Scottish Graduate School of Social Sciences

SGSSS Summer School 2025
Queen Margaret University, Edinburgh

Multilevel Models for Applied Social Research

Instructor: Paul Lambert
Univ. Stirling

Co-facilitator: Kate O'Hara
Univ. Stirling & Univ. Edinburgh

0930-1500hrs, 18/Jun/2025

Github resources:
https://github.com/paul-lambert/SGSSS-2025/

0930-1100	Talk 1: Classical perspectives on multilevel modelling Talk 2: Realistic complexity
1115-1230	Lab 1: Implementing selected popular multilevel models Talk 3: Case study on effect scores from random effects residuals
1330-1500	Talk 4: MLMs meet econometrics
	Lab 2: Responding to complex data and to complex analytical options

(4a) Do you just mean it's a random effects model?

- Almost!

Econ 101:

i. Use panel data
ii. Try random effects model for the system
iii. Also try fixed effects model for the same system
iv. Use the Hausman test to compare results
v. Choose the fixed effects model

- Yes, overwhelmingly, 'multilevel models' just mean models that use random effects for multilevel clusters
  - (But sometimes, 'multilevel models' refer to any model designed to deal with a multilevel system, not just via random effects)

- Well, in econometrics, most uses of random effects are the two-level random intercepts model and don't really reflect the variety of ways that MLMs are used constructively elsewhere…
  - Econ. focus is usually on **optimising β estimates**, ignoring things like variance decomposition estimates and higher level residuals
  - Econ. typically use estimators such as '**gls**' which don't prioritise accuracy of the random part variance
  - Econ. examples rarely model '**random slopes**', or 3+ level systems
  - In econ. the classical **contrast is with 'fixed effects'** model, but in the multilevel modelling tradition the boundaries are not so crisp (see 4c)

(4a) Do you just mean it's a random effects model?

The MLM tradition outside economics is associated with particular substantive interests

- Commitment to studying the multiple levels of the system (e.g. of the REWB model, see (4c))
- Motivation of accounting for 'realistic complexity'

Non-econ debates on when random effects models are most helpful for multilevel systems

E.g. Luke 2020, p12, random effects most helpful if any of:

- The clusters can be thought of as a random sample from a wider population of clusters
- The units of the levels can be considered replaceable
- The interest is in variability between clusters
- The are a relatively large number of clusters

Most influential in practice as e.g. in the MAIHDA tradition (Evans et al. 2018)

E.g. Paterson & Goldstein 1992, Bryan and Jenkins 2016

- 25*25 research design approximate guide (30*30 for non-linear outcomes)

Widely disregarded…

E.g. Gelman and Hill 2007

"Advice is sometimes given that multilevel models can only be used if the number of groups is higher than some threshold, or if there is some minimum number of observations per group. Such advice is misguided. …" [2007: 275] "A question that commonly arises is when to use fixed effects … and when to use random effects. The statistical literature is full of confusing and contradictory advice. … Our advice […] is always to use multilevel modelling ("random effects")" [2007: 246]

Just for Bayesianists…

(4b) Why not just use robust standard errors?

- Widely used in economics, robust standard errors reformulate the standard error calculation within a model (based upon properties of individual level residuals)
- 'Sandwich estimator', 'Huber-White standard errors'
- The reformulation can take account of cluster structures, but doesn't have to

E.g. The major UK household surveys (e.g. BHPS, UKHLS, LFS) have a natural mulitilevel structure (respondents in households) yet this is rarely modelled with random effects. It is often ignored outside econ traditions, or modelled with robust standard errors in econ.

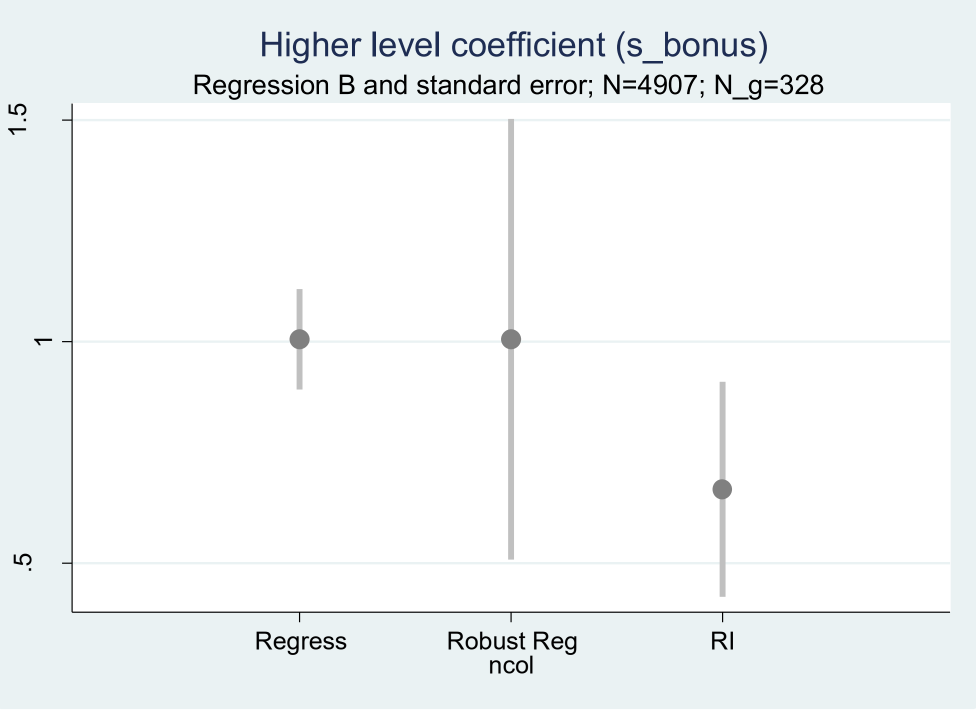
- Well, sometimes but not always!

In applied research, you may see robust standard errors used in response to any of…

- Clustering in the data
- Heteroscedasticity in the model errors
- Non-normality of errors
- Non-random sampling
- Measurement error
- *(e.g. for Stata users, they are implemented automatically if you use 'svy' and have set a 'psu' identifier)*

(4b) Why not just use robust standard errors?

- Many but not all models can incorporate robust standard errors (cf. algorithm/software)
- Usually serve to **widen standard errors**
  - don't impact beta coefficients
  - in some scenarios, might be overly conservative (s.e.'s too big)
  - De facto, equivalent to pretending you have a smaller sample size than you really do (no longer treating as a simple random sample)
  - Standard advice is to add robust s.e.s to random effects if have small average cluster sizes and/or low numbers of clusters
- Particularly in economics, it is common to use robust standard errors *rather than* use random effects parameter(s) to take account of clustering (much quicker to estimate)
  - 'GEE's' or 'Marginal models' are in effect doing this
  - Sometimes use random effects for some structures, robust s.e. for others



(4b) Why not just use robust standard errors?

```
. regress gcse lrt girl sch_2 sch_3

      Source |       SS           df       MS      Number of obs   =     4,059
-------------+----------------------------------   F(4, 4054)      =    580.15
       Model |  147413.623         4  36853.4058   Prob > F        =    0.0000
    Residual |  257526.693     4,054  63.5240979   R-squared       =    0.3640
-------------+----------------------------------   Adj R-squared   =    0.3634
       Total |  404940.316     4,058  99.7881508   Root MSE        =    7.9702

------------------------------------------------------------------------------
        gcse | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         lrt |   .5910456   .0126259    46.81   0.000     .5662919    .6157993
        girl |   1.326415   .3431437     3.87   0.000     .6536643    1.999165
       sch_2 |   1.825545   .4256816     4.29   0.000     .9909749    2.660114
       sch_3 |   1.680108   .3259095     5.16   0.000     1.041146    2.319069
       _cons |  -1.608613   .2395156    -6.72   0.000    -2.078195   -1.139031
------------------------------------------------------------------------------
```

```
. regress gcse lrt girl sch_2 sch_3, robust cluster(school)

Linear regression                               Number of obs     =      4,059
                                                F(4, 64)          =     189.65
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3640
                                                Root MSE          =     7.9702

                                 (Std. err. adjusted for 65 clusters in school)
------------------------------------------------------------------------------
             |               Robust
        gcse | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         lrt |   .5910456   .0228708    25.84   0.000      .545356    .6367352
        girl |   1.326415   .3065043     4.33   0.000     .7141018    1.938727
       sch_2 |   1.825545   .8978494     2.03   0.046     .0318842    3.619205
       sch_3 |   1.680108   .8797399     1.91   0.061    -.0773747     3.43759
       _cons |  -1.608613   .4846955    -3.32   0.001    -2.576904   -.6403227
------------------------------------------------------------------------------
```

- *Robust standard errors make sense if your only concern is to deal reasonably with cluster structure*
- *But blunt instruments, & don't give other insights that might follow from a random effects adjustment*

```
. mixed gcse lrt girl sch_2 sch_3 ||school:lrt, stddev cov(un)

Performing EM optimization ...

Performing gradient-based optimization:
Iteration 0:   log likelihood = -13986.738
Iteration 1:   log likelihood = -13986.738

Computing standard errors ...

Mixed-effects ML regression                     Number of obs     =      4,059
Group variable: school                          Number of groups  =         65
                                                Obs per group:
                                                              min =          2
                                                              avg =       62.4
                                                              max =        198
                                                Wald chi2(4)      =     820.18
Log likelihood = -13986.738                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
        gcse | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         lrt |   .5544234   .0199381    27.81   0.000     .5153455    .5935014
        girl |   1.682633   .3382158     4.98   0.000     1.019743    2.345524
       sch_2 |    1.79864   .9914847     1.81   0.070     -.144634    3.741915
       sch_3 |   1.748186   .7876956     2.22   0.026     .2043314    3.292042
       _cons |  -1.888208   .5135382    -3.68   0.000    -2.894724   -.8816913
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
school: Unstructured         |
                     sd(lrt) |   .1210464   .0189266      .0890962     .164454
                   sd(_cons) |   2.820425   .2901373      2.305428    3.450466
             corr(lrt,_cons) |   .5886686   .1389937      .2531162    .7977935
-----------------------------+------------------------------------------------
                sd(Residual) |   7.417498   .0837029      7.255244     7.58338
------------------------------------------------------------------------------
```

(4c) Why not just use the fixed effects model?

Outcome		'Fixed part'	'Random part'	
Y_{ij}	=	$\beta_p X_{pij} + \beta_q Z_{qj} +$	ε_{ij}	*No cluster model*
Y_{ij}	=	$\beta_p X_{pij} + \beta_q Z_{qj} +$	$\mu_j + \varepsilon_{ij}$	*Random effects*
Y_{ij}	=	$\beta_p X_{pij} + \gamma_j +$	ε_{ij}	*Fixed effects*

- Fair, but if you'd just give random effects a try...!

- The fixed effects model is a popular tool in econometrics.
  - It fully models cluster-to-cluster averages, meaning its remaining β parameters identify on deviations from the cluster average, or 'within effects'
  - In econ, the 'within' parameter is often the default preference
- In default expression, random effects model parameters reflect an unknown mix of 'within' and 'between' processes
  - Though they can be readily disentangled, by fitting cluster means of X vars
- In default expression, fixed effects models preclude higher level explanatory variables (would be collinear to the cluster fixed effects)
  - Though we can work-around by using cluster means + cluster level variables

(4c) Why not just use the fixed effects model?

Special case of panel data

Person (Level 2)	i	**1**			**2**		**3**	**4**				**5**	**6**		**7**	**.**
Time point (Level 1)	t	**1**	**2**	**3**	**1**	**2**	**1**	**1**	**2**	**3**	**4**	**10**	**6**	**7**	**1**	**.**
		4	**5**	**6**	**3**	**4**	**2**	**5**	**6**	**7**	**8**	**11**	**8**	**9**	**2**	
		7	**8**	**9**	**5**	**6**	**3**	**9**	**10**	**11**	**12**	**12**	**10**	**11**	**3**	

- Multilevel models with random effects make sense for panel data, thinking for records from different time points t nested within individuals i, but…
  - Common to compare and contrast with other popular panel models (including fixed effects, which are widely used for panel data analysis in economics)
  - Common to model 'growth' ('growth curves') as an effect of time (linear, quadratic, etc with explicit random slopes for the growth coefficients
  - Common to add an additional specification for structured correlation amongst the level 1 residuals (e.g. autoregressive, meaning that closer time points have higher correlations)

(4c) Why not just use the fixed effects model?

Popular models with panel data include…

(most can be considered for any multilevel data scenario)

'Pooled regression' model	$\mathbf{Y_{it} = \beta X_{it} + \varepsilon_{it}}$ (common to use robust standard errors to adjust for repeated contacts)
Random effects model	$\mathbf{Y_{it} = \beta X_{it} + \mu_i + \varepsilon_{it}}$ (μ_i's are random effects described by a variance)
Fixed effects model	$\mathbf{Y_{it} = \beta X_{it} + \gamma_i + \varepsilon_{it}}$ (γ_i's are fixed effects/dummies)
Between effects model	$\mathbf{\bar{Y}_i = B\dot{X}_i + \bar{e}_i}$
Random effects growth model	$\mathbf{Y_{it} = \beta X_{it} + \zeta\mu_i + \varepsilon_{it}}$ (μ is a random effect, ζ indicates 1+ explanatory variables incl. a time effect)
Two way error components model	$\mathbf{Y_{it} = \beta X_{it} + \mu_i + \zeta_t + \varepsilon_{it}}$ (μ and ζ are random effects)
'Hybrid' or 'REWB' model	$\mathbf{Y_{it} = \beta X_{it} + B\dot{X}_i + \mu_i + \varepsilon_{it}}$ (some or all group means modelled, so model features both within (fe) and between parameters and random effects adjustments)

(4c) Why not just use the fixed effects model?

- *Nowadays, it is pretty easy to implement all options in packages such as Stata, so it is natural to compute them and compare results*

Variable	lin2	clus2	be2	fe2	re2
lninc	-.2985***	-.2985***	-.4575***	-.1574***	-.1826***
fem	1.147***	1.147***	1.038***	0	1.164***
age	.09416***	.09416***	.1125***	.01852	.06013***
age2	-.000918***	-.000918***	-.001081***	.000299*	-.000469***
hied	-.1179**	-.1179	-.07051	.04218	-.02425
noed	.5503***	.5503***	.6927***	-.1867	.5079***
convot	-.6563***	-.6563***	-.5992***	-.1973*	-.502***
labvot	.01591	.01591	.104	-.07397	.04198
_cons	10.56***	10.56***	11.21***	10.64***	10.22***
N	103306	103306	103306	103306	103306
r2	.02755	.02755	.0438	.00203	
r2_p					
ll	-317544	-317544	-61415	-277029	

legend: * p<0.05; ** p<0.01; *** p<0.001

(model: influences on GHQ score in the BHPS)

(4c) Why not just use the fixed effects model?

Econ 101:

i. Use panel data
ii. Try random effects model for the system
iii. Also try fixed effects model for the same system
iv. Use the Hausman test to compare results
v. Choose the fixed effects model

- By common convention in Econ., the 'within parameter' is desired so the FE model is favoured
  - FE is unbiased (as estimator of the within parameter) but may be inefficient, in which case an RE model may be more efficient, but is only better if the Hausman test shows it is still largely unbiased
  - The RE model relies on an assumption of no correlation between random effects and explanatory variables ('NCRX') if it is to represent the within parameter accurately. The Hausman test serves to evaluate this assumption
  - RE models can be slow to estimate or are unstable contingent on estimation tool

- But (simplifying a bit), the within parameter can be retrieved, and CRX risks avoided, with a carefully specified RE model (REWB specification) plus the RE model offers other summarising statistics and outputs – so RE models are often better…
  - (e.g. Bell et al. 2019)

(4c) Why not just use the fixed effects model?

…Retrieving the within-effect in a random effects model…

- Group means + individual scores disentangle within from between
- All possible group means + scores = REWB model
- For any one measure, note the individual score coefficient defaults to (between - within), unless converted to deviation-from-mean

The **within effect** of employment on ghq = **-1.5**

The **between effect** of household profile of employment on ghq = **-1.8** = *(-0.23 – 1.53)(n.s.?)*

*The parameter **-1.67** in the 1st model is some combination of w/b*

Variable	no_gpmean	orig_var	dev_var
ghq			
fem	1.184***	1.193***	1.193***
age	.1906***	.1908***	.1908***
age2	-.00196***	-.001973***	-.001973***
cohab	-.683***	-.6713***	-.6713***
gdn	-.5321***	-.5299***	-.5299***
emp_10hrs	-1.665***	-1.53***	
hh_emp		-.2322	-1.762***
dev_emp			-1.53***
_cons	8.596***	8.664***	8.664***
lns1_1_1			
_cons	.7664***	.7656***	.7656***
lnsig_e			
_cons	1.599***	1.599***	1.599***
Statistics			
N	13790	13790	13790
ll	-42739	-42739	-42739
bic	85563	85572	85572

(4d) What if I want to do something that's not a GLM?

- Fair enough! No need for MLMs all of the time!

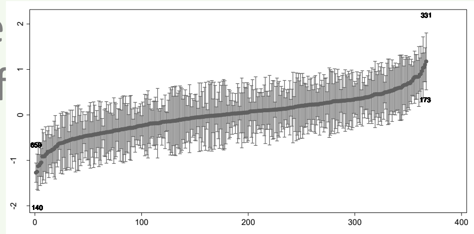
- Effort in econometrics often concerns model specifications designed to tease out a key fixed part parameter in a carefully design system that might not resemble a single GLM
  - The plausible causal effect given an instrumental variable model / regression discontinuity design / propensity score matching controls
  - An isolated empirical relationship net of a selection model process / recursive relationship / multiprocess system
  - Within-unit change in a panel data context

- At present, multilevel models with random effects work neatly for GLM's but there are only limited scenarios where they can be usefully combined with analytical systems that involve more than one GLM
  - Some niche programmes for adding cluster random effects to specialist econ models
  - Structural Equation Models (SEMs) offer flexible options for combining multiprocess systems and random effects, yet are not that widely used to this end

(4d) What if I want to do something that's not a GLM?

X Typical in other Socsci
X Typical in Econ.

- *In a longer course on multilevel models and random effects models I argue:*

Random effects definitely help if we want to tell stories about variation in higher level units (e.g. with random intercepts, random slopes, and higher level residuals)...	E.g. of occupations with above or below average health net of many fixed part predictors 	X
Random effects probably help, but not definitely, to improve efficiency in the analysis of explanatory factors (fixed part influences)	• Down-weighting / robust s.e. tools achieve much the same • De facto view that if ICC < c5%, there's no real impact on Betas and their standard errors	X
Random effects may or may not help to describe the effects of influences at the higher level, contextual effects, or disentangle 'within' and 'between' effects	• All about the fixed part specification; random effects might aid fixed part estimate but sometimes curtails options	X
Random effects don't necessarily help (and can mislead) if your desire is to describe the fixed part effects completely net of the cluster effects	• You want the within effect, but you'll only get this with careful specification (REWB model)	X
Random effects definitely don't help when the model can't be sensibly estimated or described	• E.g. too few units, or too many error effects • Other elements such as multi-process systems that can't readily be combined	X X

Summary: MLMs meet econometrics

(4a) Do you just mean it's a random effects model?	**(4b) Why not just use robust standard errors?**
(4c) Why not just use the fixed effects model?	**(4d) What if I want to do something that's not a GLM?**

References cited

- Bell, A., Fairbrother, M., & Jones, K. (2019). Fixed and random effects models: making an informed choice. *Quality & Quantity*, 53(2), 1051-1074
- Bryan, M. L., & Jenkins, S. P. (2016). Multilevel Modelling of Country Effects: A Cautionary Tale. *European Sociological Review*, 32(1), 3-22 https://doi.org/https://doi.org/10.1093/esr/jcv059
- Evans, C. R., Williams, D. R., Onnela, J. P., & Subramanian, S. V. (2018). A multilevel approach to modeling health inequalities at the intersection of multiple social identities. *Social Science and Medicine*, 203, 64-73,
- Gelman, A., & Hill, J. (2007). *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Luke, D. A. (2020). *Multilevel Modelling, Second Edition*. Sage.
- Paterson, L., & Goldstein, H. (1992). New statistical methods for analysing social structures : An introduction to multilevel models. *British Educational Research Journal*, 17, 387-393.