobility. There are now 6 levels in the table with the sixth level being the default and levels 4 and 5 representing mobility. Cells (4,1), (1,4), (4,7), and (7,4) form the new mobility level. The rationale for picking these cells is that they form a pathway of occupational mobility through the table between skilled workers (7), small employers (4) and larger employers and managers (1) . Immobility is still divided into three subsets for the elite, non-manual, and manual classes. Cells in interaction levels 1 and 3, which identify the immobile elite and manual groups, were assigned on the basis that they form groups of immobile occupations. Further to this, it is expected that immobility will be highest at level 1 and lower at level 2 as movement between the manual occupational groups is recognised as part of level 4, the first of the two mobile levels. The assignment of cells (4,4), (5,5), and (6,6) to the same interaction level, 3, does not reflect a belief that they form a coherent group. This is essentially a weak grouping of cells based on the recognition that immobility is significant for all classes in the Caversham data. Because of the involvement of small employers in much of the mobility proposed by the model it is likely that level 3 is less dense than levels, 1 and 2. Of the two mobile groups, level 5 should have the higher density (indicating more mobility) as this traces what we believe to be the central route for occupational mobility in Caversham.

Table 8: Design Matrix for Model II

	1	2	3	4	5	6	7	8	9
1	1			5		4			
2		1		4		4	4		
3			1	4		4	4		
4	5	4	4	3			5		
5					3				
6	4	4	4			3			
7	4	4	4	5			2	4	4
8							4	2	4
9							4	4	2

Model II fits for 1911-1922, and further improves the fit for 1902-1911.

1902-1911
L^2 48.1482 DF 59 P 0.84
1911-1922
L^2 70.4414 DF 59 P 0.15

The level parameters (in logged form and relative to the default level 6, which has a parameter value of 0) for model II are shown in table 9. In both periods the ranking predicted above is matched by the actual parameters.

Table 9: Level Parameters for Model II

level	1902-11	1911-22
1	4.56	5.78
2	3.79	3.92
3	3.37	3.60
4	0.86	0.96
5	1.52	1.52

Model II, then, identifies an immobile elite which recruits from the small employers, white collar and skilled, of which the small employers are the most important contributor. At the opposite corner of the table, self-recruitment (meaning, in terms of this data, continued employment in the same group) is similarly important in the manual groups, but they also form a broader group with upward and downward mobility occurring between the skilled, semi-skilled and unskilled. The skilled connect the elite and manual groups together taking from and giving to non-manual occupational groups. The position of the middle three classes is less clear. There does appear to be a divided between the manual groups and the white collar group, with no transfer between white collar workers and the manual classes. The small employers, by virtue of participating in the skilled-small employer-elite pathway occupy a pivotal position, but again white collar workers do not fit into this system.

Bibliography

Baron, J., N. (1980). Indianapolis and beyond: A structural model of occupational mobility across generations. American journal of sociology, 85(4), 815-839.

Bishop, Y., Fienberg, S., & Holland, P. (1975). Discrete multivariate analysis: Theory and practice. Cambridge, Massachusetts: MIT Press.

Erikson, R., & Goldthorpe, J. H. (1993). The constant flux: A study of class mobility in industrial societies. Oxford: Clarendon Press.

Gilbert, G. N. (1981). Modelling society: An introduction to loglinear analysis for soical researchers. London: George Allen & Unwin.

Grusky, D. B., & Fukumoto, J. K. (1989). Social history update: A sociological approach to historical social mobility. Journal of Social History, 123, 221-232.

Harraway, J. (1995). Regression methods applied. Dunedin, New Zealand: Otago University Press.

Hauser, R. M. (1978). A structural model of the mobility table. Social Forces, 56(3), 919-953.

Hauser, R. M. (1979). Some exploratory methods for modeling mobility tables and other cross-classified data. In D. R. Heise (Ed.), Sociological Methodology 1980, (pp. 141-458). San Francisco: Jossey-Bass.

Hauser, R. M. (1981). Hope for the mobility ratio. Social forces, 60(2), 572-584.

Hout, M. (1983). Mobility Tables. London: Sage.

Jones, F. L., & Davis, P. (1986). Models of society. Sydney: Croom Helm.

Knoke, D., & Burke, P. J. (1980). Log-linear models. London: Sage Publications.

Reynolds, H. T. (1977). The analysis of cross-classifications. New York: The Free Press.

Reynolds, H. T. (1977). The analysis of nominal data. London: Sage Publications.

Van Leeuwen, M. H. D., & Maas, I. (1997). Social mobility in a Dutch province, Utrecht 1850-1940. Journal of Social History, 1997(Spring), 619-644.