THE MULTILEVEL MODEL FRAMEWORK FOR INVESTIGATING VARIATIONS IN SOCIO-ECONOMIC STATUS BY OCCUPATIONAL CLASSIFICATION.

Mark Tranmer, University of Manchester
Paul Lambert, David Griffiths, University of Stirling
Vernon Gayle, University of Edinburgh.

1. Introduction.

Hierarchical multilevel models have been used to analyse ego-net data for which there is no overlap between the alters of one ego-net to the next (or the overlap is so minimal it can be reasonably ignored). Snijders et. al (1995) proposed such a method using an interval (continuous) response variable, based on networks of cocaine users in an urban area. de Miguel Luken and Tranmer (2010) applied a logistic multilevel model to ego net data, where the ego is a recent immigrant to Spain and the alters are people those immigrants name in their support network. In this case the there was a binary tie from ego to alter to indicate if alter was another recent immigrant to Spain or someoe who had lived in Spain All their lives.

In these ego-net analyses, the unit of analysis is the tie between alter and ego and the aim of the analyses is to model variations in the values of these ties between and within ego nets. In the models, ego is at level 2 and alter at level 1. Both Snijders et al (1995) and de Miguel and Tranmer (2010) also explained how homophily (similarity) of ego and alter's attributes can be included in such a model framework to assess whether homophily explains some of the variation in the alter-ego tie values between and within the ego-nets

In this working paper, we consider the use of this ego-net idea where the ego is the occupational group of the husband in a partnership and the alter is the occupational group of his wife. The tie value we model here is the similarity of a socio economic score – such as Camsis. We have recoded this to take the value 1 if the Camsis scores for husbands and wives are very similar (within 10 units of Camsis), and 0 otherwise. Our data consist of a series of husband and wife pairs. We can hence sort these records by husband's occupational group and use a multilevel model assess tie variation to wives occuaptional group witihn and between each of the husband's occupational groups.

We propose the use of the multilevel modelling framework in this situation. In particular, the multilevel logistic model can be used to model variations in the log odds of the probability of Camsis score similariy for husband and wife in different occupational groups. Here, using ego-net terminology, we treat the husband's occupational group as the ego and the "alters" are the occupational groups of wives within husband's occupational group.

Model 1 measures the extent of the variation in similarity in different husband occupational groups. The bigger the value of σ^2_{u0} , the greater the variation in the (log odds of the) similarity of Camsis score between groups. We can fit Model 1 (and the models that follow) in MLwiN using MCMC estimation, as it is a non-linear model. We can evaluate the goodness of fit of this models with other models, such as a baseline single level logisitc regession that does not include an occupational group level using the Deviance Information Criterion (DIC) measure – the smaller the value the better the model fit, having taken model complexity into account.

Model 1: Null Model (measuing extent of variation in similarity of Camsis score in different occupational groups).

y_{ij} similarity of s.e.s measure for H and W for person i, in occupational group j
1 if similar, 0 if not.

$$y_{ij} \sim \text{Binomial}(1, \pi_{ij})$$
$$\text{Logit}(\pi_{ij}) = \beta_{0j}$$
$$\beta_{0j} = \beta_0 + u_{0j}$$
$$u_{0j} \sim N(0, \sigma^2_{U0})$$
$$\text{Var}(y \mid \pi_{ij}) = \pi_{ij}(1 - \pi_{ij})$$

2. Extending the Model for Homophily.

Model 1 can be extended to include explanatory variables for attributes of husband and wife. We illustrate this with educational status; in particular whether husband or wife have a degree/diploma. This leads to four possibilities: neither husband nor wife have a degree/diploma. Only husband has a degree/diploma. Only wife has a degree/diploma. Both husband and wife have a degree/diploma. By introducing a series of indicator variables, x_{1ij}, x_{2ij} and x_{3ij} (which is the interaction of x_{1ij} and x_{2ij}) can allow for these four possibilities in the model.

Model 2: Allowing for Homophily

Definitions of variables:

y_{ij} similarity of s.e.s measure for H and W for person i, in occupational group j
1 if similar, 0 if not.
$x_{1ij} = 1$ if h has a degree/diploma, 0 otherwise.
$x_{2ij} = 1$ if w has a degree/diploma, 0 otherwise
$x_{3ij} = 1$ if both h and w have degree/diploma, 0 otherwise
Hence $x_{3ij} = x_{1ij} \times x_{2ij}$

Because we are working in a multilevel framework, we can give each of the indicator variables random coefficients – this will allow us to measure the extent of between occupational group variation (i.e. the total level 2 variance) for different combinations of attributes: neither husband or wife have a degree/diploma, only husband has a degree/diploma, only wife has a degree/diploma, both husband and wife have a degree/diploma – as we illustrate in our example. In Model 2 the fixed coefficients indicate whether the log odds of husband and wife having similar s.e.s increase, decrease or stay the same for each combination of degree/diploma by husband/wife. We can use the estimates of variance and covariance components from the fitted model to calculate the estimated total variance in log odds for husband and wife for each of the four combinations.

Model 2

$$y_{ij} \sim \text{Binomial}(1, \pi_{ij})$$
$$\text{Logit}(\pi_{ij}) = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \beta_{3j}x_{3ij}$$
$$\beta_{0j} = \beta_0 + u_{0j}$$
$$\beta_{1j} = \beta_1 + u_{1j}$$
$$\beta_{2j} = \beta_2 + u_{2j}$$
$$\beta_{3j} = \beta_3 + u_{3j}$$
$$\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \\ u_{3j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma^2_{u0} & & & \\ \sigma_{u01} & \sigma^2_{u1} & & \\ \sigma_{u02} & \sigma_{u12} & \sigma^2_{u2} & \\ \sigma_{u03} & \sigma_{u13} & \sigma_{u23} & \sigma^2_{u3} \end{bmatrix}$$

$$\text{Var}(y \mid \pi_{ij}) = \pi_{ij}(1 - \pi_{ij})$$

Calculating total variation for husband and wife education combinations.

1. Neither husband or wife have a degree/diploma = $\hat{\sigma}^2_{u0}$

2. Only husband has a degree/diploma = $\hat{\sigma}^2_{u0} + 2\hat{\sigma}_{u01} + \hat{\sigma}^2_{u1}$

3. Only wife has a degree/diploma = $\hat{\sigma}^2_{u0} + 2\hat{\sigma}_{u02} + \hat{\sigma}^2_{u2}$

2. Both husband and wife have a degree/diploma =

$$\hat{\sigma}^2_{u0} + 2\hat{\sigma}_{u01} + \hat{\sigma}^2_{u1} + 2\hat{\sigma}_{u02} + 2\hat{\sigma}_{u12} + \hat{\sigma}^2_{u2} + 2\hat{\sigma}_{u03} + 2\hat{\sigma}_{u13} + 2\hat{\sigma}_{u23} + \hat{\sigma}^2_{u3}$$

3. Illustrative example.

The data are from the British Household Panel Survey (BHPS). The response variable is whether the Camsis score within 10 units for husband and wife partnership. The explanatory variables are whether or not the husband, wife, or both have a degree/diploma.

Table 1: Model results - Model 1 is a null model, Model 2A is a model that only contains fixed effects for the degree/diploma status of husband wife. Model 2B includes random coefficients on the degree diploma variables. All models fitted in MLwiN with MCMC estimation based on chains of 10,000 and default gamma priors for the variance/covariance components.

	Model 1		Model 2A		Model 2B	
Response	y		y		y	
		S.E.		S.E.		S.E.
Fixed Part						
cons	-0.318	0.059	-0.151	0.058	-0.223	0.079
degdip			-0.265	0.028	-0.218	0.099
spdegdip			-0.48	0.033	-0.474	0.136
bothdegdip			0.552	0.044	0.228	0.152
Random Part						
cons/cons	0.868		0.883		1.763	
bothdegdip/cons					0.648	
bothdegdip/bothdegdip					3.801	
degdip/cons					-0.671	
degdip/degdip					2.007	
spdegdip/cons					-1.586	
spdegdip/degdip					0.921	
spdegdip/spdegdip					3.657	
bothdegdip/degdip					-1.898	
bothdegdip/spdegdip					-2.183	
Level: spjsoc						
DIC:	58292.161		58041.867		54840.575	
Units: mrjsoc	352		352		352	
Units: spjsoc	44628		44628		44628	

Model 1A is a baseline model to allow us to see whether the goodness of fit improves for the more realistically complex models. We see that the DIC reduces for Model 1B, which includes degree/diploma combinations for husband and wife – we can also see that the log odds of similar camsis scores change, on average, for different combinations of husband and wife degree/diploma. Model 2 is the most complex, but leads to a large reduction in the DIC and is therefore the best fit to the data, despite the increase complexity of the model. From Model 2, we can estimate the total variance in the chance of the Camsis score being the same (or very similar) for the four different combinations of husband/wife and degree/diploma status. We can see that the estimated extent of variation is different for each of these combinations. We can also general residual caterpillar plots to identify husband's occupational groups that are associated with especially high or low level 2 residuals for different combinations of husband/wife educational status. (See Figure 1 below)

Total variance estimates for different husband/wife degree/diploma combinations.

Estimated total between husband occupational group variance for neither husband nor wife has a degree/diploma: 1.763

Estimated total between husband occupational group variance for only husband has a degree/diploma: 2.428

Estimated total between husband occupational group variance for only wife has a degree/diploma: 2.248

Estimated total between husband occupational group variance for both husband and wife have a degree/diploma: 1.690

4. Conclusion.

Based on previous applications of multilevel models to non overlapping ego-nets we have shown how the multilevel models can be applied to "ego-nets" of social economic status within occupational groups. By, for example, making husband's occupational group the ego and wife's occupational group the alter within those egos – where the unit of analysis is the tie between husband and wife for each husband wife pair in a dataset. The hierarchy could also be swapped around so that wife's occupational status was the ego and husband's occupational status the ego. Other groupings or classifications could be used.

There is clearly more to be done with the application and interpretation of this model framework in this context, and we intend to extent this working paper into a journal paper on this topic. We hope this working paper provides an indication of the potential of multilevel models in the analysis of socio-occupational status in partnerships and a way to investigate homophily in this context. We illustrated homophily with educational status here. Other variables could also be used.

References:

Snijders, T., Spreen, M., Zwaagstra, R., (1995). The use of multilevel modelling for analysing personal networks: networks of cocaine users in an urban area. Journal of Quantitative Anthropology 5, 85–105.

de Miguel Luken, V., & Tranmer, M. (2010). Personal support networks of immigrants to Spain: A multilevel analysis. Social Networks, 32(4), 253-262.

Figure 1: Level 2 residuals for the constant and three indicator variables of husband/wife degree/diploma status.