Multilevel Masterclass

Kelvyn Jones, University of Bristol

26th November 2008, 10:00am - 5:00pm (1-hour lunch break)

Multilevel Modeling: An Introduction

Time	Session
10:00 - 11:00	1: What is multilevel modelling?
11:00 - 11:45	2: Varying relations & random effects: Theory
11.45:12.00	Break
12:00- 13:00	3: Varying relations & random effects: A Demonstration using MLwiN; using the software
13:00 – 14:00	*Lunch*
14:00 - 14:45	4: Variance Functions
14:45 - 15:45	5: Logit Models
15:45 - 16:00	*Break*
16:00 - 16:45	6: Using MCMC estimation (including Spatial Models)
16:45 - 15:00	7: Resources for Going Further

1 What is multilevel modelling?

Session outline

- *Realistically complex modelling*
- *Structures that generate dependent data*
- *Dataframes for modelling*
- *Distinguishing between variables and levels (fixed and random classifications)*
- *Why should we use multilevel modelling as compared to other approaches?*

Multilevel Models:

AKA

- random-effects models,
- hierarchical models,
- variance-components models,
- random-coefficient models,
- mixed models

- First known application: 1861: several telescopic observations per night for several different nights; separated the variance into between and within-night variation (technically: one-way, random-effects model)

- Increasingly widespread use since late 1980's associated with development of effective algorithms, linked to software, for model estimation

ESRC National Centre for Research Methods

Learning Environment for Multilevel Methodology and Applications

ESRC Economic & Social Research Council

Realistically complex modelling

Statistical models as a formal framework of analysis with a complexity of structure that matches the system being studied

Three KEY Notions

Modelling contextuality: micro & macro

eg individual house prices vary from n'hood to neighbourhood

eg individual house prices varies differentially from n'hood to neighbourhood according to size of property

Modelling heterogeneity

standard regression models 'averages', ie the general relationship

ML **additionally** models ***variances***

Eg between-n'hood AND between-house, within-n'hood variation

Modelling dependent data deriving from complex structure

series of structures that ML can handle routinely, ontological depth!

ESRC National Centre for Research Methods

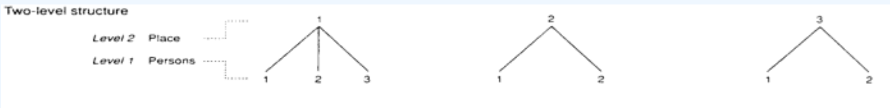
Learning Environment for Multilevel Methodology and Applications

ESRC Economic & Social Research Council

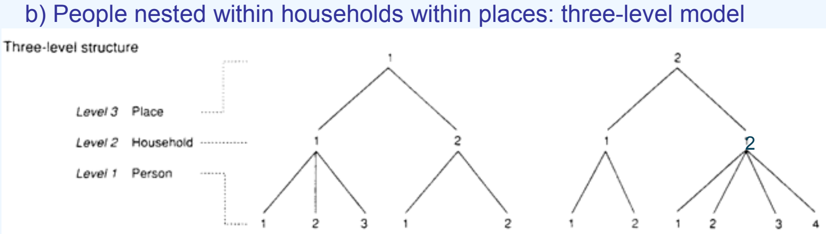
Modelling data with complex structure

1: Hierarchical structures : model all levels simultaneously

a) People nested within places: two-level model



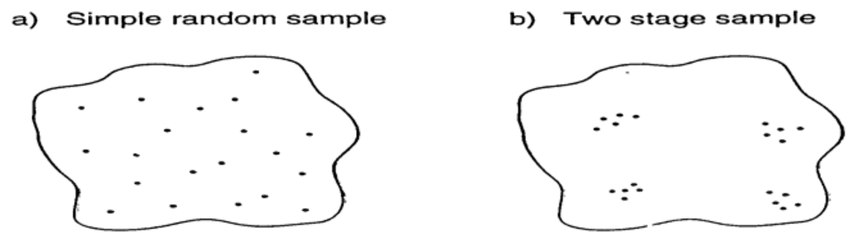
b) People nested within households within places: three-level model



Note imbalance allowed!

Multistage sampling designs

- for efficient collection of data
- most large-scale surveys are **not** SRS



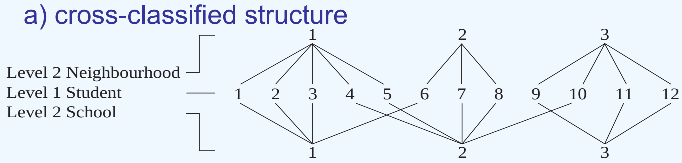
- Two-level structure imposed by design
- respondents nested within PSU's

Multistage sampling designs

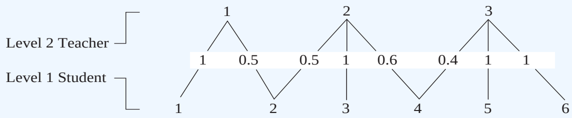
- Multistage designs (usually) generate dependent data
  - individuals living within the same PSU can be expected to be more alike than a random sample
- The 'design effect'
  - Inferential procedures (SE's, confidence limits, tests) are likely to be incorrect
  - incorrect estimates of precision
  - Type 1 errors: finding a relationship where none exists
- Multilevel models model this dependency and automatically corrects for the 'design effect'

Non- Hierarchical structures

a) cross-classified structure

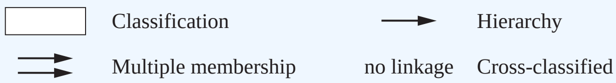


b) multiple membership with weights

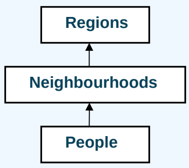


- So far, unit diagrams now……

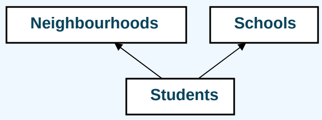
CLASSIFICATION DIAGRAMS



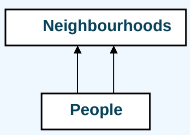
a) 3-level hierarchical structure



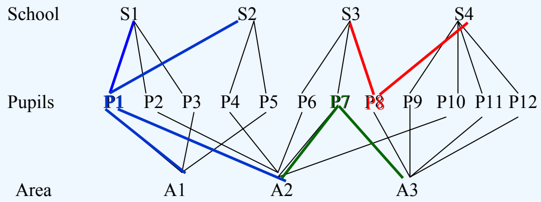
b) cross-classified structure



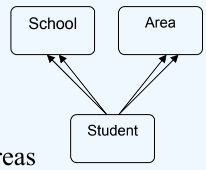
c) multiple membership structure



Combining structures: crossed-classifications and multiple membership relationships



Pupil 1 moves in the course of the study from residential area 1 to 2 and from school 1 to 2



Pupil 8 has moved schools but still lives in the same area

Pupil 7 has moved areas but still attends the same school

Now in addition to schools being crossed with residential areas pupils are ***multiple members*** of both areas and schools.

A data-frame for examining neighbourhood effects on price of houses

Classifications or levels		Response	Explanatory variables		
House i	*N'hood j*	House Price ij	*No of Rooms ij*	*House type ij*	*N'hood Type j*
1	1	75	6	Semi	Suburb
2	1	71	8	Semi	Suburb
3	1	91	7	Det	Suburb
1	2	68	4	Ter	Central
2	2	37	6	Det	Central
3	2	67	6	Ter	Central
1	3	82	7	Semi	Suburb
2	3	85	5	Det	Suburb
1	4	54	9	Terr	Central
2	4	91	7	Terr	Central
3	4	43	4	Semi	Central
4	4	66	55	Det	Central

Questions for multilevel (random coefficient) models

•What is the between-neighbourhood variation in price taking account of size of house?

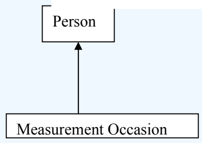
Are large houses more expensive in central areas?

• Are detached houses more variable in price

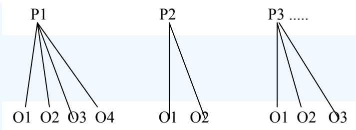
Form needed for MLwiN

Two level repeated measures design: classifications, units and dataframes

Classification diagram



Unit diagram



Classifications or levels		Response	Explanatory variables	
Occasion i	*Person j*	*Income$_{ij}$*	*Age$_{ij}$*	*Gender$_j$*
1	1	75	25	F
2	1	85	26	F
3	1	95	27	F
1	2	82	32	M
2	2	91	33	M
1	3	88	45	F
2	3	93	46	F
3	3	96	47	F

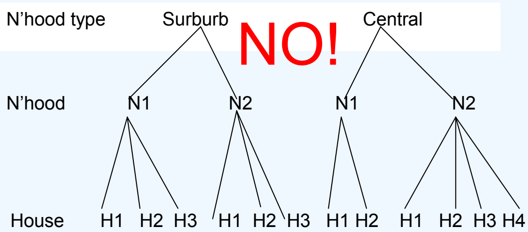
a) in long form

Form needed for MLwiN

Person	*Inc-Occ1*	*Inc-Occ2*	*Inc-Occ3*	*Age-Occ1*	*Age-Occ2*	*Age-Occ3*	*Gender*
1	75	85	95	25	26	27	F
2	82	91	*	32	33	*	M
3	88	93	96	45	46	47	F

b) in short form :

Distinguishing Variables and Levels



Classifications or levels			Response	Explanatory Variables	
House I	*Nhood j*	*Type k*	*Price ijk*	*Rooms ijk*	*House type ijk$_{ijk}$*
1	1	Suburb	75	6	Det
2	1	Suburb	71	4	Det
3	1	Suburb	91	7	F
1	2	Central	68	9	F
2	2	Central	37	6	M
Etc					

N'hood type is not a random classification but a fixed classification, and therefore an attribute of a level; ie a VARIABLE

Random classification: if units can be regarded as a *random* sample from a wider population of units. Eg houses and n'hoods

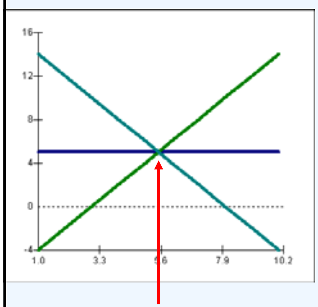
Fixed classification is a small *fixed* number of categories. Eg Suburb and central are not two types sampled from a large number of types, on the basis of these two we cannot generalise to a wider population of types of n'hoods,

Analysis Strategies for Multilevel Data

What are the alternatives; and why use multilevel modelling?

I Group-level analysis. Move up the scale: analyse only at the macro level; Aggregate to level 2 and fit standard regression model.

- *Problem*: Cannot infer individual-level relationships from group-level relationships (ecological or aggregation fallacy)



Same mean $\bar{Y}_j$, $\bar{X}_j$ but three very different within school relations

Example: research on school effects

Response: Current score on a test, turned into an average for each of j schools; $\bar{Y}_j$

Predictor: past score turned into an average for each of j schools $\bar{X}_j$

Model: regress means on means

Means on means analysis is meaningless!
Mean does not reflect **within group** relationship

Aitkin, M., Longford, N. (1986), "Statistical modelling issues in school effectiveness studies", *Journal of the Royal Statistical Society*, Vol. 149 No.1, pp.1-43.

I Group-level analysis Continued Aggregate to level 2 and fit standard regression model.

- *Problem*: Cannot infer individual-level relationships from group-level relationships (ecological or aggregation fallacy)

Level	Black illiteracy	Foreign-born illiteracy
Individual	0.20	0.11
State	0.77	-0.52

Robinson (1950) demonstrated the problem by calculated the correlation between illiteracy and ethnicity in the USA for 2 aggregate and individual

2 scales of analysis for 1930 USA

- Individual: for 97 million people; States: 48 units
- very different results! The *ECOLOGICAL* FALLACY

Analysis Strategies continued

II Individual-level analysis. Move down the scale; work only at the micro level; Fit standard OLS regression model

- *Problem*: Assume independence of residuals, but may expect dependency between individuals in the same group; leads to underestimation of SEs; Type I errors

Bennet's (1976) "teaching styles" study uses a single-level model: test scores for English, Reading and Maths aged 11 were significantly influenced by teaching style; PM calls for a return to 'traditional' or formal methods

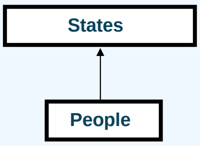
Re-analysis: Aitkin, M. et al (1981) Statistical modelling of data on teaching styles (with Discussion). *J. Roy. Statist. Soc. A* 144, 419-461

Used multilevel models to handle dependence of pupils within classes; no significant effect

Also *atomistic* fallacy………….

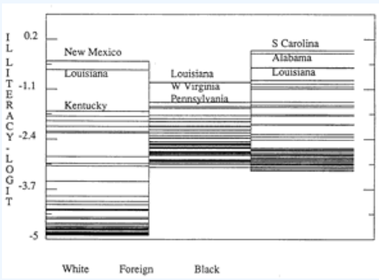
What does an individual analysis miss?

- Re-analysis as a two level model (97m in 48 States)

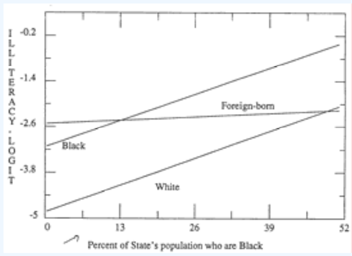


Who is illiterate? Individual model

Does this vary from State to State?



Cross-level interactions?



Analysis Strategies (cont.)

III Contextual analysis. Analysis individual-level data but include group-level predictors

Problem: Assumes all group-level variance can be explained by group-level predictors; incorrect SE's for group-level predictors

- Do pupils in single-sex school experience higher exam attainment?
- Structure: 4059 pupils in 65 schools
- Response: Normal score across all London pupils aged 16
- Predictor: Girls and Boys School compared to Mixed school

Parameter	**Single level**	**Multilevel**
Cons (Mixed school)	-0.098 (0.021)	-0.101 (0.070)
Boy school	0.122 (**0.049**)	0.064 (**0.149**)
Girl school	0.245 (**0.034**)	0.258 (**0.117**)
Between school variance(σ_u^2)		0.155 (0.030)
Between student variance (σ_e^2)	0.985 (0.022)	0.848 (0.019)

SEs

ESRC National Centre for Research Methods | Learning Environment for Multilevel Methodology and Applications | ESRC Economic & Social Research Council

Analysis Strategies (cont.)

IV Analysis of covariance (fixed effects model). Include dummy variables for each and every group

Problems

- What if number of groups very large, eg households?
- No single parameter assesses between group differences
- Cannot make inferences beyond groups in sample
- Cannot include group-level predictors as all degrees of freedom at the group-level have been consumed
- Target of inference: individual School versus schools

ESRC National Centre for Research Methods | Learning Environment for Multilevel Methodology and Applications | ESRC Economic & Social Research Council

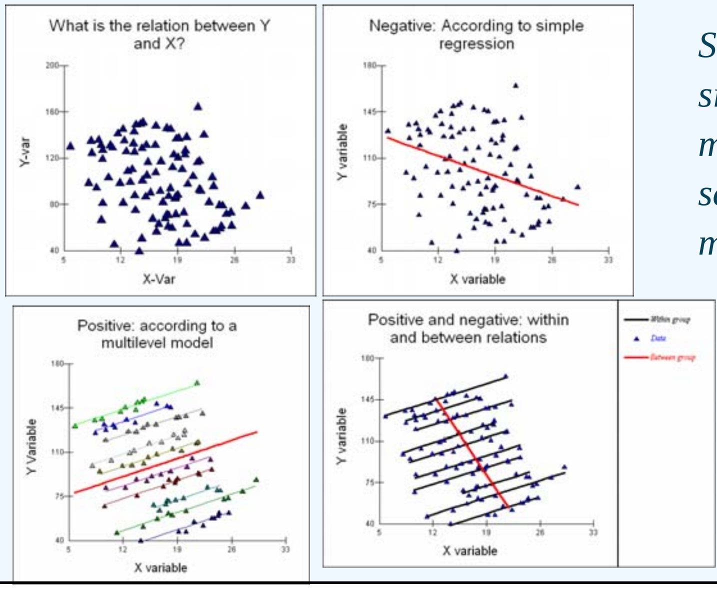
Analysis Strategies (cont.)

V Fit single-level model but adjust standard errors for clustering (GEE approach)

Problems: Treats groups as a nuisance rather than of substantive interest; no estimate of between-group variance; not extendible to more levels and complex heterogeneity

VI Multilevel (random effects) model. Partition residual variance into between- and within-group (level 2 and level 1) components. Allows for un-observables at each level, corrects standard errors, Micro AND macro models analysed simultaneously, avoids ecological fallacy and atomistic fallacy: *richer set of research questions* ***BUT*** (as usual) need well-specified model and assumptions met.

Why should we use multilevel model?



Sometimes: single level models can be seriously misleading!

Some reading

- Johnston, RJ, Jones, K, Propper, C & Burgess, SM (2007) Region, local context, and voting at the 1997 General Election in England, ***American Journal of Political Science***, **51 (3)**, 641-655
- Jones, K, Subramanian, SV & Duncan, C. (2003) Multilevel methods for public health research', in Kawachi, I and Berkman L F (Eds.), ***Neighbourhoods and Health***, Oxford University Press, 65-111.
- Jones, K & Duncan, C. (2001) Using multilevel models to model heterogeneity: potential and pitfalls, ***Geographical Analysis***, **32**, 279-305
- Bullen N, Jones K, Duncan C. (1997) Modelling complexity: analysing between individual and between-place variation—a multilevel tutorial. ***Environment and Planning***. 29: 585–609
- Jones, K., and Duncan, C. (1998) 'Modelling context and heterogeneity: Applying multilevel models', in E. Scarbrough and E. Tannenbaum (Eds.), ***Research Strategies in the Social Sciences***. Oxford University Press.

2 MODELLING HETROGENEITY: *varying relations & random effects*

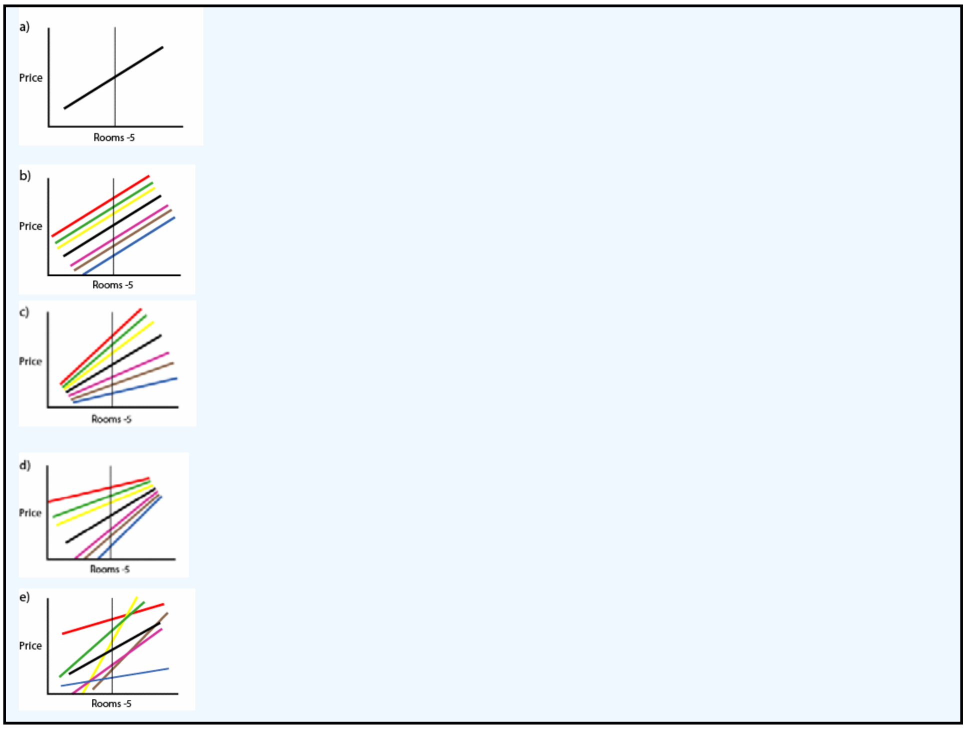
VARYING RELATIONS

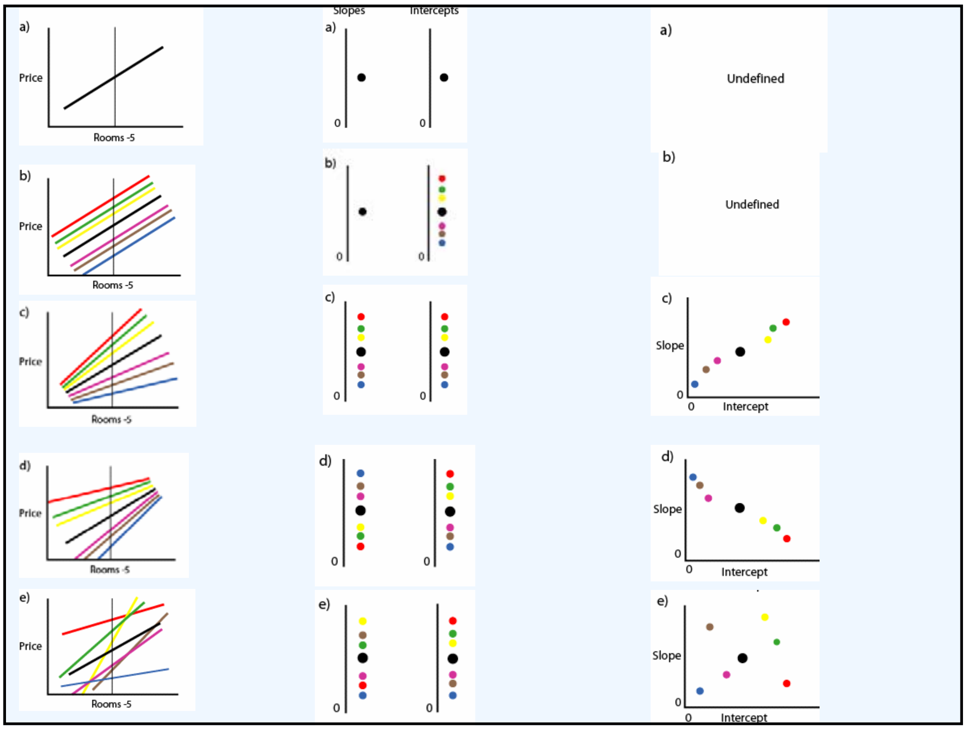
- **Single response:** house price
- **Single predictor**
  - size of house, number of rooms

Rooms	1	2	3	4	5	6	7	8
	-4	-3	-2	-1	0	1	2	3

- **Two level hierarchy**
  - houses at level 1 nested within
  - neighbourhoods at level 2 are the contexts

Set of characteristic plots………………


a)
Price
Rooms -5
b)
Price
Rooms -5
c)
Price
Rooms -5
d)
Price
Rooms -5
e)
Price
Rooms -5


a)
Price
Rooms -5
b)
Price
Rooms -5
c)
Price
Rooms -5
d)
Price
Rooms -5
e)
Price
Rooms -5
Slopes
Intercepts
a)
0
0
b)
0
0
c)
0
0
d)
0
0
e)
0
0
a)
Undefined
b)
Undefined
c)
Slope
0
0
Intercept
d)
Slope
0
0
Intercept
e)
Slope
0
0
Intercept

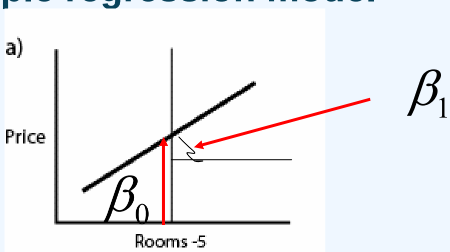
General Structure for Statistical models

- Response = general trend + fluctuations
- Response = systematic component + stochastic element
- Response = fixed + random

- Specific case: the single level simple regression model

Response		***Systematic Part***				***Random Part***
House Price	=	Price of average-sized house	+	Cost of extra room	+	house residual variation
		Intercept		***Slope***		***Residual***

Simple regression model



y is the outcome, price of a house

x_1 is the predictor, number of rooms, which we shall deviate around its mean, 5

Rooms	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
x_1	**-4**	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**

Simple regression model (cont)

$$y_i = \beta_0 + \beta_1 x_{1i} + (e_i)$$

y_i **is the price of house i**

x_1 **is the individual predictor variable**

β_0 **is the intercept;** β_1 **is the fixed slope term:**

e_i **is the residual/random term, one for every house**

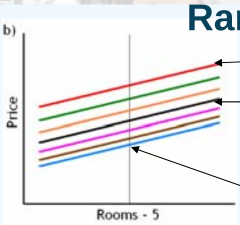
Summarizing the random term: ASSUME IID
Mean of the random term is zero
Constant variability (Homoscedasticy)
No patterning of the residuals (*i.e*, they are independent)

$$e_i \sim N(0, \sigma_e^2)$$

σ_e^2 **between house variance; conditional on size**

ESRC National Centre for Research Methods
Learning Environment for Multilevel Methodology and Applications
E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Random intercepts model



Premium $+u_{0j}$

Citywide line $\hat{Y} = \beta_0 + \beta_1 x_1$

Discount $-u_{0j}$

Differential shift for each district j : index the intercept

Micro-model

$$y_{ij} = \beta_{0j} + \beta_1 x_{1ij} + e_{ij}$$

Macro-model: index parameter as a response

$$\beta_{0j} = \beta_0 + u_{0j}$$

Price of average = citywide + differential for
district j price district j

Substitute macro into micro…….

Random intercepts COMBINED model

Substituting the macro model into the micro model yields

$$y_{ij} = (\beta_0 + u_{0j}) + \beta_1 x_{1ij} + e_{ij}$$

Grouping the random parameters in brackets

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + (u_{0j} + e_{ij})$$

- Fixed part $\beta_0 + \beta_1$
- Random part (Level 2) $u_{0j} \sim N(0, \sigma^2_{u0})$
- Random part (Level 1) $e_{0ij} \sim N(0, \sigma^2_{e0})$
  - District and house differentials are independent $Cov[u_{0j}, e_{0ij}] = 0$

The meaning of the random terms

- Level 2 : between districts

$[u_{0j}] \sim N(0, \sigma^2_{u0})$

σ^2_{u0}

- Between district variance conditional on size

- Level 1 : within districts between houses

$[e_{0ij}] \sim N(0, \sigma^2_{e0})$

σ^2_{e0}

- Within district, between-house variance conditional on size

Variants on the same model

- **Combined model**

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + (u_{0j} + e_{ij})$$

- **Combined model in full**

$$y_{ij} = \beta_0 x_{0ij} + \beta_1 x_{1ij} + (u_{0j} x_{0ij} + e_{0ij} x_{0ij})$$

x_{0ij} • **Is the constant ; a set of 1's**

- **In MLwiN**

$y_{ij} \sim \mathrm{N}(XB, \Omega)$

$y_{ij} = \beta_{0ij} x_0 + \beta_1 x_{1ij}$

$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$

$\left[u_{0j}\right] \sim \mathrm{N}(0, \Omega_u) : \Omega_u = \left[\sigma^2_{u0}\right]$

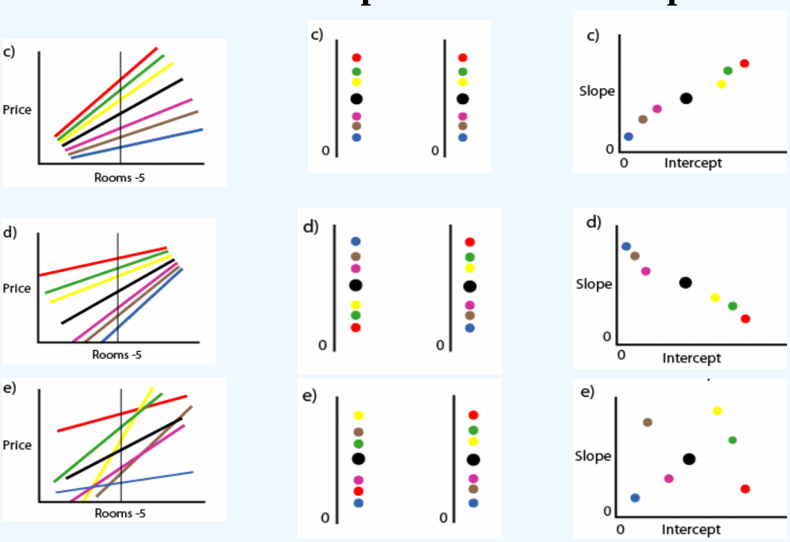
$\left[e_{0ij}\right] \sim \mathrm{N}(0, \Omega_e) : \Omega_e = \left[\sigma^2_{e0}\right]$

*-2*loglikelihood(IGLS Deviance)* = 9357.242(4059 of 4059 cases in use)

Differentials at each level

ESRC National Centre for Research Methods
Learning Environment for Multilevel Methodology and Applications
E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Random intercepts and random slopes



Random intercepts and slopes model

Micro-model

$$y_{ij} = \beta_{0j}x_{0ij} + \beta_{1j}x_{1ij} + e_{0ij}x_{0ij}$$

Note: Index the intercept and the slope associated with a constant, and number of rooms, respectively

Macro-model (Random Intercepts)

$$\beta_{0j} = \beta_0 + u_{0j}$$

Macro-model (Random Slopes)

$$\beta_{1j} = \beta_1 + u_{1j}$$

Slope for district j = citywide slope + differential slope for district j

Substitute macro models into micro model…………

Random slopes model

Substituting the macro model into the micro model yields

$$y_{ij} = (\beta_0 + u_{0j})x_{0ij} + (\beta_1 + u_{1j})x_{1ij} + e_{0ij}x_{0ij}$$

Multiplying the parameters with the associated variable and grouping them into fixed and random parameters yields the combined model:

$$y_{ij} = \beta_0 x_{0ij} + \beta_1 x_{1ij} + (u_{0j}x_{0ij} + u_{1j}x_{1ij} + e_{0ij}x_{0ij})$$

Characteristics of random intercepts & slopes model

$$y_{ij} = \beta_0 x_{0ij} + \beta_1 x_{1ij} + (u_{0j} x_{0ij} + u_{1j} x_{1ij} + e_{0ij} x_{0ij})$$

Fixed part β_0 and β_1

Random part (Level 2) $\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u0u1} & \sigma^2_{u1} \end{bmatrix})$

Random part (Level 1) $e_{0ij} \sim N(0, \sigma^2_{e0})$

Interpreting varying relationship plot through mean and variance-covariances

	Intercepts: terms associated with **Constant** x_0		**Slopes:** terms associated with **Predictor** x_1		**Intercept/Slope:** terms associated with $x_0 x_1$
Graph	**Mean**	**Variance**	**Mean**	**Variance**	**Covariance**
	β_0	σ^2_{u0}	β_1	σ^2_{u1}	σ_{u0u1}
A	+	**0**	+	**0**	**undefined**
B	+	+	+	**0**	**undefined**
C	+	+	+	+	+
D	+	+	+	+	-
E	+	+	+	+	**0**

Random intercepts and slopes model in MLwiN

$$y_{ij} \sim \mathrm{N}(XB, \Omega)$$

$$y_{ij} = \beta_{0ij} x_0 + \beta_{1j} x_{1ij}$$

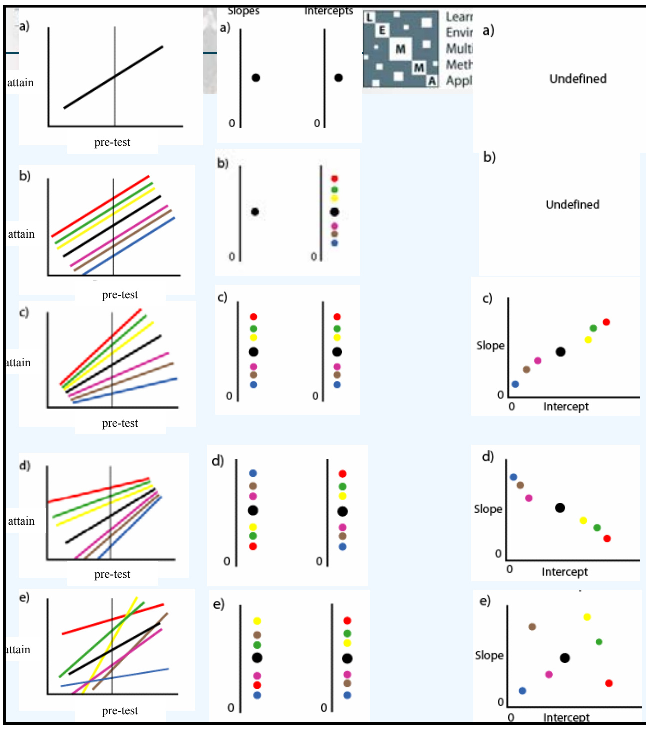
$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim \mathrm{N}(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u10} & \sigma^2_{u1} \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \end{bmatrix} \sim \mathrm{N}(0, \Omega_e) : \Omega_e = \begin{bmatrix} \sigma^2_{e0} \end{bmatrix}$$

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL



$$\begin{cases} y_i = \beta_0 + \beta_1 x_i \\ e_i \sim N(0, \sigma_e^2) \end{cases}$$

$$\begin{cases} y_{ij} = \beta_{0j} + \beta_1 x_{ij} + e_{ij} \\ \beta_{0j} = \beta_0 + u_{0j} \\ u_{0j} \sim N(0, \sigma^2_{u0}) \\ e_{ij} \sim N(0, \sigma^2_e) \end{cases}$$

$$\begin{cases} y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + e_{ij} \\ \beta_{0j} = \beta_0 + u_{0j} \\ \beta_{1j} = \beta_1 + u_{1j} \\ \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma^2_{u0} & \\ \sigma_{u01} & \sigma^2_{u1} \end{bmatrix} \\ e_{ij} \sim N(0, \sigma^2_e) \end{cases}$$

Type of questions tackled by ML: fixed AND random effects

- Even with only 'simple' hierarchical 2-level structure
- EG 2-level model: current attainment given prior attainment of pupils(1) in schools(2)
- Do Boys make greater progress than Girls (**F**: ie averages)
- Are boys more or less variable in their progress than girls? (**R**: modelling variances)
- What is the between-school variation in progress? (**R**)
- Is School X different from other schools in the sample in its effect? (**F**)……….

Type of questions tackled by ML cont.

- Are schools more variable in their progress for pupils with low prior attainment? (**R**)
- Does the gender gap vary across schools? (**R**)
- Do pupils make more progress in denominational schools? (**F**)) (correct SE's)
- Are pupils in denominational schools less variable in their progress? (**R**)
- Do girls make greater progress in denominational schools? (**F**) (cross-level interaction) (correct SE's)

More generally a focus on variances: segregation, inequality are all about differences *between* units

3 Varying relations & random effects:

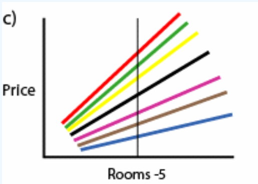
A Demonstration using MLwiN

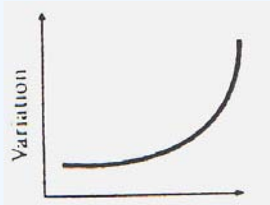
4 Variance functions: the core of random coefficient modelling

What are Variance Functions?

Functions that structuring the variance in terms of other variables

Example: between-district variance in price as a variance function of the size of the house





size

Regression: models many means (the fixed part) but single overall variance, called the error term and treated as disturbances

Multilevel: explicit modelling of (complex) variances, that is modelling heterogeneity

WHY?

ESRC National Centre for Research Methods
Learning Environment for Multilevel Methodology and Applications
E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

WHY Use variance functions?

Substantively

- Some research questions are about variances as well as means: segregation, inequality are all about differences *between* units; lots of such work remains descriptive

- Impact heterogeneity: eg 2 interventions that reduce the mean of the problem, but one is preferred becomes it narrows the variance; ie more consistent in effect.

- Impact heterogeneity eg 2 interventions that have same mean effect but one is more consistent for men, the other for women; the variance is structured by gender

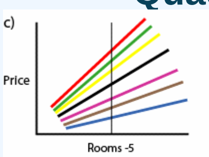
- First lesson in data analysis; middle **and** scatter around the middle; yet in modelling: homoscedastic assumption, ie all variation consigned to one term and labelled error!

WHY Use variance functions?

Technically

- Gives correct estimate of standard errors in the presence of heteroscedasticity
- Multilevel models are about variance functions at higher levels; eg between district variance in terms of price BUT can get confounding across levels; need explicit modelling of variance at all levels
- Example: between-district variation in price in terms of size, requires model simultaneously between-house variation in terms of size
- Example: detached houses are more expensive than non-detached (fixed), have bigger differences between neighbourhoods (level 2) and bigger differences between houses (level 1)

Random Intercepts and Random Slopes Model: Quadratic variance function at level-2



$$y_{ij} = \beta_0 x_{0ij} + \beta_1 x_{1ij} + (u_{0j} x_{0ij} + u_{1j} x_{1ij} + e_{0ij} x_{0ij})$$

- Total variance at level-2, sum of **TWO** random terms: u_{0j} u_{1j}

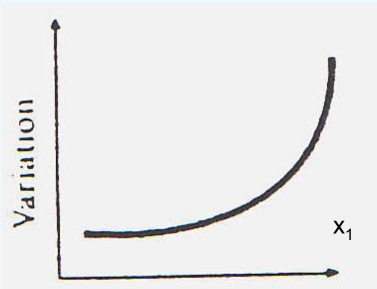
$$Var(u_{0j} + u_{1j}) = \sigma^2_{u0} x^2_{0ij} + 2\sigma_{u0u1} x_{0ij} x_{1ij} + \sigma^2_{u1} x^2_{1ij}$$

- Clearly (if random slopes are required) results in a quadratic function at level-2

Between district-heterogeneity as increasing QUADRATIC function of size

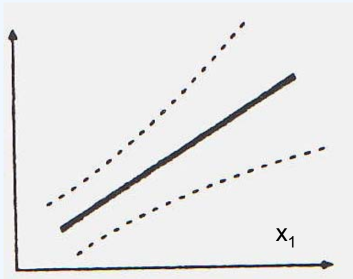
$$Var(u_{0j} + u_{1j}) = \sigma^2_{u0} x^2_{0ij} + 2\sigma_{u0u1} x_{0ij} x_{1ij} + \sigma^2_{u1} x^2_{1ij}$$

- Variance function by Size

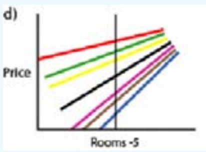


- Population bounds for districts

y

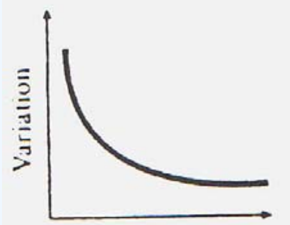


Between district-heterogeneity as decreasing QUADRATIC function of size



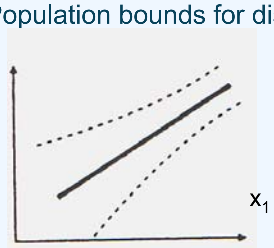
$$Var(u_{0j} + u_{1j}) = \sigma^2_{u0} x^2_{0ij} + 2(-\sigma_{u0u1}) x_{0ij} x_{1ij} + \sigma^2_{u1} x^2_{1ij}$$

- Variance function by Size



x_1

- Population bounds for districts

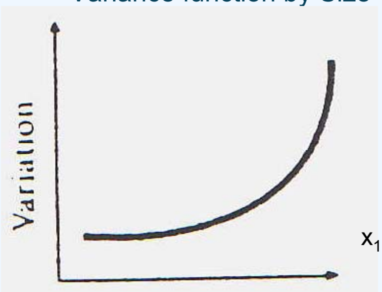


Between house-heterogeneity as increasing quadratic function of size

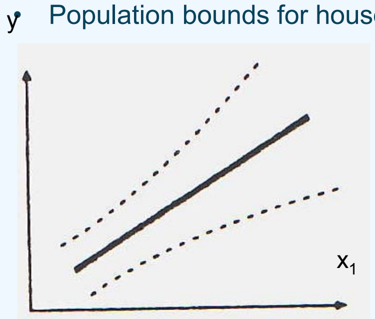
- Micro model $y_{ij} = \beta_{0j} x_{0ij} + \beta_{1j} x_{1ij} + (e_{0ij} x_{0ij} + e_{1ij} x_{1ij})$

$$Var(e_{0ij} + e_{1ij}) = \sigma_{e0}^2 x_{0ij}^2 + 2\sigma_{e0e1} x_{0ij} x_{1ij} + \sigma_{e1}^2 x_{1ij}^2$$

- Variance function by Size



- Population bounds for houses

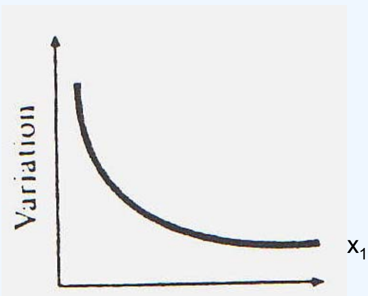


Between house-heterogeneity as decreasing quadratic function of size

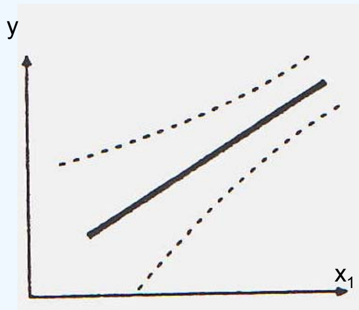
- Micro model $y_{ij} = \beta_{0j} x_{0ij} + \beta_{1j} x_{1ij} + (e_{0ij} x_{0ij} + e_{1ij} x_{1ij})$

$$Var(e_{0ij} + e_{1ij}) = \sigma_{e0}^2 x_{0ij}^2 + 2(-\sigma_{e0e1}) x_{0ij} x_{1ij} + \sigma_{e1}^2 x_{1ij}^2$$

- Variance function by Size



- Population bounds for houses



Variance Partitioning Coefficient

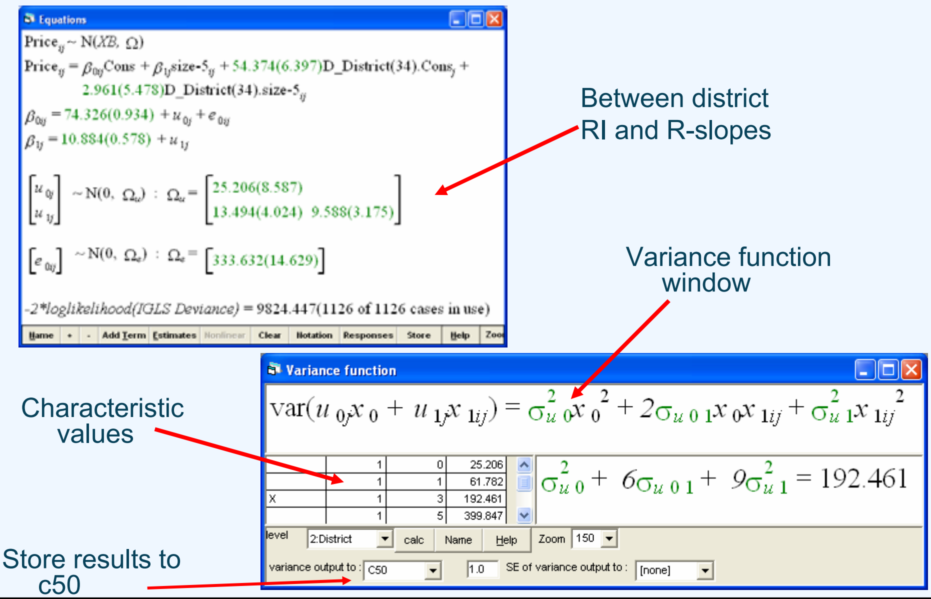
- 1 the percentage of the variance that lies between districts
- 2 the degree of similarity with district; the degree of dependency AKA the intra-class correlation

Given by between-district variation divided by the (between-house variation plus between-district variation)

$$\rho = \frac{\sigma^2_{u0} + 2\sigma_{u0u1}x_{1ij} + \sigma^2_{u0}x^2_{1ij}}{\sigma^2_{e0} + 2\sigma_{e01}x_{1ij} + \sigma^2_{e1}x^2_{1ij} + \sigma^2_{u0} + 2\sigma_{u0u1}x_{1ij} + \sigma^2_{u0}x^2_{1ij}}$$

ESRC National Centre for Research Methods
Learning Environment for Multilevel Methodology and Applications
ESRC Economic & Social Research Council

In MLwiN (starting with model 4)



Between district RI and R-slopes

Variance function window

Characteristic values

Store results to c50

Model 5

Quadratic variance function at both levels

Equations

$$\text{Price}_{ij} \sim \text{N}(XB, \Omega)$$

$$\text{Price}_{ij} = \beta_{0ij}\text{Cons} + \beta_{1ij}\text{size-5}_{ij} + 54.704(6.184)\text{D\_District(34).Cons}_j + 3.204(4.769)\text{D\_District(34).size-5}_{ij}$$

$$\beta_{0ij} = 74.084(0.909) + u_{0j} + e_{0ij}$$

$$\beta_{1ij} = 10.536(0.559) + u_{1j} + e_{1ij}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim \text{N}(0,\ \Omega_u) \ : \ \Omega_u = \begin{bmatrix} 27.000(8.056) & \\ 10.045(3.798) & 8.656(3.032) \end{bmatrix}$$

$$\begin{bmatrix} e_{0ij} \\ e_{1ij} \end{bmatrix} \sim \text{N}(0,\ \Omega_e) \ : \ \Omega_e = \begin{bmatrix} 211.777(13.686) & \\ 32.953(4.439) & 26.520(4.710) \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = 9677.463(1126 of 1126 cases in use)

Name | + | - | Add Term | Estimates | Nonlinear | Clear | Notation | Responses | Store | Help | Zoom

Model 4 and 5 comparison

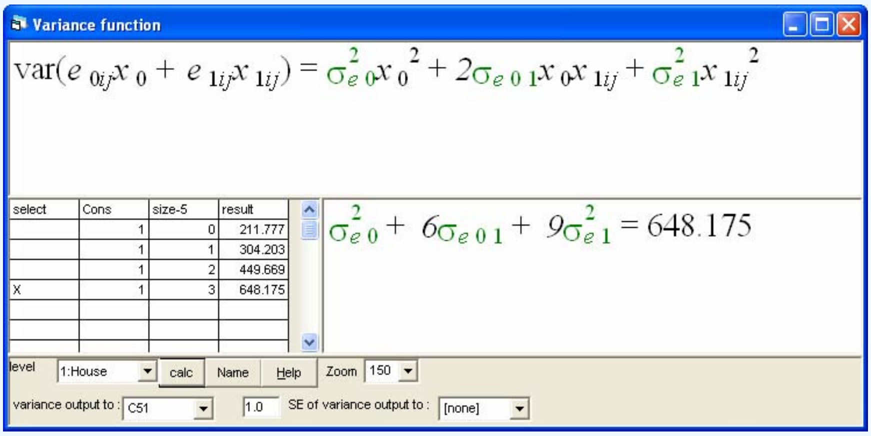
Results Table

Copy

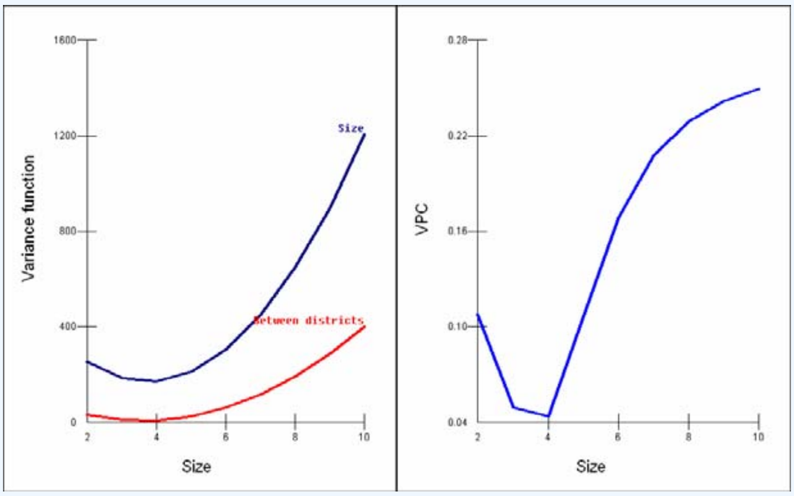
	Model 4	Standard Error	Model 5	Standard Error
Response	Price		Price	
Fixed Part				
Cons	74.326	0.934	74.084	0.909
size-5	10.884	0.578	10.536	0.559
D_District(34).C	54.374	6.398	54.704	6.184
D_District(34).s	2.961	5.478	3.204	4.769
Random Part				
Level: District				
Cons/Cons	25.213	8.575	27.000	8.056
size-5/Cons	13.494	4.023	10.045	3.798
size-5/size-5	9.586	3.178	8.656	3.032
Level: House				
Cons/Cons	333.632	14.630	211.777	13.686
size-5/Cons			32.953	4.439
size-5/size-5			26.520	4.710
-2*loglikelihood	9824.447		9677.463	
DIC:				
Units: District	50		50	
Units: House	1126		1126	

Changed level 2
Variance function window

Variance function at level 1

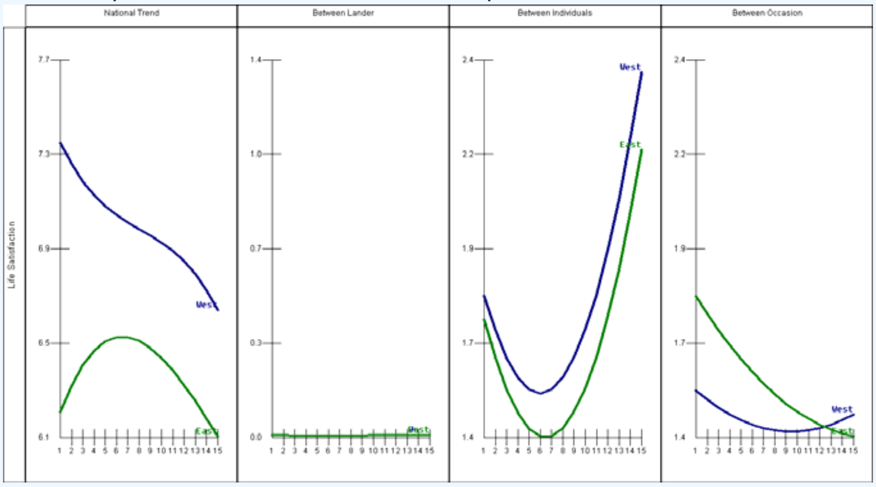


Two-level hierarchical model : houses districts



Exemplifying variance heterogeneity

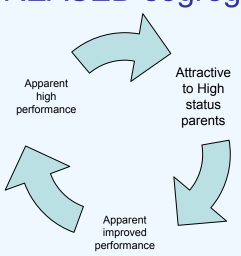
- How has life satisfaction changed in Germany 1991-2006?
- Structure: 15 occasions for 16k individuals in 16 lander
- Response: Life satisfaction on a 10 point score

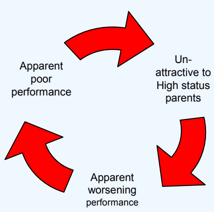


Motivation: are we become a segregated society? EG in relation to schools

Virtuous and Vicious circles

Following 1988 Education Reform Act with emphasis on choice, league tables, competition expectation of INCREASED segregation





Choice → increased polarization in terms of ability

Choice → increased polarization in terms of socio-economic background; poverty; ethnicity etc

Research Questions

- Has ***school*** FSM segregation increased?
- Has ***LEA*** segregation increased?
- Has segregation been ***differential*** between types of schools and types of LEA's

Approach: ***model*** the variance not calculate an Index

π_{jk}

Anatomy of a simple model

Dependent variable: *observed* FSM or not, in 2001 for pupil i in school j

Distributed as a Binomial variable with a denominator equal to no of pupils in each school, with an underlying propensity of having a FSM, π

Model Log-odds of propensity

$$\log e\left(\frac{\pi}{1-\pi}\right)$$

$$\text{fsm01}_{ij} \sim \text{Binomial}(\text{denom01}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j}\text{cons}$$

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$\left[u_{0j}\right] \sim \text{N}(0, \Omega_u) : \Omega_u = \left[\sigma^2_{u0}\right]$$

$$\text{var}(\text{fsm01}_{ij}|\pi_{ij}) = \pi_{ij}(1 - \pi_{ij})/\text{denom01}_{ij}$$

As an underlying average β_0

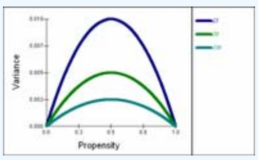
& allowed to vary school difference u_{0j}

With a variance of σ^2_{u0}

School differences assumed to come from a Normal distribution

KEY measure of segregation; between-school variance on logit scale; if assumption met, complete summary

Between pupil variance: allows for stochastic fluctuations determined by n and π

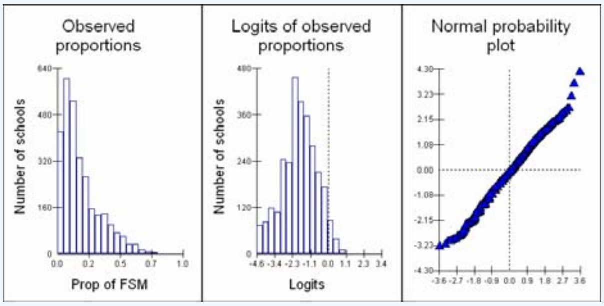


Results from simple model

$$\text{fsm01}_{ij} \sim \text{Binomial}(\text{denom01}_{ij}, \pi_{ij})$$
$$\text{logit}(\pi_{ij}) = \beta_{0j}\text{cons}$$
$$\beta_{0j} = -1.84(0.02) + u_{0j}$$
$$\left[u_{0j}\right] \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[1.18(0.03)\right]$$
$$\text{var}(\text{fsm01}_{ij}|\pi_{ij}) = \pi_{ij}(1 - \pi_{ij})/\text{denom01}_{ij}$$

Logit: -1.84 when transformed median of 0.137 (95% CI's 0,133 and 0.142); and mean of 0.182 (0.177 and 0.187)

"Significant" between school segregation; Equivalent to a D of 0.374 (see next slide)

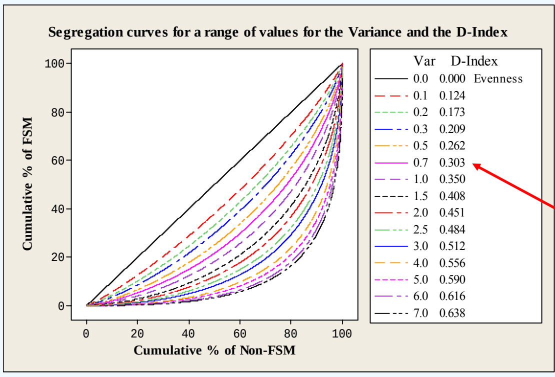


Distributional assumptions for school differences

ESRC National Centre for Research Methods
Learning Environment for Multilevel Methodology and Applications
E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Linking models to indexes

- Using model parameters we can derive expected values of any function of underlying school probabilities
- Consequently, derive index by simulation from model parameters.



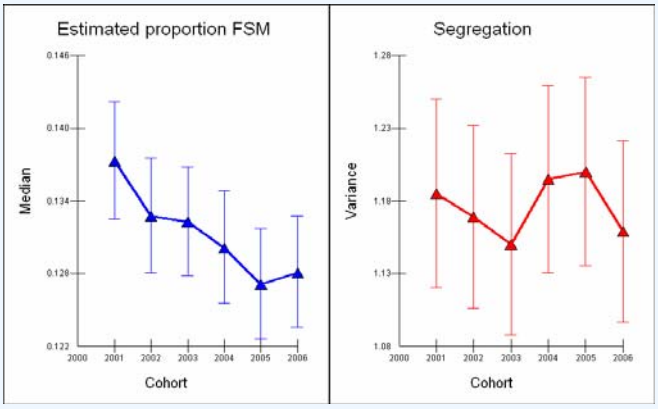
EG: Converting logit Variance to D

(simulate 500k Logits with a given underlying mean and variance; convert to proportions, and calculate Index)

Variance of 0.7 ***equals*** D-Index of 0.30

ESRC National Centre for Research Methods | LEMMA Learning Environment for Multilevel Methodology and Applications | E·S·R·C Economic & Social Research Council

Results for simple model repeated for each entry cohort 2001-2006



Median: small improvement

Segregation: changes smaller than uncertainty

ESRC National Centre for Research Methods | LEMMA Learning Environment for Multilevel Methodology and Applications | E·S·R·C Economic & Social Research Council

Three-level model: partitioning between LA, and between school variance

$$\mathrm{FSM01}_{ijk} \sim \mathrm{Binomial}(\mathrm{den01}_{ijk}, \pi_{ijk})$$
$$\mathrm{logit}(\pi_{ijk}) = \beta_{0jk}\,\mathrm{cons}$$
$$\beta_{0jk} = \beta_0 + v_{0k} + u_{0jk}$$
$$\left[v_{0k}\right] \sim \mathrm{N}(0,\ \Omega_v) \ :\ \Omega_v = \left[\sigma^2_{v0}\right]$$
$$\left[u_{0jk}\right] \sim \mathrm{N}(0,\ \Omega_u) \ :\ \Omega_u = \left[\sigma^2_{u0}\right]$$
$$\mathrm{var}(\mathrm{FSM01}_{ijk}|\pi_{ijk}) = \pi_{ijk}(1 - \pi_{ijk})/\mathrm{den01}_{ijk}$$

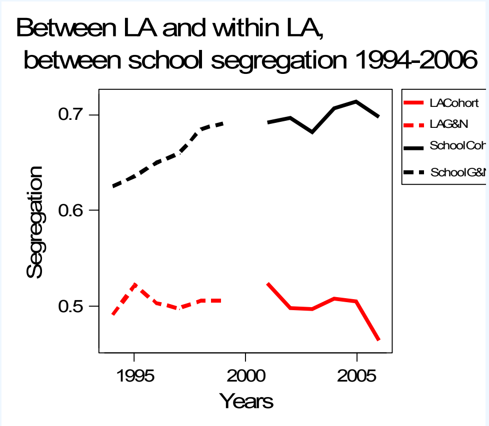
3 Changes

- Pupils (i) in schools (j) In LA's (3)
- Average + LA difference + School difference
- Between LA difference
- Within LA, between school

Modelling at two scales *simultaneously*

Results for 3 level model

- 3 level model applied to each cohort separately
- compared with Goldstein and Noden (earlier and overall school and not entry cohort)



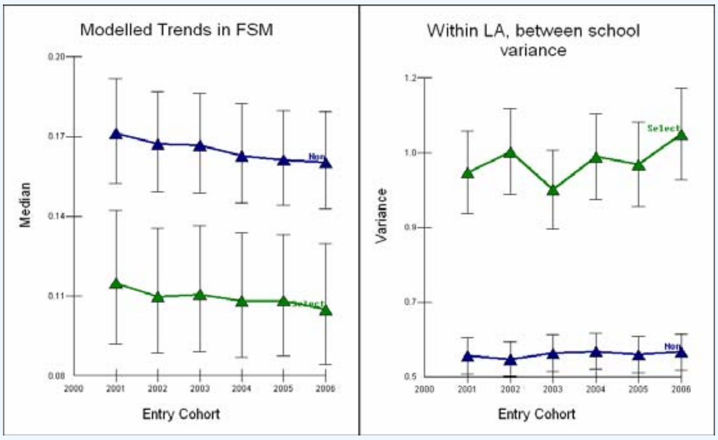
- Greater segregation between schools than between LA's
- LA's: trendless fluctuations
- Continued increasing between-school segregation

Area characteristics 1

- Are LA's that are selective (Grammar/Secondary) more segregated than totally Comprehensive systems?
- 3 level model, with a different variance for schools within different LA characteristics

$$\mathrm{fsm01}_{ijk} \sim \mathrm{Binomial}(\mathrm{denom01}_{ijk}, \pi_{ijk})$$
$$\mathrm{logit}(\pi_{ijk}) = \beta_{1j}\mathrm{Non}_k + \beta_{2j}\mathrm{Select}_k + v_{0k}\mathrm{cons}$$
$$\beta_{1j} = \beta_1 + u_{1jk}$$
$$\beta_{2j} = \beta_2 + u_{2jk}$$
$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_v) \ :\ \Omega_v = \begin{bmatrix} \sigma^2_{v0} \end{bmatrix}$$
$$\begin{bmatrix} u_{1jk} \\ u_{2jk} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u) \ :\ \Omega_u = \begin{bmatrix} \sigma^2_{u1} & \\ 0 & \sigma^2_{u2} \end{bmatrix}$$
$$\mathrm{var}(\mathrm{fsm01}_{ijk}|\pi_{ijk}) = \pi_{ijk}(1 - \pi_{ijk})/\mathrm{denom01}_{ijk}$$

- Average FSM
  - for English pupils living in a ***non-*selecting** LA
  - for English pupils living in a ***selecting*** LA
- Between LA variance
- Within LA
  - between school variance for schools located in a ***non-selecting*** LA
  - between school variance for schools located in a ***selecting*** LA



- Pupils going to school in Selecting LA's are less likely to be in poverty
- Slight decline in poverty in both types of area

- Schools in Selecting areas are more segregated
- Slight evidence of an increase

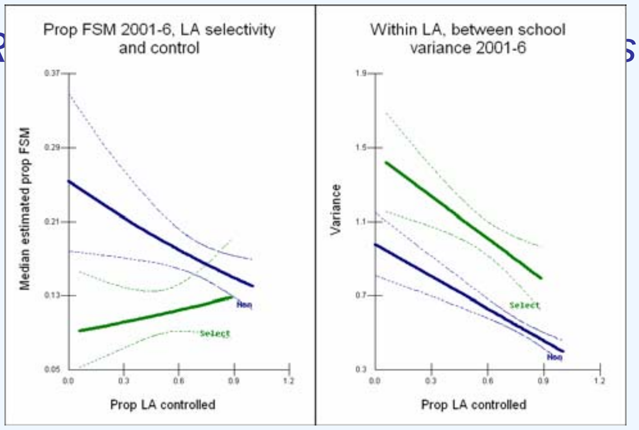
Area characteristics 2

- Is there more segregation in areas that are selective and where less schools are under LA ***control*** in terms of ***admission policies***?
- Variance function for Selective/Non-selective, ***structured*** by the proportion of pupils in an LA who go to Community or Voluntary Controlled schools (contra Voluntary Aided,Foundation, CTC's, Academies)

$$\text{fsm2001-6}_{ijk} \sim \text{Binomial}(\text{denom01-6}_{ijk}, \pi_{ijk})$$
$$\text{logit}(\pi_{ijk}) = \beta_{1j}\text{NoSel} + \beta_{2j}\text{nonprop}_k + \beta_{3j}\text{Sel}_k + \beta_{4j}\text{selprop}_k + v_{0k}\text{cons}$$
$$\beta_{1j} = \beta_1 + u_{1jk}$$
$$\beta_{2j} = \beta_2 + u_{2jk}$$
$$\beta_{3j} = \beta_3 + u_{3jk}$$
$$\beta_{4j} = \beta_4 + u_{4jk}$$
$$\begin{bmatrix} v_{0k} \end{bmatrix} \sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} \sigma^2_{v0} \end{bmatrix}$$
$$\begin{bmatrix} u_{1jk} \\ u_{2jk} \\ u_{3jk} \\ u_{4jk} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma^2_{u1} & & & \\ \sigma_{u12} & \sigma^2_{u2} & & \\ 0 & 0 & \sigma^2_{u3} & \\ 0 & 0 & \sigma_{u34} & \sigma^2_{u4} \end{bmatrix}$$
$$\text{var}(\text{fsm2001-6}_{ijk} | \pi_{ijk}) = \pi_{ijk}(1 - \pi_{ijk})/\text{denom01-6}_{ijk}$$

FSM over the period 2001-6

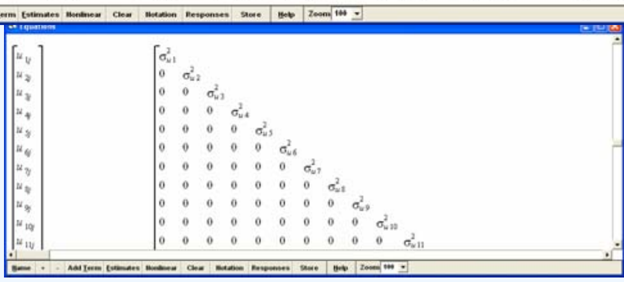
- Average FSM in selecting and ***non***-selecting LA's and how this changes with degree of LA control
- Between LA variance
- Within LA between schools
  - variance function for ***non-selecting*** LA
  - variance function for ***selecting*** LA



- Pupils going to school in Non-Selecting LA's with low LA control are more likely to be in poverty
- Schools in Selecting areas are more segregated
- Segregation decreases with greater LA control for both types of LA

Area characteristics 3

- Which of England's LA's have the most segregated school system?
- Model with 144 averages and 144 variances, one for each LA!



LA's with highest segregation

(not including estimates less than 2* SE)

LA	Variance	D equiv Index	Median prop FSM 2001-6	Select	Prop LA control
Buckinghamshire	2.12	0.46	0.03	Select	0.77
Southend-on-Sea	1.92	0.45	0.09	Select	0.21
Slough	1.76	0.43	0.11	Select	0.37
Trafford	1.75	0.43	0.08	Select	0.40
Oldham	1.72	0.43	0.18	Non	0.75
Calderdale	1.59	0.42	0.12	Select	0.32
Sutton	1.50	0.41	0.05	Select	0.39
Telford &Wrekin	1.46	0.40	0.15	Select	0.53
Solihull	1.42	0.40	0.08	Non	0.85
Barnet	1.42	0.40	0.16	Select	0.41
Knowsley	1.38	0.40	0.34	Non	0.67
Wirral	1.38	0.40	0.18	Select	0.74
Milton Keynes	1.36	0.39	0.12	Non	0.43
Croydon	1.30	0.39	0.16	Non	0.31
Stockton-on-Tees	1.29	0.39	0.16	Non	0.69

Benefits of multilevel approach

- Explicit and separate modelling of trends and segregation
- Separate modelling of segregation at any level: eg increasing LEA (local economy?), but decreasing School (admission policies?)
- Segregation for different types of schools and different types of areas
- Explicit modelling of binomial fluctuations
- Confidence intervals and significance testing

5 Logit models: an example of a non-linear multilevel model

- *Generalised models*
- *Multilevel logit*
- *Extra-binomial variation*
- *Estimation: quasi-likelihood and MCMC*
- *VPC*
- *Population average and subject specific estimates*
- *Application: teenage employment in Glasgow*

Introduction to ***generalized*** multilevel models

So far: Response is linearly related to predictors and all random terms are assumed Normal

- **Now** a range of other data types for the response
  All can be handled routinely by MLwiN

- **Achieved by 2 aspects**
  a non-linear link between response and predictors
  a non-Gaussian level 1 distribution

Response	Example	Model
Binary Categorical	Yes/No	Logit or probit or log-log model with binomial L1 random term
Proportion	Proportion un-employed	Logit etc. with binomial L1 random term
Multiple categories	Travel by train, car, foot	Logit model with multi-nomial random term; can handle ordered and unordered categories
Count	No of crimes in area	Log model with L1 Poisson random term; can include an Offset
Count	LOS	Log model with L1 NBD random term; can include an Offset

Focus on specifying **binomial** multilevel models with response that is either binary or a proportion

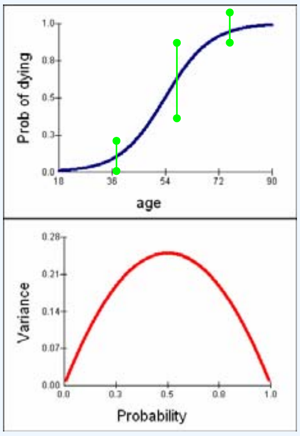
Implementation in MLwiN

Reference:

Subramanian S V, Duncan C, Jones K (2001) Multilevel perspectives on modeling census data *Environment and Planning A* **33**(3) 399 – 417

Modelling proportions

- **Proportions:** employment rate; conceived as the underlying of being employed
- **Linear probability model**: that is use standard regression model with linear relationship and Gaussian random term
- **But 3 problems**
  - (1) Nonsensical predictions: predicted proportions are unbounded, outside range of 0 and 1
  - (2) Anticipated non-linearity as approach bounds
  - (3) Heterogeneity: inherent unequal variance; dependent on mean (ie fixed part) and on denominator
- Logit model with Binomial random term **resolves** all three problems (could use probit, clog-clog)



The multilevel logistic RI model

- The underlying probability or proportion is non-linearly related to the predictor(s)

$$\pi = \frac{e^{\beta_0 + \beta_1 x_1 + \ldots + u_{0j}}}{1 + e^{\beta_0 + \beta_1 x_1 + \ldots + u_{0j}}}$$

where *e* is the base of the natural logarithm

- linearized by the logit transformation

$$\log e\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \ldots + u_{0j};$$
$$u_{0j} \sim N(0, \sigma^2_{u0})$$

- The logit transformation produces a linear function of the parameters; bounded between 0 and 1; thereby solving problems 1 and 2
- NB parameters on the logit scale

Solving problem 3:assume Binomial variation for Level-1 random part

- Variance of the response is presumed Binomial:

$$Var(y \mid \pi) = \frac{\pi(1-\pi)}{n}$$

Ie *Observed proportion* depends on underlying proportion and the denominator

- Achieved in practice by replacing the constant variable at level 1 by a binomial weight, z, and constraining the level-1 variance to 1 for exact binomial variation
- The random (level-1) component can be written as

$$y_i = \pi_i + e_i z_i, \quad z_i = \sqrt{\frac{\hat{\pi}_i(1-\hat{\pi}_i)}{n_i}}, \quad \sigma^2_{ei} = 1$$

Moving between Proportions, Odds and Logits

	Proportion/Probability	Odds
A	5 out of 10	5 to 5
B	6 out of 10	6 to 4
C	8 out of 10	8 to 2

	Proportion (p)	Odds (p/1-p)	Log of odds Loge (p/1-p)
A	0.5	1.0	0
B	0.6	1.5	0.41
C	0.8	4	1.39

MLwiN: Logit calculation

	Logit	Odds
A	e^{0}	1.0
B	$e^{0.41}$	1.5
C	$e^{1.39}$	4

MLwiN: Expo calculation

	Logit	Proportion
A	$e^{0}/(1+ e^{0})$	0.5
B	$e^{0.41}/(1+ e^{0.41})$	0.6
C	$e^{1.39}/(1+ e^{1.39})$	0.8

MLwiN: Alogit calculation

Binomial and Extra-binomial variation

- Binomial variation is all due to underlying probability & values of the denominator, level 1 random variation is not freely estimated
- More general approach: allow the variance to be estimated from the data; (un-constrain the parameter $\sigma^2_{ei} \neq 1$)
- **Over-dispersion**: (more than 1) unexplained variation in the response that is not adequately modeled by the fixed means
- **Under-dispersion**: (Less than 1); suggesting greater dependency that that expected from a binomial assumption; suggest missing an important level in the model structure
- Can only be done with proportions NOT binary data
- Application: French KM and Jones K (2006) Impact of definition on the study of avoidable mortality **Social Science & Medicine**, 62 (6), 1443-1456

Estimation

Quasi-likelihood (Marginal Quasi-Likelihood versus Predictive Quasi-Likelihood; 1st and 2nd order)

model linearised and Goldstein's IGLS applied

1st and 2nd order Taylor series expansion (linearise the model)

MQL versus PQL are higher-level effects included in the linearisation

MQL1 crudest approximation. Estimates may be biased downwards (esp. if within cluster sample size is small and between cluster variance is large eg households); but stable.

PQL2 best approximation, but may not converge.

Advice Start with MQL1 to get starting values for 2nd PQL

MCMC methods

good quality estimates even where cluster size is small;

get deviance of model (DIC) for sequential model testing,

Advice: start with MQL1 and then switch to MCMC, then wait!

A simulation study

- To assess the bias of different estimators for binary outcome
- Rodriguez, G, 'Multilevel Generalized Linear Models', in de Leeuw, J and Meijer, E (Eds.) 2008, *Handbook of Multilevel Analysis,* chapter 9, 335-376
- Generated 100 datasets with real structure (2449 births to 1558 mothers living in 161 communities in Guatemala; ie pathological case) with known coefficients, all fixed and random coefficients set equal to 1
- Data and papers available from http://data.princeton.edu/multilevel/

Results from simulation study

Estimation Method	Fixed Parameters Individual	Family	Community	Random Parameters Family	Community
True Value	1.000	1.000	1.000	1.000	1.000
MQL-1	0.738	0.744	0.771	0.100	0.732
MQL-2	0.853	0.859	0.909	0.273	0.763
PQL-1	0.808	0.806	0.831	0.432	0.781
PQL-2	0.933	0.940	0.993	0.732	0.924
5quadrature	0.983	0.988	1.037	0.962	0.981
20quadrature	0.983	0.990	1.039	0.973	0.979
Gibbs	0.971	0.978	1.022	0.922	0.953

Quadrature (eg gllamm in Stata) good quality estimates, but computational burden increases rapidly with the dimensionality of the problem eg with 12 quad points, 3 level RI model requires evaluation of equivalent of 144 likelihoods; BUT 12 point and 3 level And random intercept & slope, equivalent to almost 21,000 likelihoods

Population average and cluster specific estimates

Mlwin gives directly

CLUSTER SPECIFIC Estimates (often called the unit or **subject specific** because of use in repeated measures studies)

- the fixed effects *conditional* on higher level random effects
- effect of change for particular 'individual' of unit change in a predictor

NOT the POPULATION-AVERAGE estimates (Eg GEE)

- ie the marginal expectation of the dependent variables across the population; "averaged " across the random effects
-effect of change in the whole population if everyone's predictor variable changed by one unit

In non-linear models these are different and the PA will generally be smaller than CS, especially as size of random effects grows

Can derive PA from CS but not vice-versa, new version give both by simulation)

- **Median** of the predictive distribution is US
- **Mean** of the predictive distribution is PA

Ritz, J and Donna Spiegelman (2004) Equivalence of conditional and marginal regression models for clustered and longitudinal data, *Statistical Methods in Medical Research*, 13(4), 309-323

Variance Partitioning Coefficient

For 2-level *Normal* response random intercept model:

$$\text{VPC} = \frac{\text{Level 2 variance}}{\text{Level 1 variance} + \text{Level 2 variance}}$$

BUT with logit, level 2 variance is on the logit scale and the level 1 variance is on the probability scale so they can not be directly compared. Also level 1 variance depends on underlying proportion

Possible solutions include
i) set the level 1 variance = variance of a standard logistic distribution:

$$\text{Then VPC} = \frac{\sigma_u^2}{\sigma_u^2 + 3.29}$$

Ignores that the level –1 variance is not constant

Variance Partitioning Coefficient (cont.)

Possible solutions include
ii) Simulation

Step 1: Simulate M (5000 say) values for random effect *u's* with same variance as estimated level 2 variance, $N(0, \hat{\sigma}_u^2)$

Step 2: using these 5000 *u*'s, combine with fixed part estimates and particular values of the predictor variables to get predicted logits; alogit to get probabilities; $\pi^*_{(m)} = [1 + \exp(-(\hat{\beta}^T \mathbf{x}^* + u_{(m)}))]^{-1}$
the variance at level 2 on the probability scale is the variance of these values.

Step 3: calculate a level 1 variance for the 5000 simulations on the probability scale: $v^*_{1(m)} = \pi^*_{(m)}(1 - \pi^*_{(m)})$
take the mean of these values to get overall level 1 variance

Step 4: use the usual VPC formula, now that level 1 and level 2 variances are on the same scale

Browne WJ, Subramanian S V, Jones K, Goldstein H. (2005)Variance partitioning in multilevel logistic models that exhibit overdispersion.*Journal of Royal Statistical Society A*,168(3) 599-613

Application: teenage employment in Glasgow

- "Ungrouped" data that is individual data
- Model binary outcome of employed or not and two individual predictors

Name	Person	District	Employed	Qualif	Sex
Craig	1	1	Yes	Low	Male
Subra	2	1	Yes	High	Male
Nina	3	1	Yes	Low	Fem
Min	4	1	No	Low	Fem
Myles	5	1	No	High	Male
Sarah	12	50	Yes	High	Fem
Kat	13	50	No	Low	Fem
Colin	14	50	Yes	Low	Male
Andy	15	50	No	High	Male

Same data as a multilevel structure: a set of tables for each district

	GENDER			
QUALIF	**MALE**	**FEMALE**	**Postcode**	**UnErate**
LOW	5 out of 6	3 out of 12	G1A	15%
HIGH	2 out of 7	7 out of 9		
LOW	5 out of 9	7 out of 11	G1B	12%
HIGH	8 out of 8	7 out of 9		
LOW	3 out of 3	-	G99Z	3%
HIGH	2 out of 3	out of 5		

- Level 1 — cell in table
- Level 2: — Postcode sector
- Margins: — define the two categorical predictors
- Internal cells: — the response of 5 out of 6 are employed

Turning a table into a model:

		GENDER		Adult
		Male	Female	Unemp rate
Qualif	Low	$p_{1j} = e_{1j} / n_{1j}$	$p_{2j} = e_{2j} / n_{2j}$	w_j
	High	$p_{3j} = e_{3j} / n_{3j}$	$p_{4j} = e_{4j} / n_{4j}$	

For 1 = 1….4 cells, within j postcodes

RESPONSE: Lij = predicted log-odds of employment for type of person i in place j

FIXED

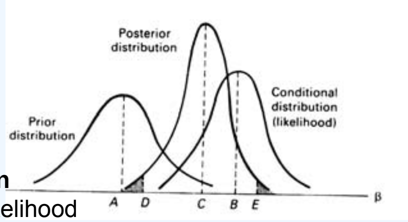
$\beta_0 x_0$	$+\beta_1 x_1$	$+\beta_2 x_2$	$+\beta_3 x_3$
unqualified male Base	Unqualified Female Contrast	qualified male Contrast	qualified female Contrast

- **Level 2 random part : between postcode sectors** $u_{0j} \sim N(0, \sigma^2_{u0})$
- **Level 1 random part: between cells (ie teenagers)** $e_i z_i, \quad z_i = \sqrt{\frac{\hat{\pi}_i(1-\hat{\pi}_i)}{n_i}}, \quad \sigma^2_{ei} = 1$

6 MCMC estimation including Spatial models

- Bayesian approach
- What does MCMC do?
- IGLS versus MCMC
- MCMC in practice in MLwiN

What is a Bayesian approach?



Full probability modelling
- everything treated as a distribution
- parameters & observations as random variables

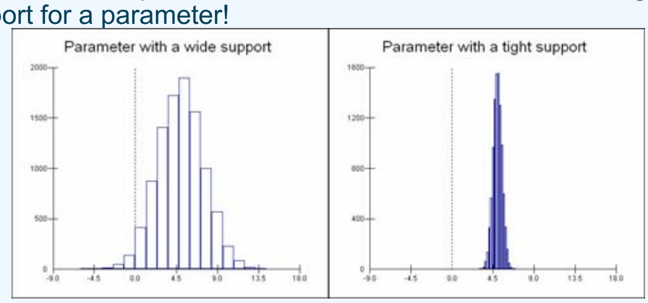
Fundamental theorem: 3 *types* of distribution
Posterior is proportional to the Prior times the Likelihood

- **Prior**
  - belief about a parameter before data; subjectivity! Includes Diffuse
- **Likelihood**
  - estimate of parameter based on data and assumptions
- **Posterior**
  - updated evidence after combining prior and likelihood; distribution**, not** just a point estimate and SE's based on asymptotics

In Practice
- celibate sex counsellor…….. due to complex **joint** posterior when several (maybe 100's) parameters need evaluating, But MCMC

The *Statistical* Holy Grail!

- What is the degree of support for a parameter?
- Normally: infer to the population, get a point estimate, and make Normality assumptions about the distribution of the parameter and calculate SE's
- Problematic assumption eg when small number of groups and variance parameter
- Bayesian solution: simulate from the Posterior distribution either with informative belief or ignorance prior and rely on the data
- The answer is a plot of the distribution which shows the degree of support for a parameter!



What does MCMC do?

What do we want to know? In a two-level, variance components model we have the following unknowns: $\beta, u, \sigma_u^2, \sigma_e^2$

a joint posterior distribution with many dimensions (notice vector of u_j's and β's)

$$p(\beta, u, \sigma_u^2, \sigma_e^2 \mid y)$$

MCMC is a numerical engine for evaluating joint posterior by simulating from each of the marginal distributions in an iterative scheme eg Gibbs

First we assume some starting values for our unknown parameters

$$\beta_{(0)}, u_{(0)}, \sigma_{u(0)}^2, \sigma_{e(0)}^2$$

Then simulate from the following conditional distributions in rotation

$p(\beta \mid y, u_{(0)}, \sigma_{u(0)}^2, \sigma_{e(0)}^2)$ to get $\beta_{(1)}$, then

$p(u \mid y, \beta_{(1)}, \sigma_{u(0)}^2, \sigma_{e(0)}^2)$ to get $u_{(1)}$, then

$p(\sigma_u^2 \mid y, \beta_{(1)}, u_{(1)}, \sigma_{e(0)}^2)$ to get $\sigma_{u(1)}^2$, then finally

$p(\sigma_e^2 \mid y, \beta_{(1)}, u_{(1)}, \sigma_{u(1)}^2)$

IGLS versus MCMC

IGLS	MCMC
Fast to compute	Slower to compute
Deterministic convergence-easy to judge	Stochastic convergence-harder to judge
Uses mql/pql approximations to fit discrete response models which can produce biased estimates in some cases	Does not use approximations when estimating discrete response models, estimates are less biased
In samples with small numbers of level 2 units confidence intervals for level 2 variance parameters assume Normality, which is inaccurate.	In samples with small numbers of level 2 units Normality is not assumed when making inferences for level 2 variance parameters
Cannot incorporate prior information	Can incorporate prior information
Difficult to extend to new models	Easy to extend to new models eg MM, X-class; and functions of parameters; eg ranks

MCMC in practice in MLwiN

1: using maximum likelihood (IGLS: Iterative Generalised Least Squares; Goldstein 1986, *Biometrika*) to get starting values estimates

2 Switch to MCMC use default priors (uninformative) and burn-in for 500 simulations; throw away

3 Monitor for 5000 simulations and check for convergence to the distribution: good is "white noise"; poor is "slow drift"; check prospective diagnostics; check retrospective diagnostics (especially Effective sample size eg need at least 500 of these for key parameters of interest)

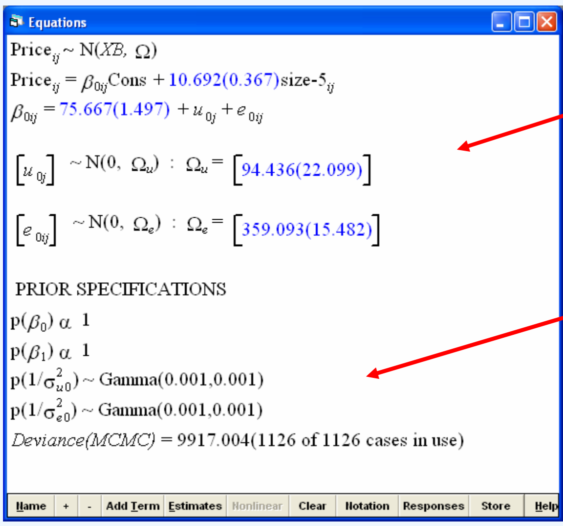
4 Increase monitoring size if suggested insufficient sample size or lack of convergence (ie substantive trending)

5 Report the mean and 95% credible intervals of the parameters and DIC

6 Repeat the above with stronger priors to assess sensitivity

Real-time demonstration (C:talks\aag\houseformcmc.wsz)

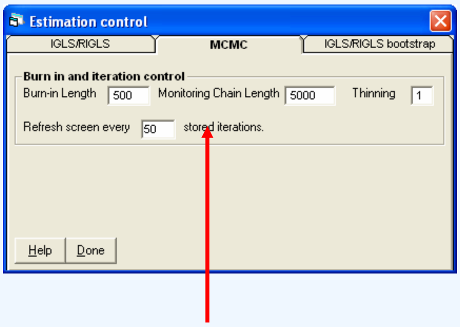
- Two-level hierarchical model :

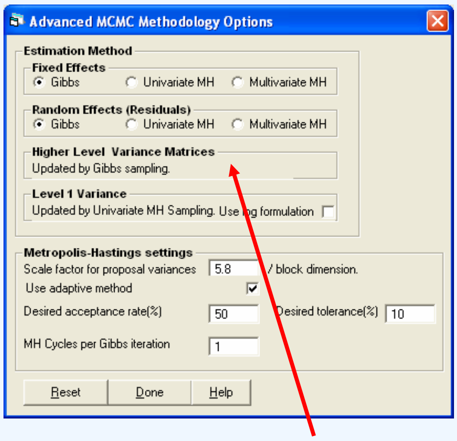


All estimates are blue; ie not converged

clicking on the + button on the tool bar, brings up default weak priors

Switch to MCMC

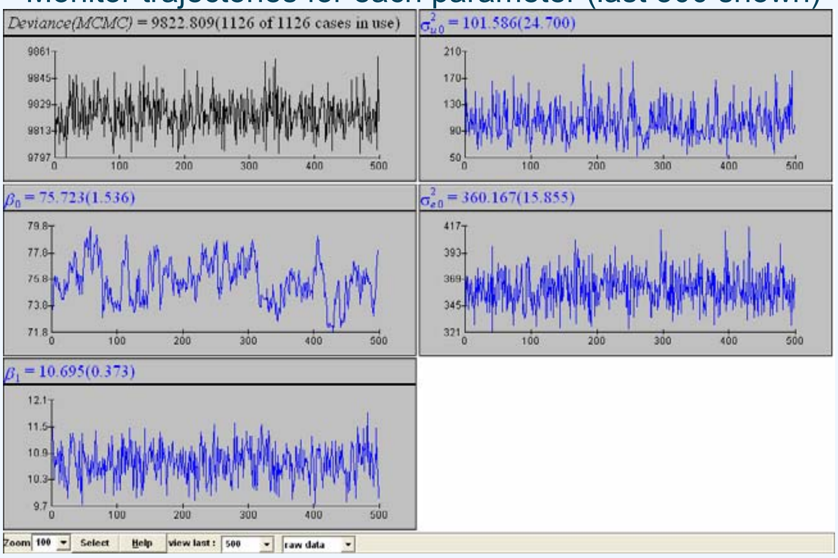




Default: burn-in 500;
monitor 50000

Default sampler Gibbs
(normal theory model)

Monitor trajectories for each parameter (last 500 shown)



ESRC National Centre for Research Methods

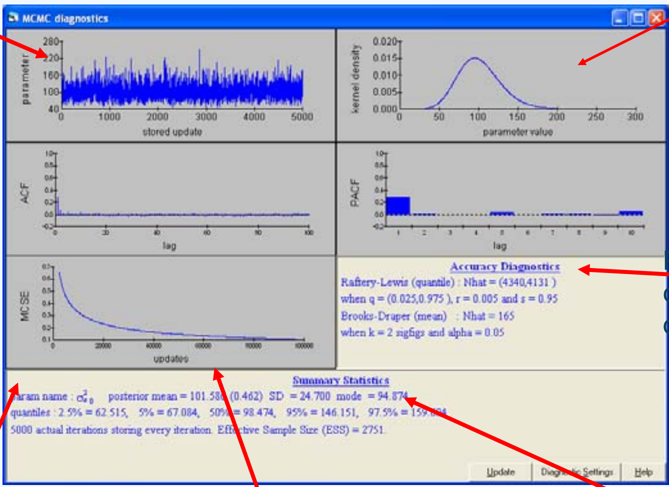
Learning Environment for Multilevel Methodology and Applications

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Trajectories showing good mixing σ_u^2

"White noise"

Posterior Distribution + skew



Prospective convergence diagnostics

95% credible intervals

Posterior mean

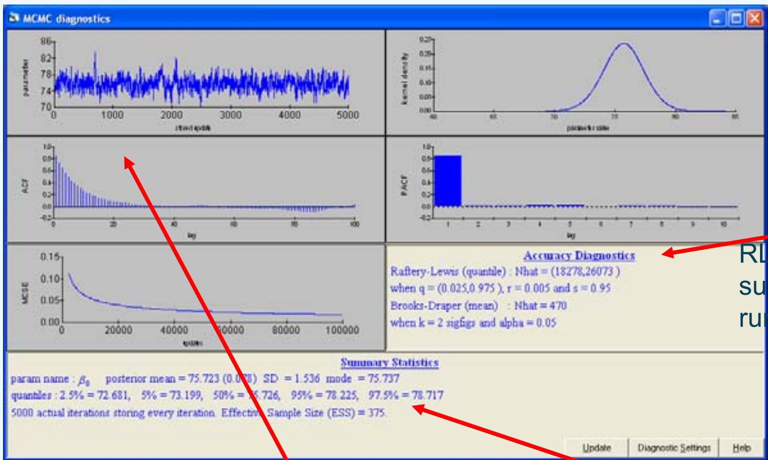
ESS = 2751

ESRC National Centre for Research Methods

Learning Environment for Multilevel Methodology and Applications

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Trajectories showing less- good mixing β_0



RL diagnostic suggests longer run needed

Auto-correlated trace; "stickiness"; exploring the parameter space slowly

ESS = 375

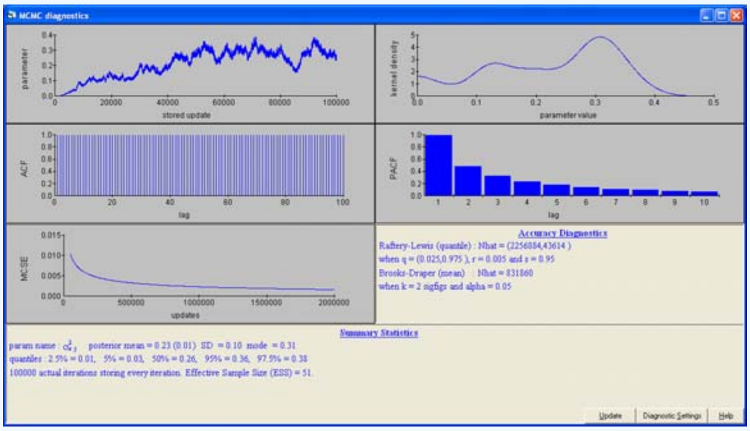
ESRC National Centre for Research Methods

Learning Environment for Multilevel Methodology and Applications

ESRC ECONOMIC & SOCIAL RESEARCH COUNCIL

Very problematic mixing despite 100K!

ESS only 51; highly auto-correlated trace
Still trending at 50k; multimodal distribution
(has not converged to a distribution); due to
very small n; only 11 schools



ESRC National Centre for Research Methods

Learning Environment for Multilevel Methodology and Applications

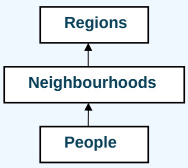
ESRC ECONOMIC & SOCIAL RESEARCH COUNCIL

Models three types of structure

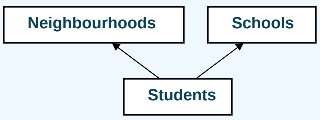
Classification — Hierarchy

Multiple membership — no linkage — Cross-classified

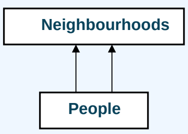
a) 3-level hierarchical structure



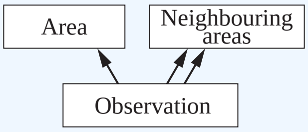
b) cross-classified structure

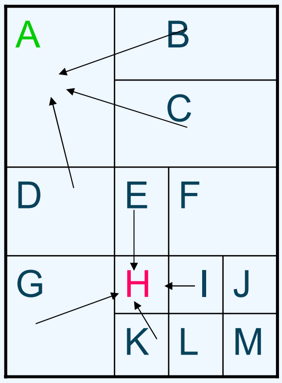


c) multiple membership structure



Spatial Models as a combination of strict hierarchy and multiple membership (including GWR)





Person in A

Affected by A(SH) and B,C,D (MM)

Person in H

Affected by H(SH) and E,I,K,G (MM)

Multiple membership defined by common boundary; weights as function of inverse distance between centroids of areas

Scottish Lip Cancer Spatial multiple-membership model

- Response: observed counts of male lip cancer for the 56 regions of Scotland (1975-1980)
- Predictor: % of workforce working in outdoor occupations (Agric;For; Fish) Expected count based on population size
- Structure areas and their neighbours defined as having a common border (up to 11); equal weights for each neighbouring region that sum to 1

Rate of lip cancer in each region is affected by both the region itself and its nearest neighbours after taking account of outdoor activity

- Model Log of the response related to fixed predictor, with an offset, Poisson distribution for counts;

  NB Two sets of random effects

  1 area random effects; (ie unstructured; non-spatial variation);

  2 multiple membership set of random effects for the neighbours of each region

MCMC estimation: 50,000 draws

$$\text{obs}_i \sim \text{Poisson}(\pi_i)$$

$$\log(\pi_i) = \text{offs}_i + \beta_{0i}\text{cons}_i + 0.048(0.015)\text{perc\_aff}_i$$

$$\beta_{0i} = -0.295(0.217) + \Sigma_{j \in neigh1(i)} w_{i,j}^{(3)} u_{0j}^{(3)} + u_{0,area(i)}^{(2)}$$

$$\left[u_{0,neigh1(i)}^{(3)}\right] \sim \text{N}(0,\ \Omega_u^{(3)}) \ :\ \Omega_u^{(3)} = \left[1.211(0.469)\right]$$

$$\left[u_{0,area(i)}^{(2)}\right] \sim \text{N}(0,\ \Omega_u^{(2)}) \ :\ \Omega_u^{(2)} = \left[0.051(0.051)\right]$$

$$\text{var}(\text{obs}_i|\pi_i) = \pi_i$$

$$Deviance(MCMC) = 270.075(56 \text{ of } 56 \text{ cases in use})$$

Poisson model

Fixed effects: Offset and Well-supported + relation

Well-supported Residual neighbourhood effect

NB: Poisson highly correlated chains

Ohio cancer: repeated measures (space and time!)

- Response: counts of respiratory cancer deaths in Ohio counties
- Aim: Are there hotspot counties with distinctive trends? (small numbers so 'borrow strength' from neighbours)
- Structure: annual repeated measures (1979-1988) for counties
  Classification 3: n'hoods as MM (3-8 n'hoods)
  Classification 2: counties (88)
  Classification 1: occasion (88*10)
- Predictor: Expected deaths; Time
- Model: Log of the response related to fixed predictor, with an offset, Poisson distribution for counts (C1);
  Two sets of random effects
  1. area random effects allowed to vary over time; trend for each county from the Ohio distribution (c2)
  2. multiple membership set of random effects for the neighbours of each region (C3)

MCMC estimation: repeated measures model, 50,000 draws

General trend

$$\log(\pi_i) = \text{logexp}_i + \beta_{0i}\text{cons}_i + \beta_{1i}\text{year}_i$$

$$\beta_{0i} = -0.10983(0.03241) + \sum_{j \in neigh1(i)} w_{i,j}^{(3)} u_{0j}^{(3)} + u_{0,county(i)}^{(2)}$$

$$\beta_{1i} = 0.00345(0.00221) + u_{1,county(i)}^{(2)}$$

$$\left[u_{0,neigh1(i)}^{(3)}\right] \sim \text{N}(0,\ \Omega_u^{(3)}) \ :\ \Omega_u^{(3)} = \left[0.05798(0.03287)\right]$$

N'hood variance

$$\begin{bmatrix} u_{0,county(i)}^{(2)} \\ u_{1,county(i)}^{(2)} \end{bmatrix} \sim \text{N}(0,\ \Omega_u^{(2)}) \ :\ \Omega_u^{(2)} = \begin{bmatrix} 0.02817(0.00727) & \\ -0.00039(0.00042) & 0.00008(0.00004) \end{bmatrix}$$

Variance function for between county time trend

$$\text{var}(\text{obs}_i|\pi_i) = \pi_i$$

Default priors

PRIOR SPECIFICATIONS

$p(\beta_0) \ \alpha\ 1$

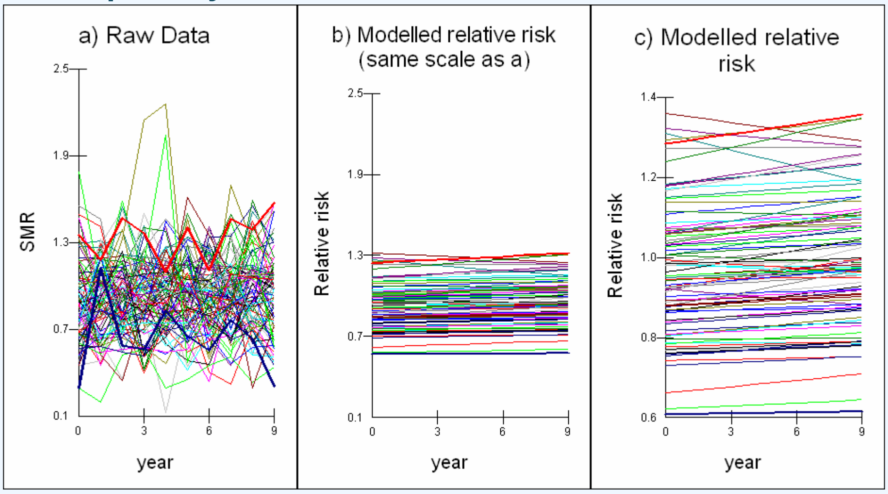
$p(\beta_1) \ \alpha\ 1$

$p(1/\Omega_{u\,0,0}^{(3)}) \sim \text{Gamma}(0.00100, 0.00100)$

$$p(\Omega_u^{(2)}) \sim \text{inverse Wishart}_2[\,2{*}S_u,\ 2],\ S_u = \begin{bmatrix} 0.00000 & \\ 0.00000 & 0.00000 \end{bmatrix}$$

ESRC National Centre for Research Methods
Learning Environment for Multilevel Methodology and Applications
E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Respiratory cancer trends in Ohio: raw and modelled



Red: County 41 in 1988; SMR = 77/49 = 1.57

Blue: County 80 in 1988: SMR= 6/19 = 0.31

Comparing Bugs and MLwiN

- IGLS estimation is far quicker; for Normal response models gives very good estimates
- Model comparison is also easier with formal test statistics
- Relatively easy to set up model and display results in MLwiN

- MCMC in MLwiN is almost always faster than in WinBUGS (examples: typically 15 fold faster)
- But MLwiN has restricted choice of MCMC algorithms, and restricted range of models (but out to Winbugs)

- Having two independent MCMC algorithms for fitting some models is useful as programming mistakes do occur

More information on spatial models in MLwiN

William J. Browne (2003) **MCMC Estimation in *MLwiN***; *Chapter 16 Spatial models*

Lawson, A.B., Browne W.J., and Vidal Rodeiro, C.L. (2003). ***Disease Mapping using WinBUGS and MLwiN*** *Wiley. London (Chapter 8: GWR)*

7 Resources for going further

Resources

Centre for Multilevel Modelling

http://www.cmm.bris.ac.uk

Provides access to general information about multilevel modelling and *MlwiN*.

Email discussion group:

http://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=multilevel

With searchable archives

ESRC National Centre for Research Methods

Learning Environment for Multilevel Methodology and Applications

ESRC Economic & Social Research Council

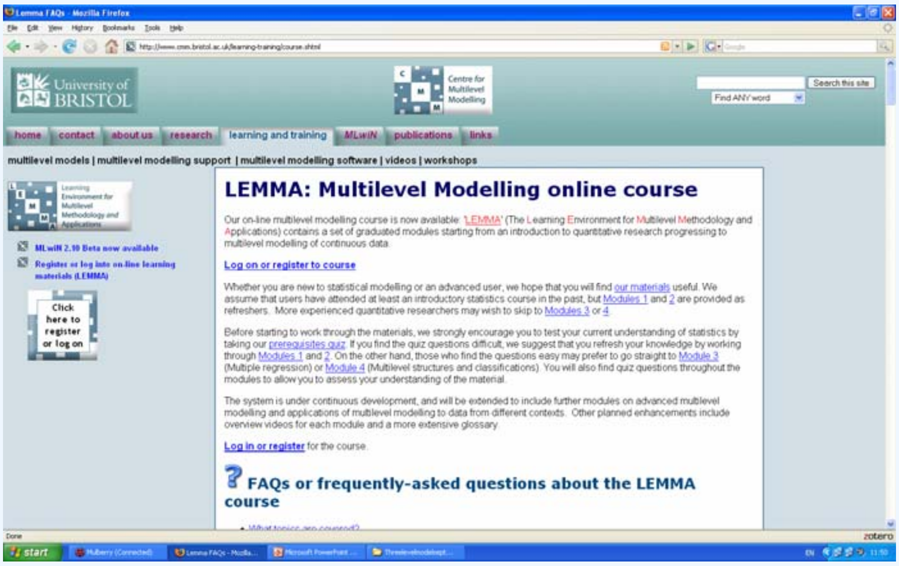
http://www.cmm.bristol.ac.uk/

ESRC National Centre for Research Methods

Learning Environment for Multilevel Methodology and Applications

ESRC Economic & Social Research Council

http://www.cmm.bristol.ac.uk/learning-training/course.shtml

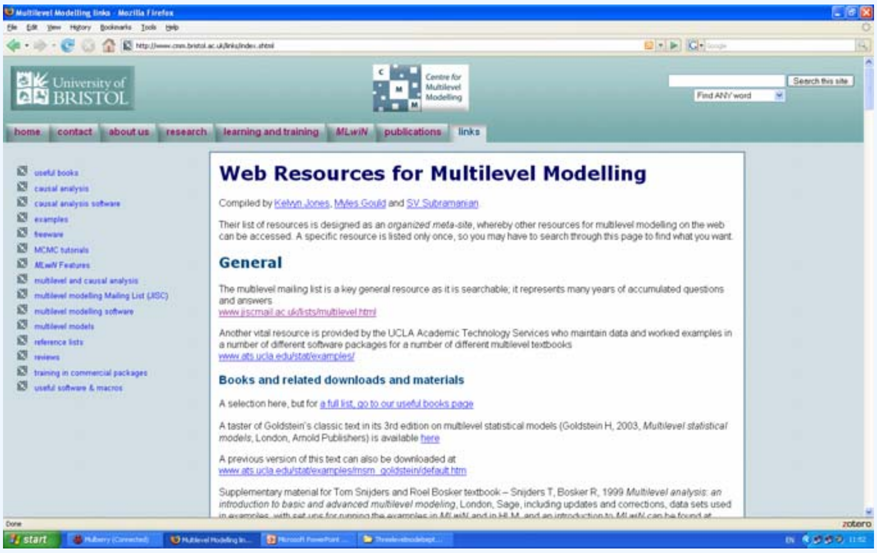


ESRC National Centre for Research Methods

Learning Environment for Multilevel Methodology and Applications

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

http://www.cmm.bristol.ac.uk/links/index.shtml

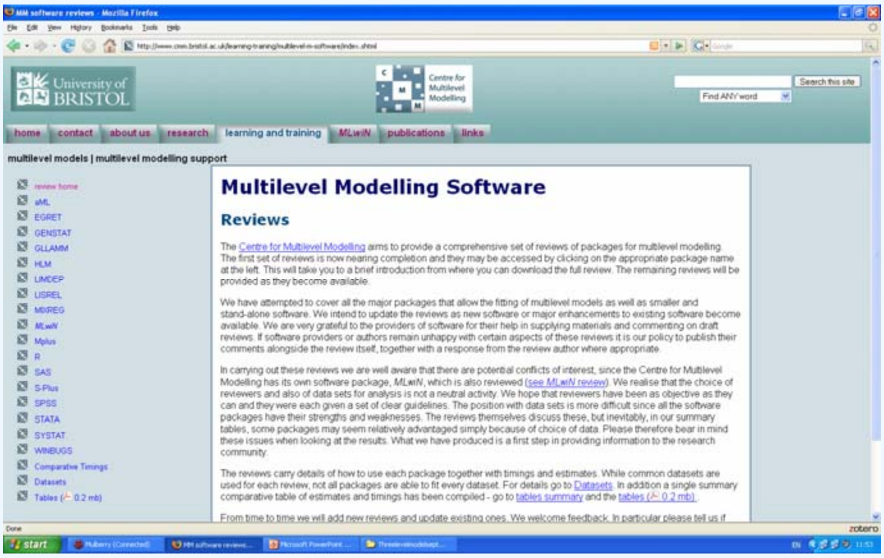


ESRC National Centre for Research Methods

Learning Environment for Multilevel Methodology and Applications

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

http://www.cmm.bristol.ac.uk/learning-training/multilevel-m-software/index.shtml



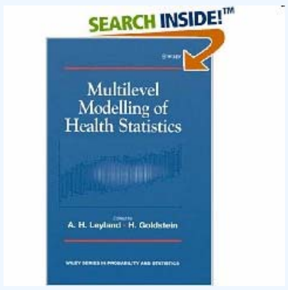
The MLwiN manuals are another training resource

http://www.cmm.bristol.ac.uk/MLwiN/download/manuals.shtml

ESRC National Centre for Research Methods
Learning Environment for Multilevel Methodology and Applications
E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Texts

- Comprehensive but demanding! : Goldstein
- Thorough but a little dated: Snijders & Bosker
- Approachable : Hox
- Authoritative: de Leeuw & Meijer
- Applications: education, O'Connell & McCoach
- Applications: health, Leyland & Goldstein



http://www.cmm.bristol.ac.uk/learning-training/multilevel-m-support/books.shtml