Issue in sampling the PLASC/NPD database for fitting of complex models:

Experiments with a cross-classified multilevel model

Antony Fielding
(University of Birmingham/ University of Bristol)

Background

- Scoping of educational resource multilevel models with crossing of area and education levels

Fielding, A. (with assistance from H.Thomas, I. Davison, F. Steele, A. Leyland, W. Browne, N. Spencer). (2006). *Using Cross-Classified Multilevel Models to Improve Estimates of the Determination of Pupil Attainment: A Scoping Study*, *Research Report for Department for Education and Skills* , School of Education, University of Birmingham, IBSN: 9780704426016 (International) 0704426013 (UK).

Levačić, R., Jenkins, A., Vignoles, A., Steele. F. and Allen, R. (2005). *Estimating the relationship between school resources and pupil attainment at Key Stage 3*. *Research Report 679*. London: Department for Education and Skills

- Frequent questions on multilevel JISCmail list about how to handle very large data sets
- Caveats surrounding purposive selection of particular subsets of the data: e.g. particular regions

Leckie (2009). The complexity of school and neighbourhood effects and movements of pupils on school differences in models of educational achievement. *Journal of the Royal Statistical Society, Series A, 3, 2009, 537-554*

- Designing multilevel studies

  MLPOWSIM: http://www.cmm.bristol.ac.uk/learning-training/multilevel-models/samples.shtml

  PINT: Bosker, R.J., Snijders, T.A.B. and Guldemond, H. (1996) PINT (Power IN Two-level designs) User Manual

Why sample?

Stress fairly complex models is what we are talking about
Summary statistics or simpler model or specific tightly formulated inferences OK with full data:

- Software +computer limitations if data or models are too large or complex
- Cleaning and preparation of data more difficult with large data sets
- Difficult to do all but simple models
- Time involved in fitting complex models to large data sets – back to old days
- Development and exploration- go back to full model for precise estimation after model selection and diagnostics on quick implementations on sample data
- Maybe two parts--- sample to estimate fixed and variance parameters – go back for estimating residuals
  Since in MCMC this is conditional my estimation framework should be OK

WE HOPE TO CONVINCE THAT SAMPLING OF RIGHT SORT IS OK

Nature of the experiment

Experimentation balancing precision and time – depends on context which is more important.

We look at sampling schemes and features and see what conclusions can be drawn

Look at a particular cross-classified model --- simple enough to handle but complex enough to illustrate major. Hierarchical models can probably be handled OK even if data set is extremely large

More complex models imply more random effects with imbalance and sparsity of data

Main factors governing difficulty of handling and time taken to run models are numbers of random effect in each crossed level but also the number of non-empty cell combinations

The model we use is for methodological exploration only. We leave its substantive import aside.

The framework

DATA : Students who took Key stage 4 in 2007: 543053 clean cases

MODEL: Response variable is a normalised KS4 score (see Leckie 2009). Students at level 1 are clustered in a crossing of the secondary school at time of KS4 and primary school for KS2

$$\text{ks4\_score}_i \sim \text{N}(XB,\ \Omega)$$

$$\begin{aligned}\text{ks4\_score}_i = {} & \beta_{0i}\text{cons}_i + \beta_1\text{ks2\_score}_i + \\ & \beta_2\text{Other White}_i + \\ & \beta_3\text{Black Caribbean}_i + \\ & \beta_4\text{Black African}_i + \beta_5\text{Indian}_i + \\ & \beta_6\text{Pakistani}_i + \beta_7\text{Chinese}_i + \\ & \beta_8\text{Other}_i + \beta_9\text{female}_i + \beta_{10}\text{age}_i + \\ & \beta_{11}\text{fsm}_i + \beta_{12}\text{sen}_i\end{aligned}$$

$$\beta_{0i} = \beta_0 + u^{(3)}_{0,se\_ks4\_school(i)} + u^{(2)}_{0,se\_ks2\_school(i)} + e_{0i}$$

$$\begin{bmatrix} u^{(3)}_{0,se\_ks4\_school(i)} \end{bmatrix} \sim \text{N}(0,\ \Omega_u^{(3)}) \ :\ \Omega_u^{(3)} = \begin{bmatrix} \Omega^{(3)}_{u0,0} \end{bmatrix}$$

$$\begin{bmatrix} u^{(2)}_{0,se\_ks2\_school(i)} \end{bmatrix} \sim \text{N}(0,\ \Omega_u^{(2)}) \ :\ \Omega_u^{(2)} = \begin{bmatrix} \Omega^{(2)}_{u0,0} \end{bmatrix}$$

$$\begin{bmatrix} e_{0i} \end{bmatrix} \sim \text{N}(0,\ \Omega_e) \ :\ \Omega_e = \begin{bmatrix} \Omega_{e0,0} \end{bmatrix}$$

TIME : What is a long time is a relative idea. Our model is not too complex

ESTIMATION: MCMC, Burn in 500, Chain length 10000, Thinning 20 Refresh rate 20

Possibilities of IGLS and RG method

MACHINE: Compaq Presario C700 Notebook PC
Windows Vista
Intel ® Core ™ Duo CPU T5450 @1.66Ghz 1.67 GhZ
2 GB Ram

Model fitted to full data set		
	Est	**SE**
Intercept	-0.0158	0.0045
KS2 score	0.6288	0.0011
Ethnicity (ref White)		
Other White	0.1444	0.0063
Black Caribbean	0.0705	0.0081
Black African	0.3385	0.0085
Indian	0.3938	0.0069
Pakistani	0.3296	0.0072
Chinese	0.5487	0.0162
Other	0.1692	0.0044
Female	0.1711	0.0018
Age	-0.0095	0.0003
FSM	-0.1852	0.0029
SEN	-0.3296	0.0026
Random effects		
Secondary School	0.0506	0.0014
Primary School	0.0281	0.0005
Pupil	0.3891	0.0008
Number of Pupils	543053	
Number of Secondary schools	3097	
Number of Primary Schools	14762	
Combinations	80472	
Time	1hr 50 mins	

Samples of students: Features

- Fixed effects still reasonably precise for smaller samples , e.g. 10% sample taking 15 minutes has standard errors of order 3 times full data set
- Main problem with this is that number of KS4 and KS2 schools is hardly reduced with this order of sample size. Number of combinations remains high
- To reduce time considerably means very imprecise fixed effects , although the variance components remain reasonably well estimated
- Even a run with only 2823 students takes 2 minutes since there still remain 1797 ks4 school effects and 2496 ks2 with 2626 combinations

Lesson for efficient sampling balancing time and precision means we might consider ways of sampling higher level units

Sampling regions and take all KS4 students and schools of that region

Might expect above problems to be ameliorated due to contiguity of students in samples and for given number of students fewer schools involved.

But 'out of area' problem still prevalent to some extent: Number of primary schools and hence random effects out of region range from 30% to 50%

They do run quite quickly

Main problem is that results differ quite considerably across regions—hence scope for inferential generalisability is limited

Sampling Local Education Authorities and take all KS4 students and KS4 schools of that LEA

- Might expect design effect on standard errors if LEA clustering is important—does not seem much in evidence e.g. fixed effect standard errors for 12 LAS are similar to student sample of 10% (similar in size in terms of students)—on these grounds LEA sampling seems OK
- SE of secondary school variance is higher but to be expected
- Main difficulty is still lack of real control over number of primary schools and combinations. Design leads to many 'out of area' primary schools and combinations and time is not much improved over student samples

Direct sampling of KS4 schools

- Design effect of clustering is not evident in fixed effects SE so control by sampling KS4 effects is OK on these grounds
- Reducing number of schools from 1002 to 144 increases standard errors of fixed effects by a factor of 2.5 but are still fairly precise
- Since number of KS4 schools reduced the SE of the KS4 school variance estimate is increased for samples of comparable student size (and to a lesser extent KS2 variance)
- Is this worth it in terms of reduction of time to execute consequent on reduction of KS4 effects implicit? It seems not. 144 KS4 schools with 55216 students take 13 minutes versus student sample of 54544 taking 16 minutes
- Role of KS2 schools –another factor is that there is less control over their number and this is probably major influence on time

5% of KS4 schools → 25% of KS2 schools
10% of KS4 schools → 45% of KS2 schools
33% of KS4 schools→ 84% of KS2 schools#

If we want fairly precise fixed effects and KS4 variance estimates, sampling in this way then time involved may be quite high since almost all primary schools have to be involved and there are large numbers of sparse combinations

Direct sampling of KS2 schools

Does experimentation with this lead us in any new directions?

- Direct contrast of results after sampling schools versus simple sampling students of similar size ---- estimated SEs are very similar ---- thus no evidence of cluster design effect
- Time comparison with direct sampling of students

Samples of KS2 schools student number	271218 (7381 schools	108805 (2953 schools)	55314 (1482 schools	24023 (736 schools)	5720 (147 schools)
Time for run (mins)	45	28	13	10	2
Direct student sample number	271316	136106	54544	27228	5532
Time for run (mins)	75	32	17	14	4

- Compare 1482 KS2 schools with 300 KS4 schools
  Almost same number of students
  Time similar
  Fixed effects precision directly comparable
  Secondary school variance more precise due to many more in sample? (2911 versus 300)
  Primary school variance similar (even though 1482 versus 6603)
- Secondary school representation + combinations

Samples of KS2 schools	7381 (50%)	2953 (20%)	1482 (10%)	736 (5%)	147 (1%)
KS4 schools in sample	3095 (99.9%)	3075 (99.3%)	2911 (94.0%)	2357 (76.1%)	848 (27.4%)
Combinations in sample (% of those in full data set 80472)	39665 (49.3%)	19322 (24.0%)	9895 (12.2%)	4664 (5.8%)	986 (1.4%)

Can we reduce the time involved by subsampling secondary schools or combinations in some way without much deleterious effect on precision?

Other experiments on sampling undertaken

Subsampling of secondary schools within primary sample

(i) At random

(ii) Selection of secondary schools with largest representation

Example 2953 Primary schools

Relative SE and Time to full sample

	Secondary selected at random			
	50%	25%	10%	5%
SE	1.4	1.9-2.2	2.8-3.2	3.9-4.2
Time	0.3	0.2	0.1	0.05
	Largest secondary in sample selected			
	50%	25%	10%	5%
SE	1.04-1.2	1.3-1.6	1.7-2.6	2.2-4.0
Time	0.50	0.35	0.21	0.14

Selecting largest slightly slower--larger sample size but also appears to have more non empty combinations

Seems quite a gain in precision

Main problem is potential selection bias: e.g. lower fsm, higher percentage white

Dropping combinations with small cell sizes in primary school samples

Need to go some way before we impact on time
e.g. drop N<10 from 2953 Primary schools sample yields only 50% reduction in time – standard errors increased by factor 1.5. Combination number reduced considerably from 19322 to 2676 though actual number of KS4 and KS2 schools not reduced by much

Considerable evidence of selection bias also.

Stratifying by LEA before sampling primary schools

For fixed number of primary school similar precision to sample of schools at random

Time reduction by 33% mainly because number of non-empty combinations much smaller (though not number of secondary schools involved in them).

Direct sampling of cells in the cross-classification

Directly reduces the number of combinations which influences estimation time

Little impact on number of ks4 schools unless sampling fraction becomes small (For 10% sample there are still 2898 ks4 schools involved)

As expected number of ks2 schools involved falls off more sharply but unless smallish fraction there is still large number (For 25% sample still 70% of schools involved

For samples of comparable size to initial sampling of students estimated SE of all estimates are comparable --- time to fit model though is considerably reduced.

How does it compare with sampling of primary schools?

50 % combinations very similar in relevant aspects to 50% primary
---precision
-- number of combinations
--- number of ks4 schools
--- number of students

even though primary school representation is higher run time is less for combinations sample but by not much less.

Ditto 10%, 5% samples

This leaves some open questions as to why this form of sampling seems slightly more efficient in time.

Other relevant issues on sampling NPD

- What is impact of more complexity such as random KS2 coefficient in this case
- Possibility with sampling of using full database for higher level variables, e.g. school context factors.
  May be possible to draw up separate data base of higher level aggregates which can be drawn on in sampling
  The availability of this means a source of sampling error is controlled
- Sampling higher level units or combinations or multistage samples of students within selected units
- Use of information on schools not evident in sample under investigation
  - data on neighbouring schools
  - historic information on schools attended by sampled students
- Adaptive sampling and topping up
- Other forms of stratification--- e.g. schools by aggregate KS4 information
- Model exploration and validation

Samples of students (i)

	50% 75minutes			25% 32minutes			10% 16minutes	
	Est	**SE**		**Est**	**SE**		**Est**	**SE**
Intercept	-0.0171	0.0049		-0.0155	0.0053		-0.0077	0.0063
KS2 score	0.6275	0.0016		0.6295	0.0022		0.6344	0.0035
Ethnicity (ref White)								
Other White	0.1651	0.0089		0.1516	0.0127		0.1349	0.0202
Black Caribbean	0.0607	0.0114		0.0536	0.0161		0.0351	0.0251
Black African	0.3346	0.0121		0.3491	0.0173		0.3231	0.0268
Indian	0.4072	0.0098		0.3928	0.0137		0.3921	0.0207
Pakistani	0.3276	0.0098		0.3204	0.0138		0.2885	0.021
Chinese	0.5586	0.0234		0.6037	0.0339		0.5204	0.0521
Other	0.1694	0.0062		0.1714	0.0088		0.1662	0.0136
Female	0.1728	0.0026		0.1709	0.0037		0.1705	0.0059
Age	-0.0098	0.0004		-0.0101	0.0005		-0.0099	0.0009
FSM	-0.1932	0.0041		-0.2131	0.0058		-0.2279	0.0092
SEN	-0.3361	0.0037		-0.3374	0.0053		-0.3541	0.0084
Variances								
Secondary School	0.0507	0.0015		0.0509	0.0017		0.0428	0.0019
Primary School	0.0261	0.0007		0.0268	0.0009		0.0252	0.0017
Students	0.3905	0.0011		0.3905	0.0016		0.3928	0.0027
No of Students	271316			136106			54544	
No of ks4 schools	3097			3097			3093	
No of ks2 schools	14710			14485			13207	
No combinations	56698			40559			24937	

Samples of students (ii)

	5% 13minutes			1% 4minutes			0.5% 2minutes	
	Est	**SE**		**Est**	**SE**		**Est**	**SE**
Intercept	-0.0141	0.0081		-0.0226	0.0151		-0.0103	0.0215
KS2 score	0.6418	0.0049		0.6495	0.0109		0.6881	0.0159
Ethnicity (ref White)								
Other White	0.2054	0.0287		0.2709	0.0588		0.0322	0.0959
Black Caribbean	0.0959	0.0371		0.1343	0.0727		-0.1322	0.1263
Black African	0.3361	0.0387		0.3811	0.0755		0.3791	0.1087
Indian	0.4258	0.0291		0.3991	0.0609		0.3563	0.0907
Pakistani	0.318	0.0303		0.2762	0.0629		0.4843	0.0828
Chinese	0.7256	0.0823		0.6352	0.1933		0.3456	0.3987
Other	0.1815	0.0195		0.2671	0.0416		0.2451	0.0596
Female	0.1678	0.0083		0.1494	0.0183		0.1745	0.0263
Age	-0.0115	0.0013		-0.0104	0.0029		-0.0091	0.0041
FSM	-0.2556	0.0131		-0.2727	0.0284		-0.2543	0.0397
SEN	-0.3432	0.0119		-0.3402	0.0262		-0.3123	0.0388
Variances								
Secondary School	0.0503	0.0028		0.0354	0.0077		0.0243	0.0112
Primary School	0.0241	0.0031		0.0169	0.0134		0.0049	0.0075
Students	0.3953	0.0044		0.3981	0.0153		0.4514	0.0169
No of Students	27228			5532			2823	
No of ks4 schools	3085			2496			1797	
No of ks2 schools	11306			4348			2496	
No combinations	16478			4807			2626	

Region analysis

	Region1			Region 2			Region 3			Region 4			Region 5	
	VA			VC			VA							
	Est	SE		Est	SE		Est	SE					Est	SE
Intercept	-0.062	0.015		-0.011	0.014		-0.022	0.015		-0.046	0.018		-0.042	0.011
KS2 score	0.634	0.004		0.651	0.003		0.574	0.003		0.615	0.005		0.628	0.002
Ethnicity (ref White)														
Other White	0.076	0.025		0.098	0.018		0.239	0.011		0.191	0.057		0.082	0.025
Black Caribbean	0.052	0.031		0.025	0.037		0.079	0.011		0.297	0.283		0.113	0.044
Black African	0.579	0.054		0.295	0.039		0.331	0.011		0.471	0.138		0.494	0.046
Indian	0.438	0.021		0.431	0.025		0.416	0.013		0.388	0.081		0.461	0.025
Pakistani	0.386	0.031		0.326	0.028		0.342	0.016		0.362	0.047		0.385	0.018
Chinese	0.532	0.065		0.445	0.052		0.553	0.029		0.677	0.085		0.601	0.043
Other	0.076	0.016		0.135	0.014		0.246	0.008		0.221	0.033		0.171	0.014
Female	0.173	0.006		0.164	0.005		0.152	0.006		0.171	0.007		0.174	0.005
Age	-0.009	0.001		-0.011	0.001		-0.008	0.001		-0.008	0.001		-0.011	0.001
FSM	-0.201	0.01		-0.205	0.011		-0.097	0.008		-0.204	0.011		-0.227	0.007
SEN	-0.319	0.009		-0.297	0.008		-0.361	0.007		-0.314	0.011		-0.318	0.007
Variances														
Secondary School	0.048	0.005		0.053	0.005		0.068	0.005		0.037	0.005		0.045	0.003
Primary School	0.032	0.002		0.026	0.002		0.027	0.002		0.036	0.003		0.029	0.002
Students	0.379	0.002		0.381	0.002		0.422	0.002		0.386	0.002		0.393	0.002
No of Students	49533			60119			63479			29942			92920	
No of ks4 schools	265			338			409			159			470	
No of ks2 schools	2849			3030			2255			1190			3548	
No of ks2 schools in region	1396			1537			1578			773			2499	

	Region 6			Region 7			Region 8			Region 9			Region 9 both schools	
	Est	**SE**		**Est**	**SE**		**Est**	**SE**					**Est**	**SE**
Intercept	0.035	0.012		-0.005	0.014		-0.024	0.014		-0.068	0.012		-0.063	0.013
KS2 score	0.624	0.003		0.666	0.004		0.621	0.003		0.643	0.001		0.644	0.003
Ethnicity (ref White)														
Other White	0.129	0.016		0.032	0.024		0.056	0.022		0.002	0.029		-0.001	0.029
Black Caribbean	0.071	0.037		0.091	0.061		0.102	0.021		0.074	0.041		0.074	0.042
Black African	0.313	0.038		0.448	0.081		0.386	0.044		0.577	0.061		0.499	0.063
Indian	0.345	0.021		0.271	0.053		0.386	0.015		0.411	0.031		0.413	0.03
Pakistani	0.328	0.023		0.211	0.072		0.263	0.016		0.341	0.018		0.337	0.018
Chinese	0.526	0.041		0.651	0.083		0.662	0.062		0.599	0.052		0.563	0.064
Other	0.131	0.013		0.071	0.019		0.121	0.012		0.135	0.017		0.126	0.017
Female	0.161	0.005		0.177	0.006		0.181	0.005		0.184	0.005		0.184	0.005
Age	-0.009	0.001		-0.011	0.001		-0.011	0.001		-0.011	0.001		-0.011	0.001
FSM	-0.217	0.008		-0.216	0.011		-0.171	0.008		-0.204	0.008		-0.203	0.008
SEN	-0.361	0.009		-0.324	0.009		-0.296	0.009		-0.359	0.008		-0.357	0.008
Variances														
Secondary School	0.055	0.004		0.046	0.004		0.056	0.005		0.042	0.004		0.041	0.004
Primary School	0.025	0.001		0.029	0.002		0.027	0.002		0.029	0.002		0.028	0.002
Students	0.384	0.002		0.377	0.002		0.378	0.002		0.386	0.002		0.385	0.002
No of Students	81297			52518			63261			59984			58622	
No of ks4 schools	476			297			371			312			312	
No of ks2 schools	3838			3364			2603			2763			1691	
No of ks2 schools in region	2013			1675			1596			1691			1691	

Samples of LEAs

	5 4.2 mins			14 12 minutes			50 49 mins			45 Stat by R	
	Est	**SE**		**Est**	**SE**		**Est**	**SE**		**Est**	**SE**
Intercept	-0.052	0.023		0.013	0.016		-0.026	0.007		-0.017	0.009
KS2 score	0.626	0.006		0.588	0.004		0.624	0.002		0.624	0.002
Ethnicity (re White)											
Other White	0.181	0.041		0.135	0.018		0.154	0.011		0.18	0.011
Black Caribbean	0.131	0.046		0.053	0.018		0.034	0.014		0.051	0.014
Black African	0.358	0.077		0.328	0.021		0.321	0.016		0.319	0.014
Indian	0.467	0.034		0.338	0.016		0.419	0.012		0.402	0.013
Pakistani	0.444	0.053		0.252	0.017		0.351	0.012		0.349	0.014
Chinese	0.647	0.086		0.619	0.049		0.535	0.028		0.571	0.027
Other	0.120	0.025		0.209	0.012		0.144	0.007		0.203	0.008
Female	0.162	0.009		0.148	0.006		0.165	0.003		0.172	0.004
Age	-0.011	0.001		-0.008	0.001		-0.009	<0.005		-0.009	0.001
FSM	-0.187	0.016		-0.161	0.001		-0.186	0.005		-0.185	0.006
SEN	-0.312	0.014		-0.344	0.008		-0.344	0.004		-0.352	0.005
Variances											
Secondary School	0.045	0.007		0.059	0.005		0.054	0.002		0.054	0.003
Primary School	0.031	0.003		0.027	0.002		0.027	0.001		0.029	0.001
Student	0.391	0.004		0.382	0.002		0.391	0.001		0.401	0.002
No of Students	**18865**			**49037**			**20384**			**143168**	
In same R for KS2 & KS4	**97.5%**			**96.0%**			**96.7%**			**97.5%**	
In same LEA for KS2 & KS4	**92.2%**			**88.1%**			**89.9%**			**87.6%**	
No of ks4 schools	**106**			**347**			**1164**			**844**	
No of ks2 schools	**1330**			**4036**			**9413**			**7119**	
Combinations	**2386**			**9486**			**29559**			**25097**	
Combs in same Region	**1951**			**7537**			**23967**			**22123**	
Combs in same LA	**1424**			5151			17304			14752	

Samples of KS4 schools

	144 6 minutes			300 13 minutes			1002 48minutes	
	Est	**SE**		**Est**	**SE**		**Est**	**SE**
Intercept	-0.041	0.019		-0.044	0.016		-0.025	0.008
KS2 score	0.631	0.005		0.615	0.004		0.629	0.002
Ethnicity (ref White)								
Other White	0.051	0.026		0.143	0.021		0.128	0.011
Black Caribbean	0.076	0.037		0.071	0.024		0.072	0.014
Black African	0.424	0.038		0.301	0.028		0.353	0.015
Indian	0.369	0.033		0.398	0.021		0.385	0.012
Pakistani	0.353	0.031		0.317	0.021		0.348	0.012
Chinese	0.598	0.075		0.642	0.049		0.548	0.028
Other	0.152	0.019		0.17	0.014		0.163	0.008
Female	0.171	0.008		0.164	0.006		0.169	0.003
Age	-0.009	0.001		-0.008	0.001		-0.009	0.001
FSM	-0.186	0.013		-0.189	0.009		-0.194	0.005
SEN	-0.317	0.012		-0.327	0.008		-0.329	0.005
Variances								
Secondary School	0.042	0.006		0.057	0.005		0.054	0.003
Primary School	0.034	0.003		0.036	0.002		0.031	0.001
Student	0.388	0.004		0.391	0.002		0.391	0.001
No of Students	**25571**			**55216**			**179549**	
In same R	**96.7%**			**97.1%**			**97.1%**	
In same LEA	**92.0%**			**90.1%**			**90.0%**	
No of ks4 schools	**144**			**300**			**1022**	
No of ks2 schools	**3693**			**6603**			**12423**	
Combinations	**4470**			**9139**			**32198**	
Combs in same Region	**3803**			**7768**			**27833**	
Combs in same LA	**2860**			**5703**			**20896**	

Samples of KS2 schools (i)

Samples of Primary Schools	7381 (50%) 45mins			2953 (20%) 29mins			1482 (10%) 13 mins	
	Est	**SE**		**Est**	**SE**		**Est**	**SE**
Intercept	-0.0214	0.0052		-0.0261	0.0064		-0.0426	0.0084
KS2 score	0.6316	0.0016		0.6365	0.0025		0.6425	0.0035
Ethnicity (R: White)								
Other White	0.1491	0.0089		0.1671	0.0141		0.1487	0.0196
Black Caribbean	0.0703	0.0114		0.1113	0.0183		0.0744	0.0273
Black African	0.3344	0.0121		0.3709	0.0192		0.3491	0.0273
Indian	0.3894	0.0103		0.3943	0.0153		0.3617	0.0226
Pakistani	0.3267	0.0102		0.3392	0.0163		0.3057	0.0236
Chinese	0.5328	0.0231		0.5437	0.0356		0.5352	0.0544
Other	0.1734	0.0063		0.1707	0.0099		0.1731	0.0143
Female	0.1722	0.0026		0.1692	0.0041		0.1741	0.0057
Age	-0.0089	0.0004		-0.0101	0.0006		-0.0102	0.0009
FSM	-0.1876	0.0041		-0.1897	0.0065		-0.1719	0.0089
SEN	-0.3222	0.0037		-0.3221	0.0058		-0.3139	0.0082
Variances								
Secondary School	0.0511	0.0016		0.0524	0.0021		0.0549	0.0028
Primary School	0.0289	0.0008		0.0307	0.0014		0.0333	0.0022
Students	0.3884	0.0011		0.3873	0.0017		0.3917	0.0024
No of Students	271218			108805			55314	
No of ks4 schools	3095			3075			2911	
No of ks2 schools	7381			2953			1482	
Combinations	39665			19322			9895	

Samples of KS2 schools (ii)

	736 (5%) 10 mins			147 (1%) 2 minutes	
	Est	**SE**		**Est**	**SE**
Intercept	-0.0294	0.0119		-0.0571	0.0248
KS2 score	0.6479	0.0052		0.6598	0.0107
Ethnicity (R: White)					
Other White	0.1456	0.0304		0.2338	0.0586
Black Caribbean	0.1051	0.0398		0.1372	0.0821
Black African	0.3622	0.0405		0.3165	0.0889
Indian	0.3868	0.0336		0.3958	0.0721
Pakistani	0.2888	0.0375		0.3661	0.0882
Chinese	0.4537	0.0782		0.6136	0.1581
Other	0.2034	0.0208		0.1315	0.0432
Female	0.1538	0.0084		0.1903	0.0174
Age	-0.0084	0.0013		-0.0134	0.0026
FSM	-0.1901	0.0133		-0.1358	0.0277
SEN	-0.3067	0.0122		-0.2657	0.0251
Variances					
Secondary School	0.0539	0.0042		0.0541	0.0086
Primary School	0.0361	0.0035		0.0348	0.0084
Students	0.3844	0.0036		0.3769	0.0075
No of Students	25023			5720	
No of ks4 schools	2357			848	
No of ks2 schools	736			147	
Combinations	4664			986	

Selected subsampling of 2953 KS2 schools sample

	Full 2953 school sample 29 minutes		25% of KS4 schools at random 6 minutes		25% largest KS4 schools 10 minutes		Cell sizes <10 omitted 15 minutes		2953 KS2 schools stratified by LA 19 minutes	
	Est	**SE**	**Est**	**SE**	**Est**	**SE**			**Est**	**SE**
Intercept	-0.0261	0.0064	-0.0161	0.0119	-0.0302	0.0093	-0.0212	0.0092	-0.0212	0.0092
KS2 score	0.6365	0.0025	0.6489	0.0048	0.6402	0.0033	0.6411	0.0037	0.6411	0.0037
Ethnicity (ref White)										
Other White	0.1671	0.0141	0.2116	0.0286	0.1368	0.0203	0.1421	0.0243	0.1467	0.0141
Black Caribbean	0.1113	0.0183	0.0848	0.0348	0.1347	0.0305	0.1524	0.0427	0.1023	0.0181
Black African	0.3709	0.0192	0.3264	0.0381	0.3021	0.0316	0.3569	0.0484	0.3287	0.0188
Indian	0.3943	0.0153	0.3721	0.0304	0.3731	0.0218	0.3752	0.0253	0.3922	0.0155
Pakistani	0.3392	0.0163	0.3534	0.0309	0.3451	0.0224	0.3331	0.0268	0.3424	0.0158
Chinese	0.5437	0.0356	0.6668	0.0731	0.5065	0.0515	0.5341	0.0627	0.5801	0.0374
Other	0.1707	0.0099	0.1674	0.0192	0.1904	0.0141	0.1861	0.0173	0.1781	0.0099
Female	0.1692	0.0041	0.1685	0.0079	0.1701	0.0052	0.1736	0.0057	0.1664	0.0041
Age	-0.0101	0.0006	-0.0118	0.0012	-0.0093	0.0008	-0.0096	0.0009	-0.0098	0.0006
FSM	-0.1897	0.0065	-0.1951	0.0124	-0.2023	0.0091	-0.2032	0.0101	-0.1855	0.0064
SEN	-0.3221	0.0058	-0.3211	0.0115	-0.3205	0.0078	-0.3211	0.0087	-0.3196	0.0058
Variances										
Secondary School	0.0524	0.0021	0.0521	0.0041	0.0364	0.0026	0.0423	0.0043	0.0521	0.0021
Primary School	0.0307	0.0014	0.0363	0.0031	0.0318	0.002	0.0248	0.0036	0.0317	0.0014
Students	0.3873	0.0017	0.3871	0.0034	0.3751	0.0023	0.3681	0.0024	0.3865	0.0017
No of Students	108805		28727		59144		73017		108669	
No of ks4 schools	3075		793		769		1887		3081	
No of ks2 schools	2953		2360		2393		2191		2952	
No of combinations	19322		5122		5859		2676		16269	

Sampling of combinations

	50% 40 minutes		25% 26 minutes		10% 11 minutes		5% 6 minutes		1% 1.5 minutes	
	Est	**SE**	**Est**	**SE**	**Est**	**SE**	**Est**	**SE**	**Est**	**SE**
Intercept	-0.0241	0.0051	-0.0383	0.0059	-0.0465	0.0081	-0.0387	0.0101	-0.0631	0.022
KS2 score	0.6284	0.0016	0.6396	0.0022	0.6468	0.0036	0.6355	0.0048	0.6284	0.0115
Ethnicity (ref White)										
Other White	0.1493	0.009	0.1607	0.0127	0.1521	0.0201	0.1739	0.0279	0.0819	0.0588
Black Caribbean	0.0645	0.0116	0.0675	0.0162	0.0809	0.0258	0.0011	0.0377	0.0668	0.0799
Black African	0.3316	0.012	0.3671	0.0171	0.3759	0.0269	0.3085	0.0396	0.4313	0.0916
Indian	0.3958	0.0097	0.3839	0.0135	0.4205	0.0229	0.3675	0.0325	0.4013	0.0747
Pakistani	0.324	0.0101	0.3383	0.0143	0.3556	0.0234	0.3504	0.0332	0.2947	0.0713
Chinese	0.5572	0.0232	0.5837	0.0319	0.5571	0.0529	0.5495	0.0667	0.8535	0.1649
Other	0.1648	0.0062	0.1702	0.0089	0.1856	0.0141	0.1217	0.0193	0.2398	0.047
Female	0.1676	0.0026	0.1703	0.0036	0.1749	0.0059	0.1813	0.0079	0.2038	0.0191
Age	-0.0099	0.0004	-0.0097	0.0005	-0.0096	0.0009	-0.0084	0.0012	-0.0093	0.0029
FSM	-0.1886	0.0041	-0.1806	0.0057	-0.1922	0.0093	-0.1914	0.0127	-0.1676	0.0301
SEN	-0.3335	0.0037	-0.3201	0.0052	-0.3093	0.0084	-0.3286	0.0113	-0.3838	0.028
Variances										
Secondary School	0.0544	0.0017	0.0561	0.002	0.0599	0.0032	0.0547	0.0045	0.0562	0.0131
Primary School	0.0293	0.0008	0.0334	0.0013	0.0393	0.0026	0.0373	0.004	0.0502	0.0138
Students	0.3849	0.0011	0.3865	0.0015	0.3832	0.0025	0.3778	0.0034	0.3751	0.0083
No of Students	268037		138152		52505		28654		4911	
No of ks4 schools	3097		3092		2898		2387		837	
No of ks2 schools	13225		10249		5798		3385		769	