Methodological developments in analyzing cross-country social mobility

Richard Breen
Yale Sociology

What are we going to look at?

How does the relationship between class origins and class destinations differ over countries?

But the methods for doing this can also be applied to address questions or test hypotheses that focus on how the relationships between a set of categorical (nominal or ordinal) variables vary over another variable (birth cohorts, countries, surveys).

We might be interested in
how the association between education, age and attitudes towards the environment differs between birth cohorts.

Or we might want to test hypotheses such as
the association between education and first occupation is stronger in countries with high levels of specific skills training (like Germany and Austria) than in countries that lack this.

In general terms: investigate how the set of associations between several (≥ 2) categorical variables varies over another variable (or could be variables), say T. A typical example of this set-up is repeated cross sections.

The main issues:

how do we model the associations between the categorical variables?

how do we model the evolution/ variation in these associations over T?

The talk deals with each of these in turn.

The particular case that we are going to look at is the analysis of two dimensional social mobility tables in which we are interested in making comparisons between different surveys taken at different times (repeated cross sections), or in different countries, or relating to different birth cohorts, etc.

Modelling contingency tables using association (log-linear and log-multiplicative) models

Assume a two-dimensional contingency table, with variables R (for Row), having I (i=1,…,I) categories and C (for Column), with categories indexed j=1,…,J.

In the case of a mobility table, R is the social class or occupational category of the respondent's family of origin and C is the respondent's own social class or occupational group.

The frequencies in each of the I×J cells of the table are denoted f_{ij}.

Supposing that the frequencies have been generated by a sampling distribution that is Poisson (in each cell), multinomial (fixed sample N cross-classified) or product multinomial (fixed row counts), an association model for these data expresses the expected value (under the model), F_{ij}, as

$$F_{ij} = \tau\tau_i^R \tau_j^C \tau_{ij}^{RC} \qquad (1)$$

Here τ_i^R are the main effects of the R variable, τ_j^C the main effects of the C variable, and τ is a grand mean or intercept term. Together these main effects fix the marginal totals of R and C to match their observed values. Our main focus of interest, however, is τ_{ij}^{RC} which is the set of parameters that captures the association between R and C. So if R and C are independent, $\tau_{ij}^{RC} = 1$ for all i × j combinations.

Logarithms are considered more transparent in this context, so the model is more usually written:

$$\log(F_{ij}) = \lambda + \lambda_i^R + \lambda_j^C + \lambda_{ij}^{RC} \qquad (2)$$

There is a one to one relationship between the λs and the τs . So now, if R and C are independent, we would have that all $\lambda_{ij}^{RC} = 0$.

Parameters and identification

Given an I × J table, we can fit at most I × J parameters. Since we have I categories in the R variable, we could fit at most I λ_i^R terms (or equally, τ_i^R terms), and, likewise, at most J λ_j^C terms and I × J λ_{ij}^{RC} terms. But not all of these are identified (think of the usual issue in defining dummy variables).

These effects can be identified in various ways. The most popular are corner parameterizations and centred parameterizations.

Corner: one category of R (e.g. i=1) is set to have $\lambda_1^R = 0$, and likewise for one category of C. This then means that $\lambda_{1j}^{RC} = 0$ and $\lambda_{i1}^{RC} = 0$.

Centred: $\sum_i \lambda_i^R = 0$ and similarly for λ_j^C .

In both cases we are left with I-1 identified R effects, J-1 C effects and therefore (I-1) × (J-1) association effects. If we fit all of these this leaves one degree of freedom for the overall effect, λ .

<u>Goodness of fit</u>

A model which fits all these terms and so uses all the available df is called SATURATED and it fits the data perfectly. Modelling tables using association models involves trying to find a more parsimonious model for the data: this a trade-off between keeping the number of fitted parameters as low as possible (maximizing df) while reproducing the data as closely as possible.

Three major criteria of goodness of fit:
1. Chi-square and likelihood-ratio chi-square:

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(f_{ij} - F_{ij})^2}{F_{ij}} \qquad \text{(3a)}$$

$$G^2 = 2\sum_{i=1}^{I}\sum_{j=1}^{J} f_{ij}\log\frac{f_{ij}}{F_{ij}} \qquad \text{(3b)}$$

G^2 is also known as 'the deviance': I use these terms interchangeably. It is equal to minus twice the difference in the log-likelihoods of the model under consideration and the saturated model.

Both these compare the observed and expected values and can be used to carry out a chi-square test of the significance of the difference between them. The test uses df equal to the number of independent observations in our data (in this case the I x J cells of the table) minus the number of parameters that we have estimated from the data. In the model of independence of R and C because all the $\lambda_{ij}^{RC} = 0$ we do not fit the (I-1) × (J-1) parameters and so we have (I-1) × (J-1) df (as in the usual chi-square test of independence).

χ^2 and G^2 are usually in agreement, but the latter is the default. Widely used to asses whether a particular model, call it M (for example the model of independence), provides an acceptable fit to the data. We use it to answer the question: suppose that in the population M held: what would then be the probability of observing the values that we do in the sample data? More generally, given any observed distribution of cell frequencies in the sample, we can calculate the probability of observing this distribution in a sample of this size, conditional on any specified distribution in the population from which the sample is drawn.

This probability is referred to as the 'p value' and by convention we say that if this probability is less than five per cent ($p < .05$) then the probability of getting such a sample distribution from a given population distribution is so small as to allow us to reject the hypothesis that the population has this particular distribution. In our example, this would mean that we reject the hypothesis that M holds in the population. On the other hand, if $p > .05$ we cannot do this and such a result is usually interpreted as saying that M holds in, or is an adequate description of, the population table.

2. Index of dissimilarity

$$\Delta = \frac{1}{2N_{ij}} \sum_{i=1}^{I} \sum_{j=1}^{J} | f_{ij} - F_{ij} | \qquad (4)$$

N_{ij} is the total observations in the table. Δ tells us what proportion of cases in the expected table would have to be reallocated in order for the observed and expected tables to be identical. This is helpful but cannot be used for hypothesis testing.

3. Bayesian information criterion, *bic*

$$bic = G^2 - df \times \log(N_{ij}) \qquad (5)$$

This quantity is an approximation to minus twice the logarithm of the odds that, given the data, the model in question is true relative to the saturated model being true. Thus if the model in question is more likely to be true than is the saturated model, *bic* will return a negative value. The preferred model according to this criterion is that with the largest negative *bic* statistic.

Bic imposes a severe penalty on more complex models because each additional parameter fitted (when it reduces the df by one) increases *bic* by a value equal to the logarithm of the sample size. In order not to worsen *bic*, each additional parameter included in the model must reduce the deviance (G^2) by at least the logarithm of the sample size. For any sample over the size of about 50, some cases in which the model with the extra parameter would be preferred according to the deviance criterion alone would be rejected by bic in favour of the simpler model. Thus using bic generally leads to a preference for simple models. It has become popular because both G^2 and χ^2 are sensitive to sample size and it may be difficult to find a parsimonious model that fits the data (according to the deviance) when N is large. *Bic* is a 'solution' to this problem.

<u>Modelling associations</u>

The association between R and C is captured using odds ratios. In this context an odds is the ratio of the number in one category of C to the number in another category of C, and an odds ratio is the ratio of these quantities in two different categories of R. In the table there are $I \times J \times (I-1) \times (J-1)/4$ possible odds ratios, but these can be written as a function of a basic set of (I-1) × (J-1) odds ratios. Such a basic set must span the space of all possible odds ratios, and there are, therefore, many possible basic sets. One commonly chosen basis is the set of odds ratios formed from adjacent rows and columns, sometimes called interstitial or local odds ratios

$$\frac{F_{ij} / F_{ij+1}}{F_{i+1j} / F_{i+1j+1}} \equiv \theta_{ij}, \qquad (6)$$

If use $\log(\theta_{ij})$ and write this expression in terms of the parameters of the log-linear model that has generated the F_{ij} we have

$$\begin{aligned}
&\log(\theta_{ij}) = \log(F_{ij}) - \log(F_{ij+1}) - \log(F_{i+1j}) + \log(F_{i+1j+1}) = \\
&\lambda + \lambda_i^R + \lambda_j^C + \lambda_{ij}^{RC} - (\lambda + \lambda_i^R + \lambda_{j+1}^C + \lambda_{ij+1}^{RC}) - \\
&(\lambda + \lambda_{i+1}^R + \lambda_j^C + \lambda_{i+1j}^{RC}) + \lambda + \lambda_{i+1}^R + \lambda_{j+1}^C + \lambda_{i+1j+1}^{RC}
\end{aligned}$$

But these cancel to leave

$$\log(\theta_{ij}) = \lambda_{ij}^{RC} - \lambda_{i+1j}^{RC} - \lambda_{ij+1}^{RC} + \lambda_{i+1j+1}^{RC} \qquad (7)$$

In words: log odds ratios are functions only of the interaction parameters of the model and not of the main (row and column) effects. One other consequence of this is that they are invariant to scalar multiplications of the frequencies in any rows or columns of the table. This property is important in studies of social mobility. The percentages moving between particular origins and destinations may vary greatly between different mobility tables and this will mainly be because the tables differ in their marginal distributions. But we are usually interested in the inequality in mobility chances as between people from different class origins. Odds ratios, because they do not depend on marginal distributions, allow us to make these comparisons.

We want to model the association between R and C as captured in (log) odds ratios: this association depends only on the association parameters of the model: therefore we concentrate on finding parsimonious specifications of the association parameters.

Example: Danish social mobility (from Svalastoga, K. 1959. *Prestige, Class and Mobility*.' London: Heinemann)

5 × 5 table of occupational mobility (father to son) using these categories for both R and C:
(1) Professional and higher administrative;
(2) Managerial and Executive; inspectional; supervisory; other higher non-manual;
(3) Inspectional; supervisory; other non-manual;
(4) Skilled manual; routine non-manual;
(5) Semi-skilled and unskilled manual.

Origin occupational category	**Destination occupational category**				
1	18	17	16	4	2
2	24	105	109	59	21
3	23	84	289	217	95
4	8	49	175	348	198
5	6	8	69	201	246

Independence:
$\log(F_{ij}) = \lambda + \lambda_i^R + \lambda_j^C$; 16 df; $G^2 = 654.21$; p < .0001; *bic*=530

Saturated:
$\log(F_{ij}) = \lambda + \lambda_i^R + \lambda_j^C + \lambda_{ij}^{RC}$; 0 df; $G^2 = 0$; *bic*= 0.

Is there something in between?

Quasi symmetry

$\log(F_{ij}) = \lambda + \lambda_i^R + \lambda_j^C + \lambda_{ij}^{RC}$; 6 df; $G^2 = 6.47$; p =.37; *bic*= -40

QS imposes the constraint that $\lambda_{ij}^{RC} = \lambda_{ji}^{RC}$ - i.e. the association parameters (and thus the odds ratios) are symmetric. QS has 6 df because it fits one parameter for each pair of symmetric cells in the table (=20/2), yielding 10 association parameters out of a possible 16.

Under the model the local log odds ratios look like this:

Origin occupational categories	**Destination occupational categories**			
	1 vs. 2	*2 vs. 3*	*3 vs. 4*	*4 vs. 5*
1 vs. 2	0.689819	-0.02632	0.245668	-0.39742
2 vs. 3	-0.02631	0.516412	0.087321	0.193267
3 vs. 4	0.245652	0.087323	0.422383	0.137605
4 vs. 5	-0.3974	0.193254	0.137608	0.332775

For example, the odds ratio of getting into Skilled manual (4) rather than unskilled manual (5) as between men originating in Professional (1) rather than Managerial (2) occupations is the same as the odds ratio of getting into Professional (1) rather than Managerial (2) occupations as between men of manual (4) rather than unskilled manual (5) origins.

Modelling the association between R and C thus amounts to imposing constraints on the λ_{ij}^{RC} parameters. These can be of two kinds: equality constraints (as in QS), or fixed value constraints, as in quasi-perfect mobility (QPM).

Quasi-perfect mobility (or quasi independence) sets $\lambda_{ij}^{RC} = 0$ when $i \neq j$. That is, independence holds in the cells not on the main diagonal of the table.

Applied to the Danish data QPM returns $G^2 = 248.7$; df = 11; p <.0001; *bic*=163.

Comparing models

Just as we use the deviance to compare the goodness of fit of a particular model with that of the saturated model (which = observed data) we can also use it to compare any models which are nested (i.e. in which model B is a simplification of model A). In this case, QPM is nested in QS (QS can be seen as QPM with some of the constraints on the association parameters removed: or QPM can be seen as QS with some extra constraints).

So, the difference in $G^2 = 248.7 - 6.47 = 242.23$; difference in df = 11-6 = 5. So the test has $G^2 = 242.23$ on 5df , p < .0001. We conclude that there is a significant improvement in the fit of the QS model compared with QPM.

Log-multiplicative (log-bilinear) models

Very many models have been developed to try to capture the pattern of association between the rows and columns of a mobility table. Some examples are 'the core model of social fluidity' in Erikson and Goldthorpe *The Constant Flux* 1992, Michael Hout's model of 'Status, Autonomy and Training' in his 1984 *AJS* paper, Breen and Whelan's 'Agriculture, Hierarchy and Property Model' in their 1994 *ESR* paper.

But one very useful and widely applied class is models begins with

Goodman, L. A. 1979. 'Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories'. *Journal of the American Statistical Association* 74, 367: 537-52.

Suppose that R and C are ordinal with fixed scores, x_i and y_j. A model for these data is

$$\log(F_{ij}) = \lambda + \lambda_i^R + \lambda_j^C + \lambda_{ij}^{RC} \text{ with } \lambda_{ij}^{RC} = \varphi x_i y_j$$

or, more simply,

$$\log(F_{ij}) = \lambda + \lambda_i^R + \lambda_j^C + \varphi x_i y_j \qquad (8)$$

$$\log(\theta_{ij}) = \varphi(x_i y_j - x_{i+1} y_j - x_i y_{j+1} + x_{i+1} y_{j+1}) = \varphi(x_i - x_{i+1})(y_j - y_{j+1}) \qquad (9)$$

If x and y are evenly spaced then we can normalize them so that $x_{i+1} - x_i = 1$ and likewise for y and then

$$\log(\theta_{ij}) = \varphi \qquad (10)$$

This is sometimes called 'uniform association'. All the local odds ratios in the table have the same value.

Now suppose that x and y are unknown: we have the log-bilinear counterpart to (8) as follows:

$$\log(F_{ij}) = \lambda + \lambda_i^R + \lambda_j^C + \varphi \mu_i \nu_j \qquad (11a)$$

Where the row and column scores to be estimated are μ_i and ν_j. But now φ is no longer separately identified and so it is sometimes convenient to absorb it into the estimated scalings to give the model

$$\log(F_{ij}) = \lambda + \lambda_i^R + \lambda_j^C + \mu_i^* \nu_j^* \qquad (11b)$$

Where $\mu_i^* = \sqrt{\varphi}\mu_i$ and $\nu_j^* = \sqrt{\varphi \nu_j}$

This model is sometimes called the RC(M) model, where M refers to the number of pairs of scalings since one might extend the model thus

$$\log(F_{ij}) = \lambda + \lambda_i^R + \lambda_j^C + \mu_i \nu_j + \eta_i \kappa_j \, .$$

But we will deal only with RC(1) models.

One useful simplification is $\mu_i = \nu_i$.

RC(1) applied to Danish data: $G^2 = 40.93$; df = 9; p < .0001; *bic* = -29.

RC(1) with equal scalings: $G^2 = 46.04$; df = 12; p <.0001; *bic*= -47.

Bic prefers this to all other models so far, although by the deviance it does not fit the data.

Scalings have to be normalized: in this case they sum to zero:

-0.6041 -0.349 -0.0124 0.3295 0.6361

These scalings look pretty equally spaced: imposing this gives us UA:

UA has $G^2 = 47.99$; df=15; p <.0001; bic= -69; $\log(\theta) = 0.24$.

The odds ratio between adjacent rows and columns is constant and equal to exp(0.24) = 1.27. If the scores run from low to high, then the odds of being in a given destination class rather than the next one are 1.27 times greater for someone from one origin category compared to someone from the next category.

These models are parsimonious and so they lend themselves to comparisons across another dimension (such as time, country, etc.).

Making comparisons of the association in several tables
Given K (k=1,…,K) tables each with I (i=1,…,I) rows and J (j=1,…,J) columns, our object is to find accurate and parsimonious models of the variation in the RC association over tables. Tables might be surveys (as in repeated cross sections), birth cohorts, countries. Call this dimension L (Layer).

The basic model for a 3-way table is a straightforward extension of the two-way model, with the added k subscript:

$$\log(F_{ijk}) = \lambda + \lambda_i^R + \lambda_j^C + \lambda_k^L + \lambda_{ij}^{RC} + \lambda_{ik}^{RL} + \lambda_{jk}^{CL} + \lambda_{ijk}^{RCL} \qquad (12)$$

Now we have three main effect sets of terms, three sets of two way associations, and one three way association. Together the main effects and the two-way associations fix the marginal distributions of the three variables to equal their observed values, and so our interest is λ_{ijk}^{RCL} since this tells us how the RC association varies over L.

This fits a maximum of (I-1) × (J-1) × (K-1) parameters: using all these, together with all the main effects and two-way interactions, would once again give a saturated model.

Example: add an occupational mobility table for England and Wales (Glass, D.V. (ed.) 1954. *Social Mobility in Britain*. London: Routledge and Kegan Paul) *and compare with Danish table.*

Take the best fitting model for the Danish data (QS) and see if the same model fits the British data. So now we (i) impose constraints on the within-table association (namely symmetry of the odds ratios) and (ii) impose $\lambda_{ijk}^{RCL} = 0$ (same pattern of association in each table).

Common QS: $G^2 = 51.68$, df=22; p=.0003; *bic*= -139.

Relax $\lambda_{ijk}^{RCL} = 0$ to give nation specific QS: $G^2 = 17.42$;df=12; p=.135;*bic*=-87.

This fits the data but all it tells us is that QS fits in each country but with different parameter values. Note that all the parsimony in this model comes from the within-table constraint of QS and not from any constraint on how QS varies over country.

Investigating λ_{ijk}^{RCL}

Suppose we have a model of the R ×C association within each table (such as the saturated model or QS or UA or RC(1)): how can we parsimoniously model differences across L? One very useful approach is due to Xie, Y. 1992. 'The Log-Multiplicative Layer Effect Model for Comparing Mobility Tables'. *American Sociological Review* 57: 380-395.

Set: $\lambda_{ijk}^{RCL} = \lambda_{ij}^{RC} \times \beta_k^L$ (13)

This means that within a given table (i.e. for fixed k), we get log odds ratios equal to

$$\log(\theta_{ij|k}) = [\lambda_{ij}^{RC} - \lambda_{i+1\,j}^{RC} - \lambda_{ij+1}^{RC} + \lambda_{i+1\,j+1}^{RC}]\beta_k^L \qquad (14)$$

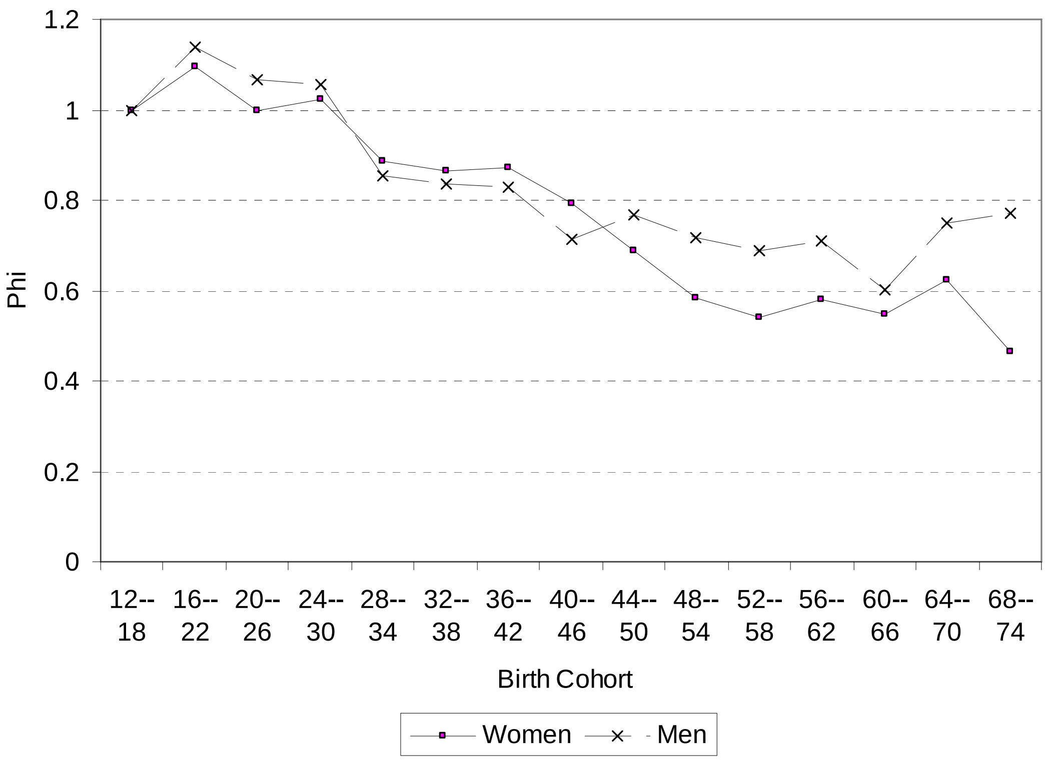
To identify the table-specific values of β_k^L we need to impose some constraints and the usual one is to set $\beta_1^L = 1$. This means that we can write

$$\log(\theta_{ij|k}) = \log(\theta_{ij1})\beta_k^L \text{ for k=2,…,K} \qquad (15)$$

So the log odds ratios in the different tables have the same pattern (ratios of odds ratios are the same in all tables) but they differ in strength over tables according to the scalar β_k^L. If this is larger than 1, the association is stronger than in the first table: less than one it is weaker and origins and destinations are less strongly related. Also known as the ‘unidiff’ model.

Breen and Jonsson (forthcoming) applied this model to look at the OD association over 15 birth cohorts in Sweden:

Figure 2: Cohort change in social fluidity for men and women. Fluidity coefficients from unidiff models.



In this case, λ_{ij}^{RC} was the saturated model. Since each table was of dimensions 6 × 6 this meant that we fitted 25 association parameters to each table, but change was modelled using 14 parameters (one for each cohort, the first one being fixed to equal 1).

But λ_{ij}^{RC} could be more parsimonious too. If it was QS then equation 15 would still hold but with the added constraint that, within each table, the odds ratios were symmetric. Applying this model to the Danish and English data we find $G^2 = 43.67$; df=21; p=.003; *bic*=-139.

The estimate of $\beta_{Denmark}$ is 1.19, showing that Denmark was a less open country than England: the symmetric odds ratios were exp(1.19)=3.3 times larger there.

Finally consider a more complicate model in which λ_{ij}^{RC} contains several different elements

$$\lambda_{ij}^{RC} = \lambda_{ii}^{RC} + \varphi \mu_i \mu_j$$

In words: there is a set of parameters that fits the cells on the main diagonal exactly (common in mobility research) and then an RC(1) model for the table as a whole which constrains the scalings of R and C to be the same. First fit this model to Denmark and England separately:

$G^2 = 10.24$ (Denmark), $G^2 = 11.12$ both with 6df.

Now we could use the log-multiplicative model to look at differences between the countries in one or more of the diagonal effects (λ_{ii}^{RC}), the overall association parameter (φ) and the scalings of rows and columns (μ_i). Some analysis suggests that the diagonal effects do not differ much between Denmark and England, so we try varying the overall association parameter. So the model is

$$\lambda_{ijk}^{RCL} = \lambda_{ii}^{RC} + \beta_k^L \varphi \mu_i \mu_j$$

But now β and φ are not separately identified so we drop β and write

$$\lambda_{ijk}^{RCL} = \lambda_{ii}^{RC} + \varphi_k^L \mu_i \mu_j$$

This model says that both countries have the same scalings of origins and destinations, but the strength of the association between them differs. The Danish association parameter is estimated to be 1.34 times larger than that for England, though the model fails to fit the data: $G^2 = 38.39$; df=22; p=.02;*bic*=-153.

So now we try the model

$$\lambda_{ijk}^{RCL} = \lambda_{ii}^{RC} + \varphi_k^L \mu_{ik} \mu_{jk}$$

This is an excellent fit to the data: $G^2 = 21.92$; df=19; p=.29;*bic*=-143. (But notice that *bic* prefers the simpler model)

Because we allowed both the scalings and the association parameter to differ they are not separately identified, and so we can absorb the former into the latter. The differences between the countries are then captured in the differences in the scalings of origins and destinations.

Scalings of Origins and Destinations by country					
	Professional	*Managerial*	*Supervisory*	*Skilled*	*Non-skilled*
England	-4.16	-2.18	0.54	2.33	3.47
Denmark	-3.36	-3.08	-0.08	2.67	3.84
First differences of above					
England	-1.98	-2.72	-1.79	-1.14	
Denmark	-0.28	-2.99	-2.76	-1.17	
Estimated Local Odds Ratios					
England	*1 vs. 2*	*2 vs. 3*	*3 vs. 4*	*4 vs. 5*	
1 vs. 2	6.75	1.57	1.35	1.21	
2 vs. 3	1.57	2.94	0.99	1.31	
3 vs. 4	1.35	0.99	1.86	1.29	
4 vs. 5	1.21	1.31	1.29	2.08	
Denmark					
1 vs. 2	4.86	1.06	1.07	1.03	
2 vs. 3	1.06	3.24	1.28	1.33	
3 vs. 4	1.07	1.28	2.64	1.41	
4 vs. 5	1.03	1.33	1.41	2.08	

We see a big gap in the scalings in England between Professional and Managerial, but not in Denmark where these two are quite close. This is reflected in the larger odds ratios in the first row or column) in England than in Denmark. But elsewhere the gaps are larger in Denmark and this is reflected in the odds ratios in the nine cells in the bottom right of the table which are all larger in Denmark than in England. In particular odds ratios involving a comparison of Supervisory with Skilled are very much larger in Denmark.

A nice recent example of this approach, applied to temporal comparisons, is Ganzeboom and Luijkx, 2004 ‘More Recent Trends in Intergenerational Occupational Class Reproduction in the Netherlands 1970-2004’ *The Netherlands Journal of Social Sciences,* Vol. 40, No. 2, pp.114-42.

Concluding remarks

All this extends readily to many tables from different countries, times, etc. And can be extended in other ways too. Recall that our aims are
to model λ_{ij}^{RC}; and
to model λ_{ijk}^{RCL}.

An extension of the first is where we are interested in the associations between three or more variables – say class origins, educational attainment and class destinations. Then we want to model the associations between each pair of these and, possibly, their three way interaction. This is often done: a good recent example is Ganzeboom and Luijkx, chapter 14 of *Social Mobility in Europe* (OUP 2004) edited by Richard Breen.

An extension of the second is where we are interested in variation over two dimensions – say country (L) and time (T). Then we want to model the term λ_{ijkl}^{RCLT}. One way to do this with log-multiplicative models is to write:

$$\lambda_{ijkl}^{RCLT} = \lambda_{ij}^{RC} \times \varphi_{kl}^{LT}$$

So that there is a scalar raising or lowering the log odds ratio in each country × time combination. Or the specification could be ‘additive’:

$$\lambda_{ijkl}^{RCLT} = \lambda_{ij}^{RC} \times \varphi_k^L \varphi_l^T$$

Here the evolution of the log odds ratios over time is the same in each country. Yet another possibility is

$$\lambda_{ijkl}^{RCLT} = \lambda_{ijk}^{RCL} \times \varphi_l^T$$

Where variation between countries is modelled non-log multiplicatively (log linearly perhaps) while temporal variation is log multiplicative.

Finally another possibility is

$$\lambda_{ijkl}^{RCLT} = \lambda_{ij}^{RC1} + \lambda_{ij}^{RC2} \times \varphi_k^L + \lambda_{ij}^{RC3} \varphi_l^T$$

Here there are three components to the row – column association: one that is constant over all countries and time points, one that varies log-multiplicatively over countries and another that varies over time. This is a case of the Goodman-Hout model (Goodman, Leo A. and Hout, Michael 1998. ‘Statistical Methods and Graphical Displays for Analyzing How the Association Between Two Qualitative Variables Differs Among Countries, Among Groups or Over Time: A Modified Regression-type Approach’, *Sociological Methodology* 28: 175-230.)

Several models of this kind appear in Chapter 3 (by Breen and Luijkx) of *Social Mobility in Europe* and in Breen and Jonsson ‘Explaining Change in Social Fluidity: Educational Equalization and Educational Expansion in Twentieth Century Sweden’ forthcoming in *AJS*.