Modelling heterogeneous variance-covariance components in two-level multilevel models with application to school effects educational research

Research Methods Festival
Oxford
9th July 2014

George Leckie
Centre for Multilevel Modelling
Graduate School of Education
University of Bristol

1. INTRODUCTION

The standard two-level random-slope multilevel model

- The standard **two-level** (e.g. students within schools or repeated measures within subjects) **random-slope multilevel model** can be written as

$$Y_{ij} = \underbrace{\beta_0 + \beta_1 X_{ij}}_{\text{fixed part}} + \underbrace{u_{0j} + u_{1j}X_{ij} + e_{ij}}_{\text{random part}}$$

where

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \mathrm{N}\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & \\ \sigma_{u0u1} & \sigma_{u1}^2 \end{pmatrix}\right\}$$

$$e_{ij} \sim \mathrm{N}(0, \sigma_e^2)$$

- Every school is modelled as having its own regression line with its own intercept, $\beta_0 + u_{0j}$, and its own slope, $\beta_1 + u_{1j}$, but all school are constrained to have a **common residual error variance**, σ_e^2

- However, it will often be substantively interesting to model this residual error variance as **heterogeneous** across students and schools, $\sigma_{e_{ij}}^2$

What we do

- We extend the standard random-slope model by modelling the level-1 variance as a log-linear function of the covariates and further random effects

Mean function: $$Y_{ij} = \underbrace{\beta_0 + \beta_1 X_{ij}}_{\text{fixed part}} + \underbrace{u_{0j} + u_{1j} X_{ij} + e_{ij}}_{\text{random part}}$$

Level-1 variance function: $$\log\left(\sigma^2_{e_{ij}}\right) = \underbrace{\alpha_0 + \alpha_1 X_{ij}}_{\text{fixed part}} + \underbrace{v_{0j} + v_{1j} X_{ij}}_{\text{random part}}$$

where

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ v_{0j} \\ v_{1j} \end{pmatrix} \sim \mathrm{N} \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u0} & & & \\ \sigma_{u01} & \sigma^2_{u1} & & \\ \sigma_{u0v0} & \sigma_{u1v0} & \sigma^2_{v0} & \\ \sigma_{u0v1} & \sigma_{u1v1} & \sigma_{v01} & \sigma^2_{v1} \end{pmatrix} \right\}, \qquad e_{ij} \sim \mathrm{N}(0, \sigma^2_{e_{ij}})$$

- We won't discus modelling the different variances and covariances of the level-2 covariance matrix as a function of the covariates, but this is possible

2. SOFTWARE

Likelihood-based methods

- Not possible to fit these models using routine commands in general-purpose packages such as **R**, **SAS**, **SPSS** and **Stata**, nor is it possible to fit these models in dedicated multilevel modelling packages such as **MLwiN**, **HLM**, and **SuperMix**

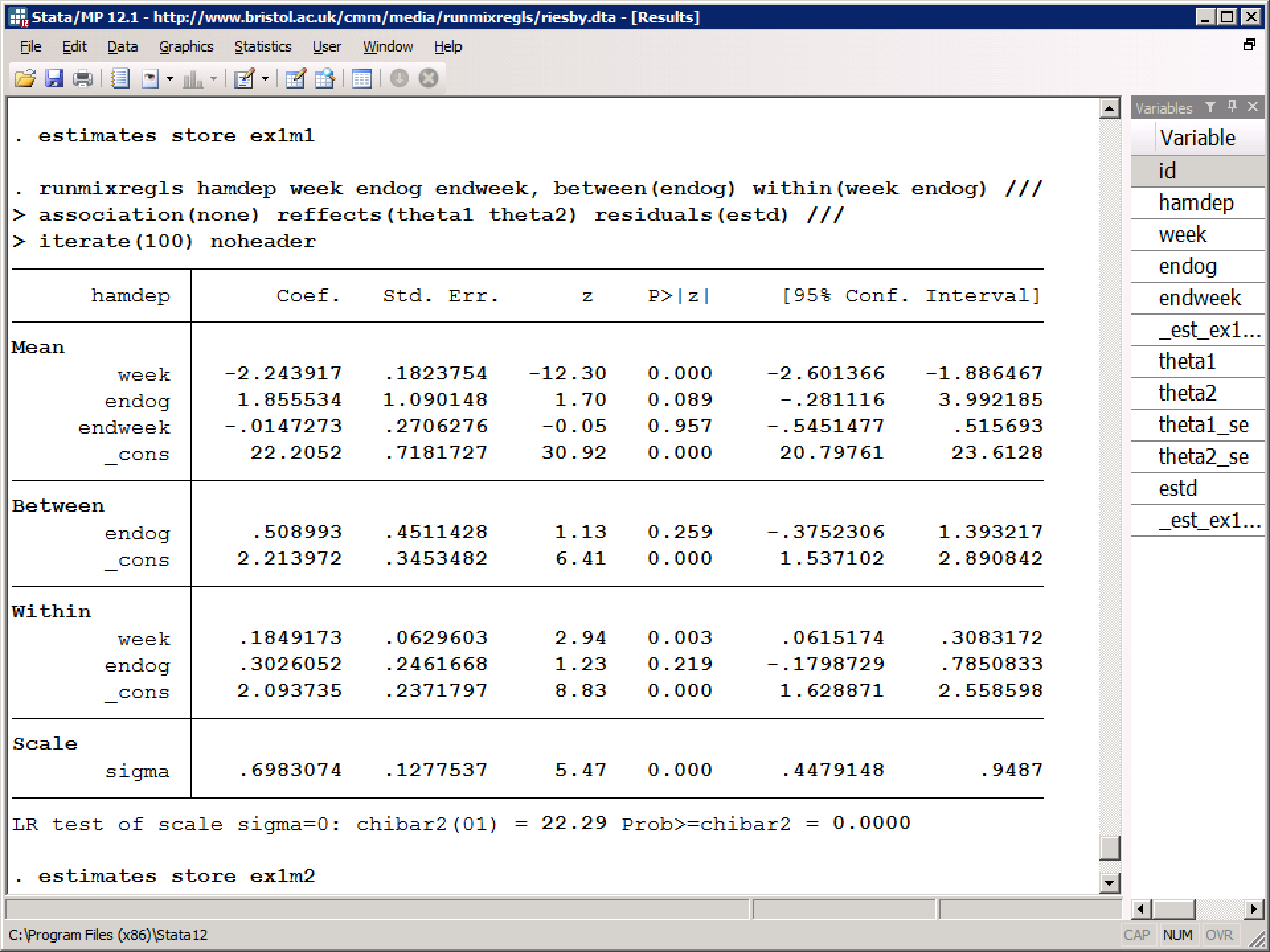
- **ASReml** and **GenStat**: Assume independent random effects

- **SAS PROC NLMIXED**: Two-level models only; slow; sensitive to starting values

- **MIXREGLS**: Developed by Don Hedeker; Two-level random-intercept models only; computationally faster and more stable than SAS; fiddly to use

  – We have written **runmixregls**, a command to call **MIXREGLS** from within **Stata**

  – http://www.bristol.ac.uk/cmm/software/runmixregls/



Stata/MP 12.1 - http://www.bristol.ac.uk/cmm/media/runmixregls/riesby.dta - [Results]

File Edit Data Graphics Statistics User Window Help

```
. estimates store ex1m1

. runmixregls hamdep week endog endweek, between(endog) within(week endog) ///
> association(none) reffects(theta1 theta2) residuals(estd) ///
> iterate(100) noheader
```

hamdep	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Mean						
week	-2.243917	.1823754	-12.30	0.000	-2.601366	-1.886467
endog	1.855534	1.090148	1.70	0.089	-.281116	3.992185
endweek	-.0147273	.2706276	-0.05	0.957	-.5451477	.515693
_cons	22.2052	.7181727	30.92	0.000	20.79761	23.6128
Between						
endog	.508993	.4511428	1.13	0.259	-.3752306	1.393217
_cons	2.213972	.3453482	6.41	0.000	1.537102	2.890842
Within						
week	.1849173	.0629603	2.94	0.003	.0615174	.3083172
endog	.3026052	.2461668	1.23	0.219	-.1798729	.7850833
_cons	2.093735	.2371797	8.83	0.000	1.628871	2.558598
Scale						
sigma	.6983074	.1277537	5.47	0.000	.4479148	.9487

```
LR test of scale sigma=0: chibar2(01) = 22.29 Prob>=chibar2 = 0.0000

. estimates store ex1m2
```

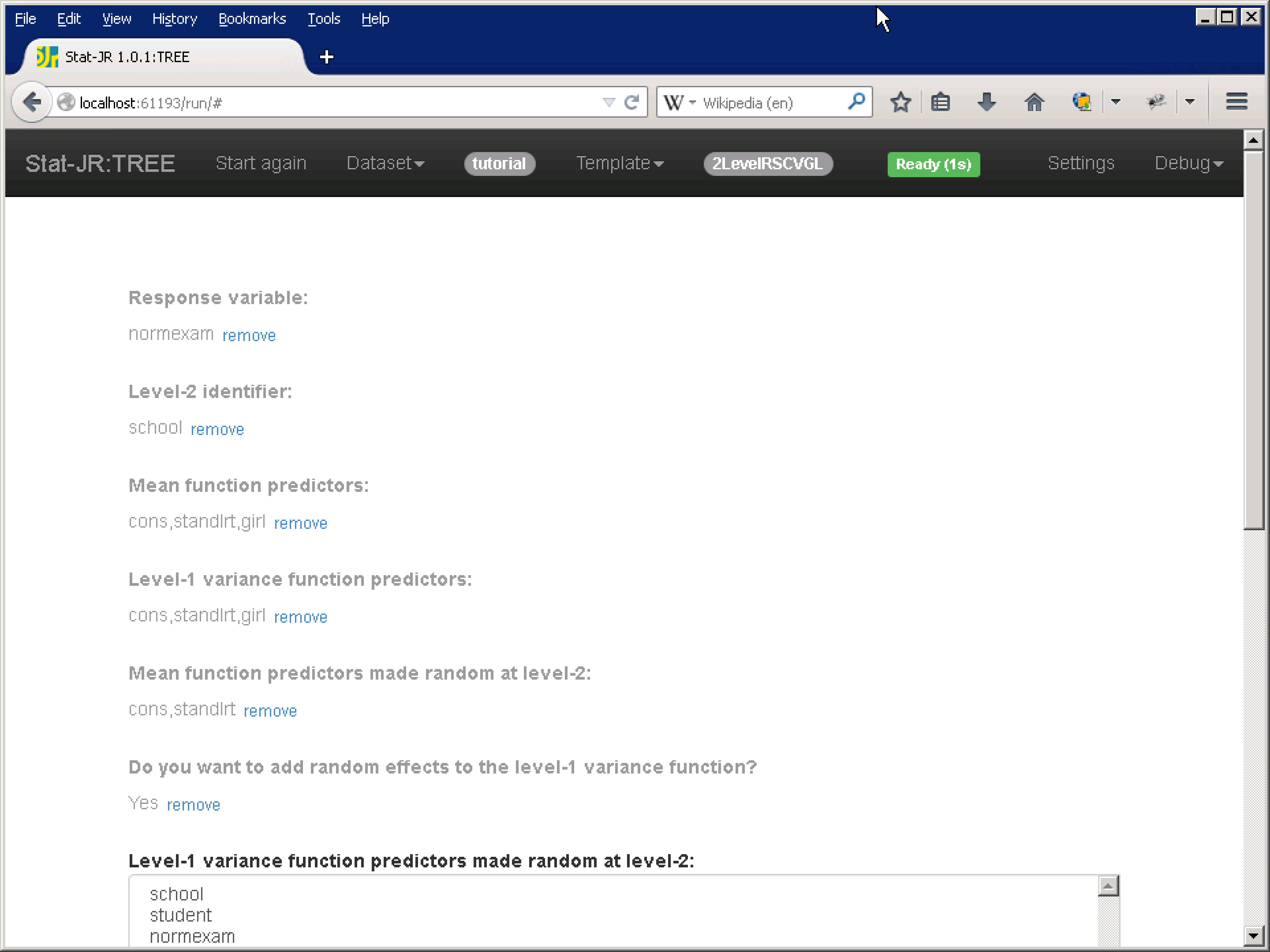
Variables

Variable
id
hamdep
week
endog
endweek
_est_ex1...
theta1
theta2
theta1_se
theta2_se
estd
_est_ex1...

C:\Program Files (x86)\Stata12 CAP NUM OVR

MCMC methods

- **WinBUGS**: Highly flexible; fiddly to use; computationally fairly slow
- **Stat-JR**: Easy to use, computationally faster than WinBUGS, developed by the MLwiN team!
  - We have developed a **2LevelRSCVGL** template to fit this calls of model
  - Need a better name!
  - http://www.bristol.ac.uk/cmm/software/statjr/


File Edit View History Bookmarks Tools Help
Stat-JR 1.0.1:TREE
localhost:61193/run/#
Wikipedia (en)
Stat-JR:TREE
Start again
Dataset
tutorial
Template
2LevelRSCVGL
Ready (1s)
Settings
Debug
Response variable:
normexam remove
Level-2 identifier:
school remove
Mean function predictors:
cons,standlrt,girl remove
Level-1 variance function predictors:
cons,standlrt,girl remove
Mean function predictors made random at level-2:
cons,standlrt remove
Do you want to add random effects to the level-1 variance function?
Yes remove
Level-1 variance function predictors made random at level-2:
school
student
normexam

3. ILLUSTRATIVE APPLICATION

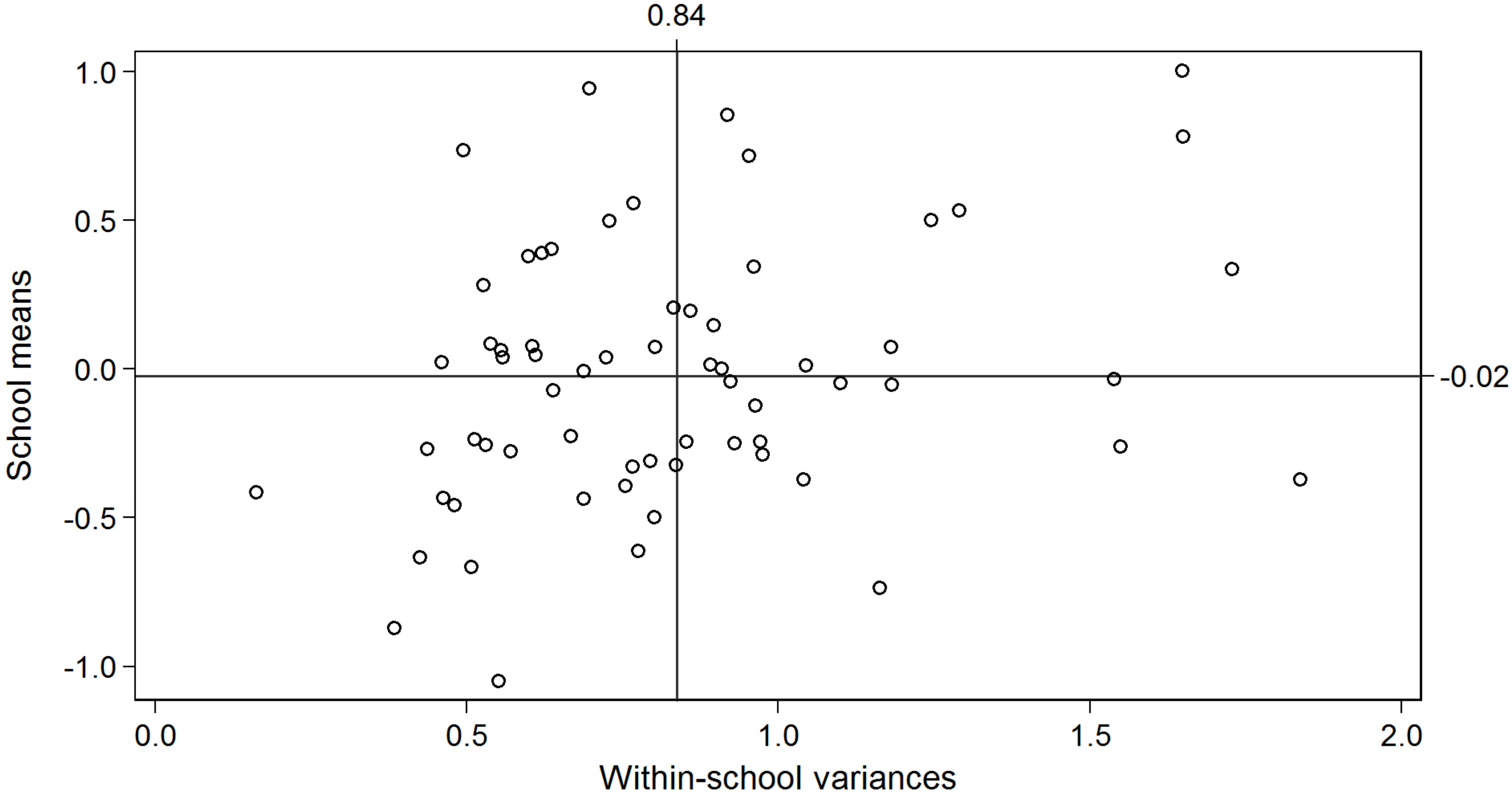
Studies of school effects

- Most studies of school effects focus on estimating **mean differences** in student achievement
  - Which schools score highest, having adjusted for intake differences?
  - What school polices and practices make some schools more effective than others?
- Rarely is anything said about whether there might be **variance differences** in student achievement
- However, just as schools influence the **mean** achievement of their students, they are likely to influence the **dispersion** in their students' achievements
  - Which schools widen initial inequalities and which schools narrow them?
  - What school polices and practices drive these differences?

Inner-London schools' exam scores dataset

- MLwiN 'tutorial' dataset
- 4,059 students (level-1) nested within 65 schools (level-2)
- 2 to 198 students per school (mean = 62 students)
- Response is a standardised age 16 exam score
- Main covariates are
  - A standardised age 11 exam score
  - Student gender

Observed school means and within-school variances



- There is substantial variability in both school means and within-school variances
- There is a moderate positive association between the two ($r = 0.29$)

Specify a log-linear level-1 variance function

- First we specify a log-linear level-1 variance function for the within-school variance and we include a new set of school random effects

Mean function

$$\mathbf{SCORE16}_{ij} = \beta_0 + u_j + e_{ij}$$

$$u_j \sim \mathrm{N}(0, \sigma_u^2)$$

$$e_{ij} \sim \mathrm{N}(0, \sigma_{e_j}^2)$$

Level-1 variance function

$$\log(\sigma_{e_j}^2) = \alpha_0 + v_j$$

$$v_j \sim \mathrm{N}(0, \sigma_v^2)$$

where u_j and v_j are allowed to covary with covariance σ_{uv} (correlation ρ_{uv})

- Every school has its own mean $\beta_{0j} = \beta_0 + u_j$ and variance $\sigma_{e_j}^2 = \exp(\alpha_0 + v_j)$

Random within-school variances

- Model 1 is simply a reparameterised variance-components model where

$$\log(\sigma_e^2) = \alpha_0$$

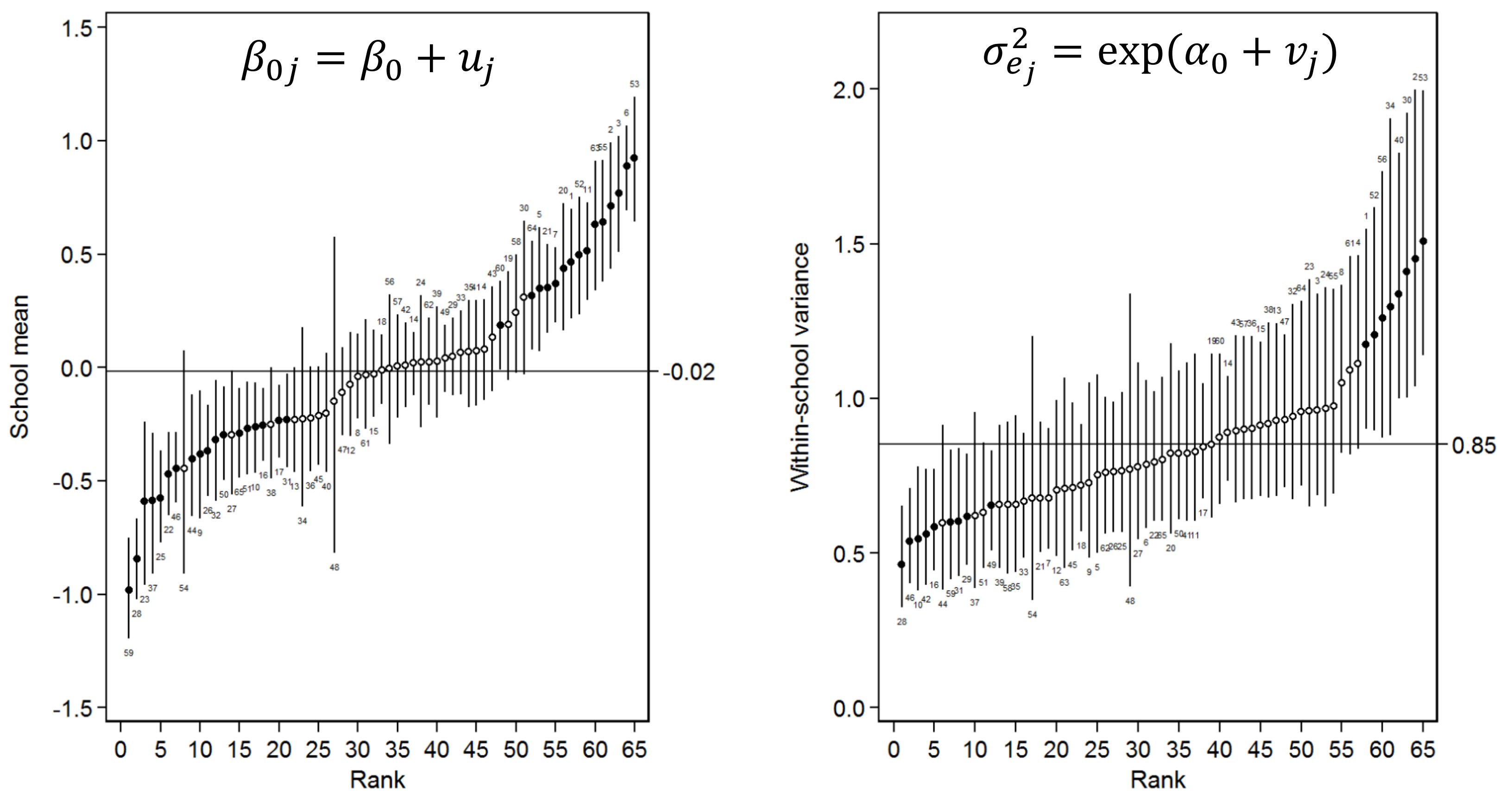
- Model 2 includes the new school random effects

$$\log(\sigma_{e_j}^2) = \alpha_0 + v_j, \qquad v_j \sim \mathrm{N}(0, \sigma_v^2)$$

			Model 1		Model 2	
		Parameter	Mean	SD	Mean	SD
Mean function	β_0	Intercept	-0.02	0.06	-0.02	0.05
	σ_u^2	Intercept variance	0.18	0.04	0.19	0.04
Level-1 variance function	α_0	Intercept	-0.17	0.02	-0.22	0.05
	σ_v^2	Intercept variance	–	–	0.11	0.03
Cross-function	ρ_{uv}	Correlation	–	–	0.36	0.14
		DIC	10910		10783	

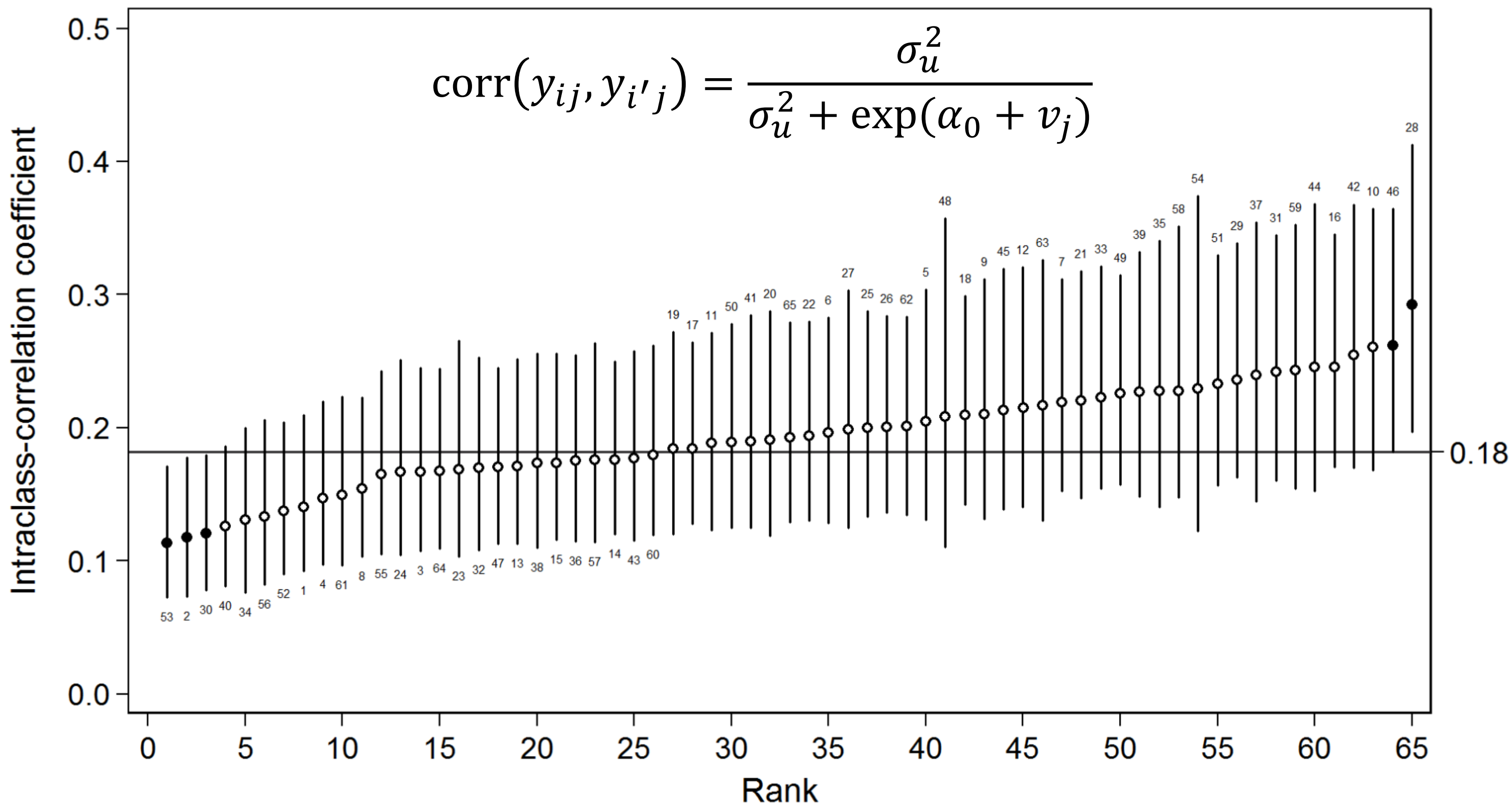
- Model 2 is preferred to Model 1 as shown by drop in DIC of 127 points
- Note that the estimated intercept has decreased from -0.17 to -0.22. Why?

‘Caterpillar’ plots of school means and within-school variances



- While 35 schools differ significantly from the population-average school mean, only 17 schools differ significantly from the population-average within-school variance

Caterpillar plot of intraclass correlation coefficients



- The expected correlation between two students from the same school ranges from 0.11 to 0.29
- Few schools differ significantly from the population-average correlation of 0.18

Add covariates to the mean function

	Parameter		Model 2 Mean	Model 2 SD	Model 3 Mean	Model 3 SD
Mean function	β_0	Intercept	-0.02	0.05	-0.09	0.05
	β_1	Age 11 scores	–	–	0.55	0.01
	β_2	Girl	–	–	0.17	0.03
	σ_u^2	Intercept variance	0.19	0.04	0.10	0.02
Level-1 variance function	α_0	Intercept	-0.22	0.05	-0.60	0.04
	σ_v^2	Intercept variance	0.11	0.03	0.06	0.02
Cross-function	ρ_{uv}	Correlation	0.36	0.14	0.03	0.01
		DIC	10783		9194	

- $\beta_1 = 0.55$ and so age 11 scores are strongly predictive of age 16 scores
- $\beta_2 = 0.17$ and so girls make more *progress* than similar initial achieving boys
- Between-school variance σ_u^2 reduces by 47%
- Population-average of the within-school variances $\mathrm{E}(\sigma_{e_j}^2)$ reduces by 33%
- Population-variance of the within-school variances $\mathrm{Var}(\sigma_{e_j}^2)$ reduces by 78%

Add a random slope to the mean function

- Are schools **differentially effective** for different types of students?
  - Are the schools that are best for high initial achievers different from the schools that are best for low initial achievers?
  - Does the gender gap vary across schools? Are there some schools where boys actually outperform girls?

- Model 4 allows the age 11 slope coefficient to vary across schools

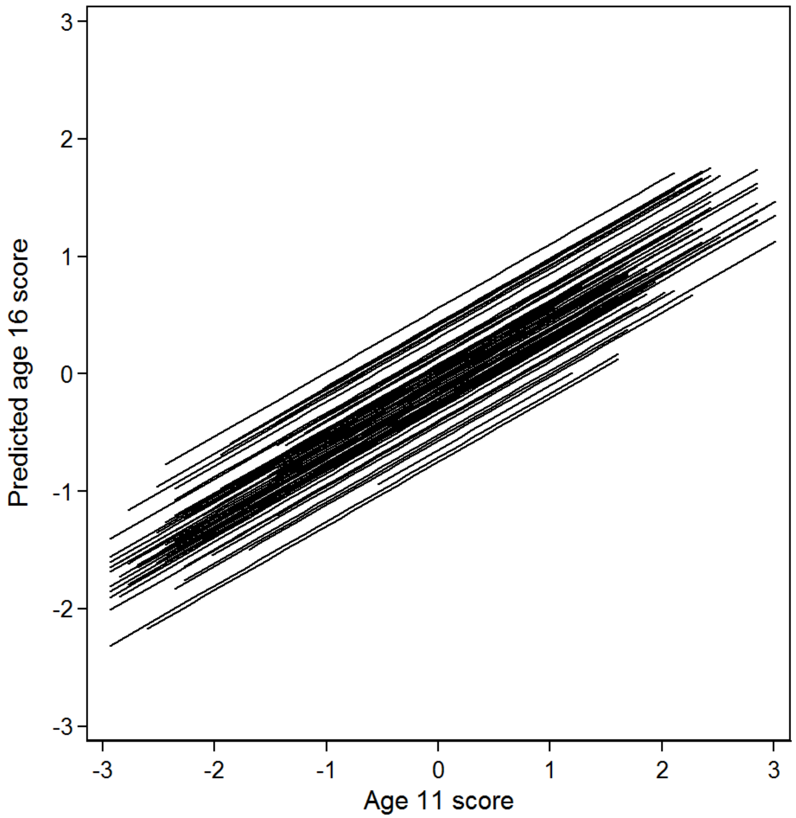
$$\mathbf{SCORE16}_{ij} = \beta_0 + \beta_1 \mathbf{SCORE11}_{ij} + \beta_2 \mathbf{GIRL}_{ij} + u_{0j} + u_{1j}\mathbf{SCORE11}_{ij} + e_{ij}$$

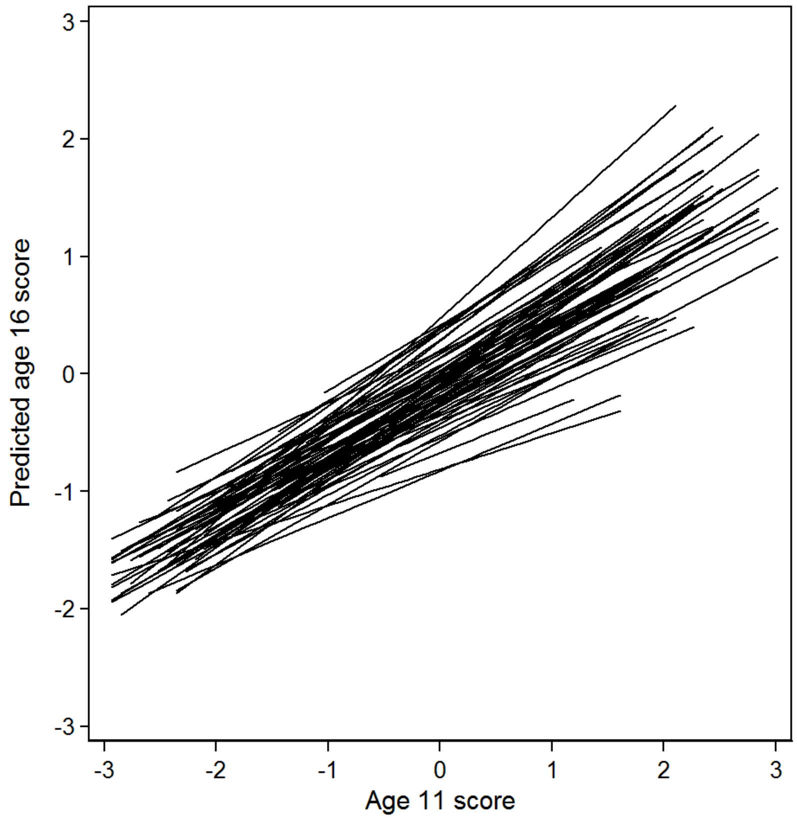
$$\log(\sigma^2_{e_j}) = \alpha_0 + v_j$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ v_j \end{pmatrix} \sim \mathrm{N}\left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u0} & & \\ \sigma_{u01} & \sigma^2_{u1} & \\ \sigma_{u0v} & \sigma_{u1v} & \sigma^2_v \end{pmatrix} \right\}$$

$$e_{ij} \sim \mathrm{N}(0, \sigma^2_{e_j})$$

Predicted mean function school lines





- Age 11 scores are more predictive of age 16 scores in some schools than in others
- Schools with steeper slopes widen initial achievement differences
- School choice matters more for high initial achievers?

Add covariates to the level-1 variance function

		Parameter	Model 4		Model 5	
			Mean	SD	Mean	SD
Mean function	…	…	…	…	…	…
Level-1 variance function	α_0	Intercept	-0.63	0.04	-0.57	0.05
	α_1	Age 11 scores	–	–	-0.07	0.02
	α_2	Girl	–	–	-0.10	0.05
	σ_v^2	Intercept variance	0.06	0.02	0.06	0.02
Cross-function	ρ_{u0v}	Intercept-intercept correlation	0.40	0.17	0.45	0.15
	ρ_{u1v}	Slope-intercept correlation	0.76	0.14	0.77	0.13
		DIC	9133		9121	

- The mean function parameters hardly change and are omitted from the table
- $\alpha_1 = -0.07$ and so, within schools, low initial achievers tend to score more variably than high initial achievers
- $\alpha_2 = -0.10$ and so, within schools, girls tend to score less variably than boys

Add a random slope to the level-1 variance function

- Do schools have **differentially dispersed** outcomes for different types of students?
  - Are the schools that are least dispersed for high initial achievers different from the schools that are least dispersed for low initial achievers?
  - Does the gender dispersion gap vary across schools?
- Model 6 adds a random slope to the level-1 variance function

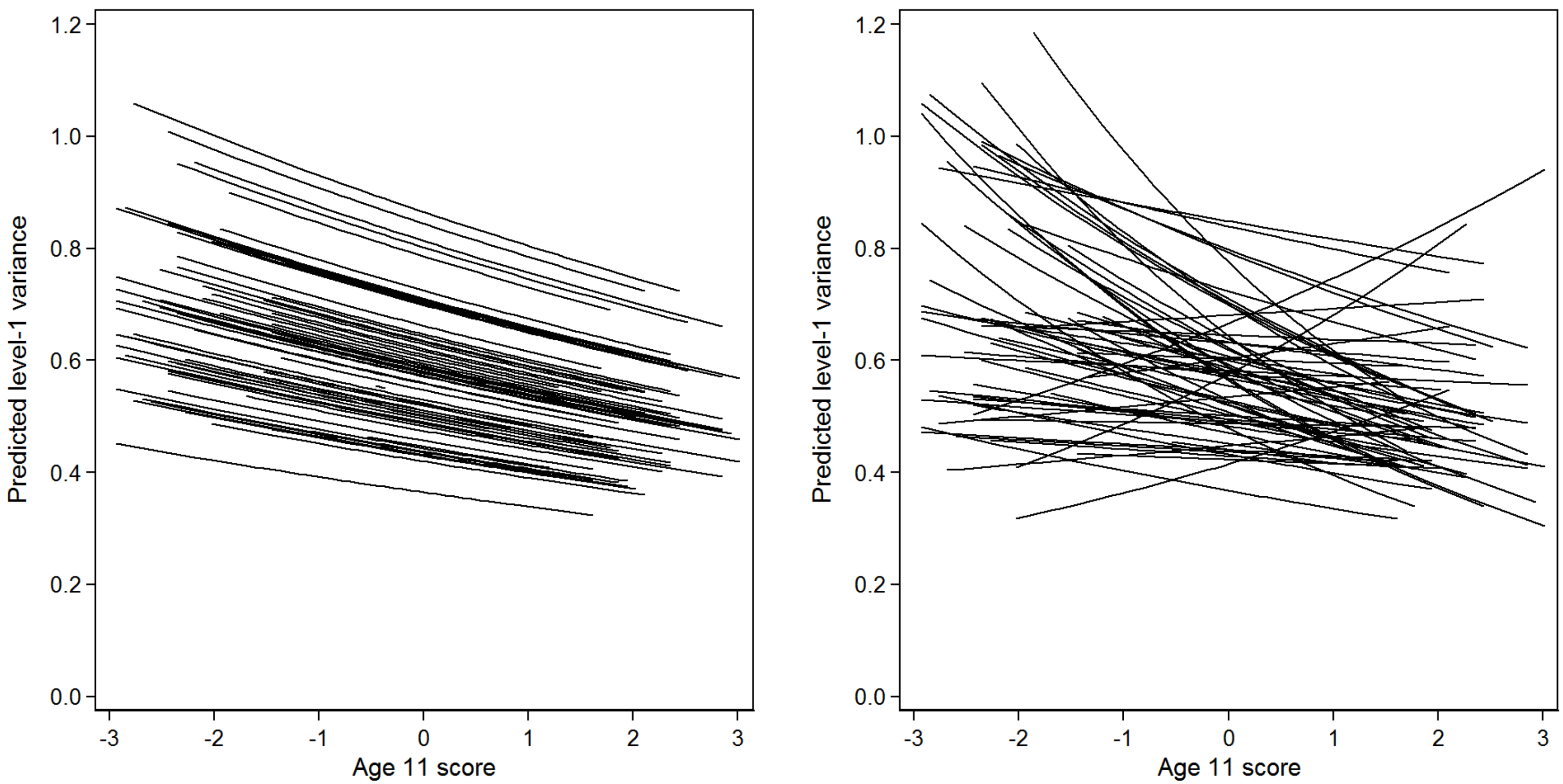
$$\mathbf{SCORE16}_{ij} = \beta_0 + \beta_1\mathbf{SCORE11}_{ij} + \beta_2\mathbf{GIRL}_{ij} + u_{0j} + u_{1j}\mathbf{SCORE11}_{ij} + e_{ij}$$

$$\log(\sigma^2_{e_{ij}}) = \alpha_0 + \alpha_1\mathbf{SCORE11}_{ij} + \alpha_2\mathbf{GIRL}_{ij} + v_{0j} + v_{1j}\mathbf{SCORE11}_{ij}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ v_{0j} \\ v_{1j} \end{pmatrix} \sim \mathrm{N}\left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u0} & & & \\ \sigma_{u01} & \sigma^2_{u1} & & \\ \sigma_{u0v0} & \sigma_{u1v0} & \sigma^2_{v0} & \\ \sigma_{u0v1} & \sigma_{u1v1} & \sigma_{v01} & \sigma^2_{v1} \end{pmatrix} \right\}$$

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

Predicted level-1 variance function school 'lines'



- Three schools actually go against the overall trend and should be examined further
- What is it about these three schools which leads their highest initial achieving students to perform more erratically than their lowest initial achieving students?

Explaining the differences between schools

- So far we have *quantified* differences in *effectiveness* and *dispersion* between schools and how the magnitude of these differences vary as function of initial achievement

- The obvious next step is to seek to *explain* these differences in terms of school-level predictors W_j

  - Entering W_j as a **main effect** into the mean function will explain away σ_{u0}^2

  - Entering W_j as a **cross-level interaction** with X_{ij} into the mean function will explain away σ_{u1}^2

  - Entering W_j as a **main effect** into the level-1 variance function will explain away σ_{v0}^2

  - Entering W_j as a **cross-level interaction** with X_{ij} into the level-1 variance function will explain away σ_{v1}^2

4. SIMULATION STUDY

Can we ignore the random effects?

- Many packages allow you to fit limited level-1 variance functions with no random effects
  - R, SAS, SPSS, Stata
  - HLM, MLwiN
- However, we have carried out simulations which show that ignoring level-2 variability in the level-1 variances leads the level-1 variance function regression coefficients to be estimated with spurious precision
  - This problem is particularly acute for the coefficients of level-2 covariates
  - We run the risk of making **Type I errors of inference** about predictors of level-1 variance
  - This problem is analogous to ignoring clustering in linear regression

5. CONCLUSION

Conclusion

- We have extended the standard two-level random-slope model to model the residual error variance as a function of the covariates and additional random effects

- We are implementing this in **runmixregls** and the new **Stat-JR** software

  - http://www.bristol.ac.uk/cmm/software/runmixregls/
  - http://www.bristol.ac.uk/cmm/software/statjr

- The principle of modelling within-group variances as randomly varying across groups applies to multilevel models more generally, including those with **additional levels**, **crossed random effects** and **discrete responses**

- The discussed methods are relevant to any study where there is interest on estimating dispersion differences on outcome variables across groups

References to our work

- Goldstein, H., Leckie, G., Charlton, C., and Browne, W. Multilevel models with random effects for level 1 variance functions, with application to child growth data. Submitted.

- Leckie, G. (2013). Modeling the residual error variance in Two-Level Random-Coefficient Multilevel Models. *Bulletin of the International Statistical Institute*, 68, 1-6.

- Leckie, G. (2014). runmixregls - A Program to Run the MIXREGLS Mixed-effects Location Scale Software from within Stata. *Journal of Statistical Software, Code Snippet*, 1-41. Forthcoming.

- Leckie, G., French, R., Charlton, C., and Browne, W. (2015). Modeling Heterogeneous Variance-Covariance Components in Two-Level Models. *Journal of Educational and Behavioral Statistics*. Forthcoming.

References to other work

- Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An Application of a Mixed-Effects Location Scale Model for Analysis of Ecological Momentary Assessment (EMA) Data. *Biometrics*, *64*, 627-634.

- Lee, Y., & Nelder, J. A. (2006). Double hierarchical generalized linear models (with discussion). *Applied Statistics*, *55*, 139–185.

- Rast, P., Hofer, S. M., & Sparks, C. (2012). Modeling individual differences in within-person variation of negative and positive affect in a mixed effects location scale model using BUGS/JAGS. *Multivariate Behavioral Research*, *47*, 177-200.

What about modelling the level-2 variance-covariance matrix?

- It is relatively easy to model a 2×2 variance-covariance matrix as a function of the covariates

$$\begin{pmatrix} u_j \\ v_j \end{pmatrix} \sim \mathrm{N}\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u_j} & \\ \sigma_{uv_j} & \sigma^2_{v_j} \end{pmatrix}\right\}$$

$$\log\left(\sigma^2_{u_j}\right) = \kappa_0 + \kappa_1 W_j$$

$$\log\left(\sigma^2_{v_j}\right) = \gamma_0 + \gamma_1 W_j$$

$$\tanh^{-1}\left(\rho_{uvj}\right) = \delta_0 + \delta_1 W_j$$

- However, simply specifying appropriate link functions will no longer ensure positive definiteness in 3×3 and larger variance-covariance matrices
  - In MCMC sampler, reject any proposed parameter values which give rise to variance-covariance matrices which are not positive definite