Comparison of joint and separate estimation analyses of the Knowledge and Skills scale of the NAEP 1986 math data

October 7, 2005

Prepared by:

Murray Aitkin and Irit Aitkin
School of Behavioural Science
University of Melbourne

Prepared for:

US Department of Education
Office of Educational Research and Improvement
National Center for Education Statistics

This project was an activity of the Education Statistics Services Institute.

Contents

1 Aim of the project

The aim of this project was to compare the joint (full maximum likelihood) and separate (constrained maximum likelihood) approaches to parameter estimation for the multi-level models used in Project 2 for the analysis of the Knowledge and Skills scale of the 1986 NAEP math test.

2 Summary

- The reporting group estimates obtained by constraining the item parameters fitted in a null model were good approximations to the full ML estimates (less than 0.5 SEs away), provided that
  - three or four levels were fitted in the model in both stages and
  - the variance components were not constrained from the first item fitting stage.
- *All* the constrained models had *substantial* downward biases in the standard errors for the parents education parameters, because of the very high correlations between these estimates and the item parameters.
- The other standard errors were close to their full ML values.
- The two-stage approach of fitting separate models for the items only, and then the reporting group variables for fixed values of the item parameters, saved negligible computing time.
- The model fitting in Stata with Gllamm was very slow. Time savings will come from different algorithms, using analytic derivatives instead of numerical, and parallelizing the numerical integration step.

3 Data and model specification

We used the same data set as in project 2: the 10,463 students clustered in 440 schools, which are themselves clustered in 94 PSUs. High-minority schools were over-sampled to ensure adequate minority student samples. This over-sampling does not require weighting in the analysis as both the school identifier and the student ethnicity are retained in all model analyses except those using the two-level model, in which the school is not identified.

The number of students per school varied from 5 to about 45, with an average of 24, and there was an average of 7 items answered per student. The item responses were coded 0 or 1 according to the manual, with items skipped coded zero and items not reached treated as missing and omitted from the data set.

We used a minimal set of reporting variables: sex, race (6 levels), region (4), size and type of community (stoc, 7) and parents education level (pared, 6), to give us some feel for the results. We used a main effect model with 20 dummy variables for these categorical variables.

In analyzing the test data, we evaluated the three main item response models: the Rasch, 2PL and MIMIC models. For the Rasch and 2PL response models we fitted three multi-level models, corresponding to two-, three- and four-level nested structures. The two-level models (of items nested within subjects) ignored the sample design altogether, the three-level model recognised the school design but ignored the PSU design, and the four-level model recognised the full survey design.

For the MIMIC model we fitted only the two-level nested structure, because of the near- singularity of this model with more levels of nesting. The models were fitted in three different ways:

- by full maximum likelihood (as reported in the final report on Project 2 – we used the results from this project in the comparison);
- by *constrained* maximum likelihood, in which the items were first fitted without any reporting group variables (the "null model") and then the "full" model was fitted with the item parameters held fixed at their estimates from the null model and the reporting group parameters were estimated by maximum likelihood. In the first null model fitting we also estimated the variance components for the models, and these were held fixed also in the second stage.
- by *less constrained* maximum likelihood, in which the items were first fitted as above without any reporting group variables and then the full model was fitted with the item parameters held fixed at their estimates from the null model. However the variance components were *not* constrained, and were estimated in the second stage with the reporting group parameters.

The reporting group parameter estimates were compared to assess the extent of the biases in estimates and standard errors resulting from the constrained maximum likelihood approach. These estimates are shown in the Appendix, where the second less constrained model is labeled Constrained ML (2). We do not give here results for the item parameter estimates. These were reported in Project 2 – they are the estimates from the null models fitted there.

4 Comparisons

We include results for the 2-level models, though these are not of direct interest because the 2-level models are inadequate to represent the sampling variability among schools in the two-stage design.

4.1 2-level models

4.1.1 Log-likelihoods

The log-likelihood decrease for the constrained model over the fully maximized model is quite large in every case – 20.95 for the Rasch, 40.14 for the 2PL and 29.96 for the MIMIC. The decrease for the less constrained model is much smaller – 0.90 for the Rasch, 21.16 for the 2PL and 9.21 for the MIMIC.

4.1.2 Parameter estimates

The Rasch parameter estimates were very little affected by the constraints, the only major change in the constrained model being the poorly determined "other" ethnic group parameter. In the less constrained model the size and type of community parameter estimates were all reduced by 0.003 and those for parents education were all reduced by .013 relative to the fully maximized model – the parameter estimates were very similar apart from an origin shift.

The 2PL estimates changed more in the constrained model, with the sex difference increased by 0.35 SE, and the lowmetro and maincity estimates increased by .7 SE and .45 SE respectively. The nores estimate increased by .5 SE. In the less constrained model the parameter estimates moved towards the full ML estimates, but were still well away, as indicated by the substantial log-likelihood difference (21.16).

The MIMIC estimates changed more than the 2PL, the black, Hispanic and American Indian gaps increasing by .5 SE, .4 SE and .4 SE respectively. The finhighschool, somecoll and nores estimates decreased by .4 SE, .3 SE and .3 SE respectively. In the less constrained model the parameter estimates moved slightly towards the full ML estimates, except for the community estimates, which were the smallest.

4.1.3 Standard errors

The Rasch standard errors were almost unchanged by the item parameter constraints, except for a slight increase for the ethnic group parameters and a substantial – 35-50 % – reduction in those of the parents education parameters. In the less constrained model they were systematically smaller than in the fully constrained model, probably because the variance estimate was correspondingly smaller.

The 2PL standard errors were similar to those of the Rasch, and were similarly smaller in the less constrained model.

The MIMIC standard errors showed the same pattern except for substantial reductions – 55-70% – in the black and Hispanic ethnic group parameters. In the less constrained model all the standard errors were further reduced by the smaller residual variance.

4.2 3-level models

4.2.1 Log-likelihoods

The log-likelihood decrease for the constrained model for the Rasch was similar to that for the 2-level model – 23.04. The less constrained model was almost the same as the fully maximized model – a log-likelihood decrease of 0.14.

For the 2PL, the constrained model converged to a vastly inferior maximum of the log-likelihood – 6,400 worse than the full model. This probably occurred because of the constraints on the variance components, as the constrained school variance component was more than twice the size of the full ML estimate. The less constrained model was very close to the fully maximized model – a log- likelihood decrease of only 3.46.

4.2.2 Parameter estimates

For the Rasch, the lomet, DK and nores estimates in the constrained model all changed by 0.5 SE, the other changes were small. For the less constrained model, the parameter estimates were close to the full ML estimates except that the region estimates were smaller by 0.008, the stoc estimates were smaller by .019, and the pared estimates were smaller by .076.

For the 2PL, many parameter estimates were changed by 0.5 to 1 SE in the constrained model. In the less constrained model, the estimates were almost the same as the full ML estimates except for the pared estimates, which were all increased by .007.

4.2.3 Standard errors

For the constrained Rasch, the student demographic standard errors were slightly increased, those of the region and stoc variables were increased more – 20 to 30% – and those of the parents education variables were *decreased* by 35% by the constraints. For the less constrained model, all the standard errors at the upper levels were slightly reduced compared to the constrained model.

For the 2PL, the changes were similar but smaller except for the stoc variable, where the standard errors were decreased by 30-50% by the constraints. For the less constrained model all standard errors except that of sex were slightly reduced compared to the constrained model.

4.3 4-level models

4.3.1 Log-likelihoods

The log-likelihood decrease for the constrained model for the Rasch was similar to that for the 2- and 3-level models – 21.50. For the less constrained model it was much smaller – 0.26.

For the 2PL, the decrease was 20.32, and 4.14 for the less constrained model.

4.3.2 Parameter estimates

For the Rasch, many estimates were changed by 0.4-0.7 SE in the constrained model. In the less constrained model the region, pared and stoc parameter estimates changed by an origin shift relative to the full ML estimates – the pared estimates decreased by .137, the stoc estimates by .32-.36 and the region estimates by .21-.26.

For the 2PL, the black estimate changed by 0.8 SE in the constrained model. Other changes were small. In the less constrained model the black estimate was changed by 0.15 SE, and all other changes were very small.

4.3.3 Standard errors

For the Rasch, the changes in the constrained model were the same as for the 3-level model – the student demographic standard errors were slightly increased, those of the region and stoc variables were increased more – 20 to 30% – and those of the parents education variables were *decreased* by 35% by the constraints. For the less constrained model, all the standard errors at the upper levels were slightly reduced compared to the constrained model.

For the 2PL, the standard errors in the constrained model were slightly increased except for the pared values, which were reduced by 35%. For the less constrained model, the standard errors were smaller than those for the full Ml estimates, and much smaller – up to 50% – for the pared estimates.

5 Computing issues

Computational time for the four-level 2PL model was very substantial. However the four-level models were unnecessary as the log-likelihood change was small and the variance component at this level was very close to zero (though formally significantly different from zero). The small variance component was at least partly responsible for the very slow computing for this model.

The two-stage approach of fitting separate models for the items only, and then the reporting group variables for fixed values of the item parameters, did not save appreciable overall computing time – the first stage model was always very slow even with no explanatory variables, and the second stage was very slow for the four-level models.

6 Conclusions

The reporting group estimates obtained by constraining the item parameters fitted in a null model were good approximations to the full ML estimates (less than 0.5 SEs away, and with very similar log-likelihoods), provided that three or four levels were fitted in the model and that the variance components were not constrained.

Estimates from the 2-level MIMIC model had large biases in the black and Hispanic group parameters.

All of the constrained models had downward biases in the standard errors for the parents education parameters: because of the very high correlations between these estimates and the item parameters, the standard errors were substantially underestimated when the item parameters were fixed.

This method of approximate likelihood maximization did not provide correct standard errors for the parents education parameters. The method can be *iterated*, when it becomes the *hill climbing* or *successive relaxation* method – by then holding the reporting group parameters fixed, and re-estimating the item parameters, and continuing this alternate fixing and *relaxing* till convergence to the full ML estimates – but it still does not give correct standard errors. To obtain correct standard errors the full joint model has to be fitted finally, and more than one iteration might be needed for this.

Extending the alternate maximization until convergence and then fitting the joint model is unlikely to save any computation time, and may take even longer than direct maximization. Time savings will come from different algorithms, using analytic derivatives instead of numerical, and parallelizing the numerical integration step.

7 Acknowledgements

This work was carried out with the use of the computers of the Victorian Partnership for Advanced Computing, and with assistance from the High Performance Computing Group at the University of Melbourne. We appreciate advice from Sophia Rabe-Hesketh on the use of Gllamm.

8 Reference

Skrondal, A. and Rabe-Hesketh, S. (2004) *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Relations Models.* Chapman and Hall/CRC, Boca Raton, FL.

9 Appendix – model parameter estimates and SEs

Rasch variance component estimates and SEs for the 2-level models

	full ML	constrained ML	constrained ML(2)
male	0	0	0
femal	.111(.031)	.109(.033)	.111(.031)
white	0	0	0
black	-.830(.045)	-.837(.048)	-.829(.045)
hispa	-.546(.045)	-.556(.047)	-.546(.044)
as/pa	-.214(.126)	-.225(.131)	-.214(.126)
amind	-.599(.113)	-.593(.123)	-.598(.113)
other	-.121(.783)	-.169(.870)	-.133(.781)
NE	0	0	0
SE	-.033(.051)	-.036(.052)	-.035(.050)
Cent	-.245(.051)	-.247(.053)	-.245(.050)
West	-.209(.046)	-.210(.048)	-.209(.045)
extru	0	0	0
lomet	-.224(.083)	-.225(.083)	-.226(.078)
himet	.508(.081)	.524(.079)	.504(.076)
manct	.141(.079)	.147(.080)	.138(.076)
urbfr	.161(.081)	.166(.079)	.157(.075)
medct	.079(.074)	.079(.074)	.076(.070)
smplc	-.066(.072)	-.066(.071)	-.069(.067)
nfnhs	0	0	0
finhs	-.105(.159)	-.111(.109)	-.117(.103)
smcol	.120(.146)	.107(.085)	.107(.081)
colgr	.593(.154)	.590(.100)	.580(.095)
DK	.591(.142)	.581(.078)	.578(.074)
nores	.165(.141)	.153(.077)	.152(.073)
s^2	1.311(.039)	1.558(0)	1.307(.037)
logL	-40,475.22	-40,496.17	-40,476.12

2PL variance component estimates and SEs for the 2-level models

	full ML	constrained ML	constrained ML(2)
male	0	0	0
femal	.015(.028)	.025(.029)	.036(.028)
white	0	0	0
black	-.792(.041)	-.769(.042)	-.768(.041)
hispa	-.514(.041)	-.498(.042)	-.499(.041)
as/pa	-.240(.114)	-.226(.118)	-.220(.115)
amind	-.588(.092)	-.574(.098)	-.573(.095)
other	-.380(.855)	-.397(1.07)	-.307(.912)
NE	0	0	0
SE	.003(.046)	.030(.048)	.021(.046)
Cent	-.219(.046)	-.210(.048)	-.215(.046)
West	-.193(.042)	-.179(.043)	-.183(.042)
extru	0	0	0
lomet	-.126(.075)	-.074(.074)	-.098(.072)
himet	.475(.074)	.484(.072)	.477(.070)
manct	.133(.072)	.161(.072)	.152(.069)
urbfr	.121(.074)	.148(.071)	.139(.069)
medct	.031(.067)	.043(.066)	.044(.064)
smplc	-.055(.066)	-.037(.064)	-.045(.062)
nfnhs	0	0	0
finhs	-.205(.152)	-.134(.098)	-.127(.095)
smcol	.061(.141)	.131(.077)	.132(.074)
colgr	.460(.149)	.517(.090)	.522(.088)
DK	.453(.138)	.512(.071)	.521(.069)
nores	.047(.137)	.119(.070)	.125(.068)
s^2	1.646(.345)	1.574(0)	1.330(.037)
logL	-40,077.26	-40,117.40	-40,098.42

MIMIC variance component estimates and SEs for the 2-level models

	full ML	constrained ML	constrained ML(2)
male	0	0	0
femal	.121(.033)	.125(.033)	.131(.031)
white	0	0	0
black	-.780(.090)	-.830(.048)	-.813(.045)
hispa	-.526(.069)	-.554(.048)	-.545(.045)
as/pa	-.188(.117)	-.209(.124)	-.194(.121)
amind	-.619(.121)	-.670(.116)	-.644(.109)
other	.230(.672)	.340(.598)	.271(.667)
NE	0	0	0
SE	-.045(.049)	-.043(.054)	-.051(.050)
Cent	-.230(.055)	-.246(.055)	-.245(.051)
West	-.213(.049)	-.232(.049)	-.229(.046)
extru	0	0	0
lomet	-.207(.085)	-.206(.087)	-.235(.081)
himet	.505(.094)	.537(.083)	.504(.078)
manct	.174(.079)	.195(.082)	.168(.077)
urbfr	.173(.082)	.179(.082)	.157(.078)
medct	.079(.073)	.064(.077)	.062(.073)
smplc	-.036(.072)	-.043(.074)	-.058(.070)
nfnhs	0	0	0
finhs	-.086(.154)	-.149(.119)	-.138(.109)
smcol	.145(.140)	.102(.090)	.108(.084)
colgr	.595(.160)	.583(.102)	.577(.097)
DK	.594(.149)	.577(.081)	.577(.077)
nores	.194(.136)	.154(.080)	.160(.076)
s^2	1.185(.244)	1.574(0)	1.322(.036)
logL	-40,080.32	-40,110.28	-40,089.53

Rasch variance component estimates and SEs for the 3-level models

	full ML	constrained ML	constrained ML(2)
male	0	0	0
femal	.107(.031)	.103(.031)	.106(.031)
white	0	0	0
black	-.703(.050)	-.680(.052)	-.704(.050)
hispa	-.492(.047)	-.485(.050)	-.492(.047)
as/pa	-.196(.129)	-.213(.134)	-.198(.129)
amind	-.489(.106)	-.464(.110)	-.490(.106)
other	-.054(.718)	-.039(.714)	-.072(.716)
NE	0	0	0
SE	-.056(.081)	-.045(.096)	-.064(.079)
Cent	-.204(.080)	-.183(.088)	-.213(.077)
West	-.204(.074)	-.190(.092)	-.211(.072)
extru	0	0	0
lomet	-.310(.127)	-.372(.161)	-.329(.118)
himet	.511(.125)	.507(.137)	.492(.117)
manct	.143(.119)	.147(.142)	.128(.114)
urbfr	.188(.125)	.195(.147)	.169(.116)
medct	.113(.112)	.125(.142)	.097(.106)
smplc	-.038(.108)	-.067(.131)	-.055(.101)
nfnhs	0	0	0
finhs	-.033(.216)	-.119(.143)	-.109(.124)
smcol	.186(.208)	.094(.130)	.111(.110)
colgr	.607(.214)	.513(.140)	.531(.120)
DK	.588(.206)	.484(.126)	.512(.105)
nores	.221(.205)	.127(.126)	.146(.105)
s^2_sch	.154(.019)	.325(0)	.154(.019)
s^2	1.168(.037)	1.246(0)	1.167(.035)
logL	-40,349.04	-40,372.08	-40,349.18

2PL variance component estimates and SEs for the 3-level models

	full ML	constrained ML	constrained ML(2)
male	0	0	0
femal	.012(.028)	.003(.028)	.013(.028)
white	0	0	0
black	-.667(.047)	-.642(.049)	-.667(.046)
hispa	-.460(.043)	-.428(.044)	-.461(.043)
as/pa	-.203(.117)	-.210(.120)	-.204(.116)
amind	-.471(.093)	-.453(.095)	-.471(.092)
other	-.200(.752)	-.208(.836)	-.198(.748)
NE	0	0	0
SE	-.020(.077)	-.059(.087)	-.019(.075)
Cent	-.172(.074)	-.135(.087)	-.172(.072)
West	-.182(.069)	-.183(.078)	-.181(.068)
extru	0	0	0
lomet	-.201(.113)	-.118(.134)	-.201(.107)
himet	.497(.116)	.536(.119)	.497(.111)
manct	.150(.106)	.109(.106)	.150(.102)
urbfr	.158(.112)	.227(.122)	.158(.106)
medct	.092(.097)	.114(.106)	.092(.092)
smplc	-.019(.095)	-.003(.095)	-.019(.089)
nfnhs	0	0	0
finhs	-.179(.206)	-.229(.120)	-.171(.111)
smcol	.045(.200)	-.011(.106)	.052(.098)
colgr	.398(.205)	.352(.115)	.405(.108)
DK	.382(.198)	.314(.104)	.389(.095)
nores	.027(.197)	-.021(.103)	.034(.094)
s^2_sch	.139(.017)	.285(0)	.139(.017)
s^2	1.682(.365)	1.665(0)	1.684(.049)
logL	-39,930.05	-46,335.50	-39,933.51

Rasch variance component estimates and SEs for the 4-level models

	full ML	constrained ML	constrained ML(2)
male	0	0	0
femal	.107(.031)	.106(.031)	.105(.031)
white	0	0	0
black	-.707(.050)	-.673(.052)	-.708(.050)
hispa	-.491(.047)	-.482(.049)	-.492(.047)
as/pa	-.199(.129)	-.218(.133)	-.205(.129)
amind	-.493(.105)	-.481(.107)	-.496(.105)
other	-.043(.717)	-.069(.732)	-.075(.714)
NE	0	0	0
SE	-.050(.095)	-.091(.108)	-.071(.092)
Cent	-.203(.098)	-.213(.122)	-.229(.093)
West	-.184(.089)	-.214(.103)	-.205(.086)
extru	0	0	0
lomet	-.300(.126)	-.366(.134)	-.336(.118)
himet	.471(.127)	.407(.151)	.435(.118)
manct	.137(.123)	.104(.144)	.103(.114)
urbfr	.179(.125)	.129(.138)	.137(.114)
medct	.069(.103)	.047(.131)	.039(.106)
smplc	-.052(.107)	-.096(.119)	-.085(.099)
nfnhs	0	0	0
finhs	.017(.213)	-.103(.145)	-.119(.127)
smcol	.234(.205)	.111(.132)	.097(.113)
colgr	.655(.211)	.525(.142)	.518(.124)
DK	.636(.202)	.498(.129)	.498(.109)
nores	.270(.202)	.143(.127)	.133(.108)
s^2_PSU	.036(.013)	.059(0)	.037(.014)
s^2_sch	.116(.019)	.257(0)	.115(.019)
s^2	1.169(.037)	1.243(0)	1.168(.035)
logL	-40,342.85	-40,364.35	-40,343.11

2PL variance component estimates and SEs for the 4-level models

	full ML	constrained ML	constrained ML(2)
male	0	0	0
femal	.012(.028)	.015(.028)	.019(.028)
white	0	0	0
black	-.670(.046)	-.632(.047)	-.662(.046)
hispa	-.454(.043)	-.439(.045)	-.450(.043)
as/pa	-.206(.116)	-.205(.120)	-.198(.117)
amind	-.471(.093)	-.444(.095)	-.467(.093)
other	-.174(.739)	-.136(.746)	-.139(.726)
NE	0	0	0
SE	-.036(.080)	-.025(.095)	-.029(.078)
Cent	-.176(.087)	-.191(.105)	-.172(.084)
West	-.198(.077)	-.182(.089)	-.192(.076)
extru	0	0	0
lomet	-.189(.120)	-.171(.132)	-.180(.111)
himet	.481(.119)	.467(.136)	.483(.111)
manct	.188(.121)	.189(.137)	.194(.114)
urbfr	.176(.122)	.161(.133)	.179(.114)
medct	.061(.110)	.048(.127)	.064(.103)
smplc	-.026(.102)	-.028(.116)	-.025(.094)
nfnhs	0	0	0
finhs	-.153(.201)	-.119(.136)	-.132(.118)
smcol	.065(.194)	.091(.125)	.084(.106)
colgr	.422(.199)	.437(.134)	.436(.115)
DK	.403(.192)	.419(.122)	.420(.103)
nores	.048(.191)	.075(.121)	.071(.102)
s^2_PSU	.019(.010)	.033(0)	.019(.010)
s^2_sch	.119(.017)	.212(0)	.120(.017)
s^2	1.653(.360)	1.634(0)	1.521(.045)
logL	-39,924.60	-39,944.92	-39,928.74