Cross-Classified Models in the Context of Value-Added Modeling [1]

J. Kyle Roberts Douglas Bates

Southern Methodist University

University of Wisconsin-Madison

March 6, 2010

[1] Paper presented at the annual meeting of the Society for Research on Educational Effectiveness, March 6, 2010, Washington, D.C. A copy of the full paper may be obtained at http://faculty.smu.edu/kyler/presen/

Background to Value-Added Modeling

- Initially, VAMs focused on model development (McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004; Tekwe et al., 2004) and the development of software routines appropriate for estimating these models.
- Work by Doran, Bates, and Phillips (under review) and work presented in the special issue of the Journal of Educational and Behavioral Statistics, volume 29(1) have made considerable strides above and beyond initial VAMs.

Current Models for VAMs

- One of the first true VAMs was introduced by Sanders, Saxton & Horn (1997) on the Tennessee VAM System.
- VAMs typically fall into one of two categories: models appropriate for monitoring school effects on student learning and models appropriate for monitoring teacher effects on student learning.
- These models typically take the form:

$$Y_{ti} = [\beta_0 + \beta_1(year_t)] + \left[\theta_{0j(i)} + \theta_{1j(i)}(year_t) + \delta_{0i} + \delta_{1i}(year_t) + \epsilon_{ti}\right]$$

where i indexes students and j could index either the schools or teachers. The notation $j(i)$ represents the school (or teacher) attended by student i in year t. In this model, the within-group errors are assumed IID where $\epsilon_{ti} \sim \mathcal{N}(0, \sigma^2)$ and school (or teacher) and student random effects are $\boldsymbol{\theta}_j = (\theta_{0j}, \theta_{1j}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$ and $\boldsymbol{\delta}_i = (\delta_{0i}, \delta_{1i}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$, respectively.

McCaffrey et al. (2003) Recommendations for Future Research

1. Develop databases that can support VAM estimation of teacher effects.
2. Develop computational tools for fitting VAM
3. Link VAM teacher-effect estimates to alternative measures of teacher effectiveness
4. Empirically evaluate the potential sources of errors we have identified
5. Estimate the prevalence of factors that contribute to the sensitivity of teacher-effect estimates
6. Incorporate decision theory into VAM
7. Use research and auxiliary data to inform modeling choices

The Strength of Cross-Classified VAMs

A case with true nesting.

	Middle School					
HS	MS 1	MS 2	MS 3	MS 4	MS 5	MS 6
HS 1	S1, S2	S3, S4, S5				
HS 2			S6, S7	S8, S9		
HS 3					S10, S11	S12, S13

A case with cross-clasification.

	Middle School					
HS	MS 1	MS 2	MS 3	MS 4	MS 5	MS 6
HS 1	S1, S2	S3	S4		S5	
HS 2		S6, S7		S8, S9, S10		
HS 3	S11		S12		S13	S14, S15

Notating the Cross-Classified Model

When each student is nested inside one and only one middle school and these middle schools are nested inside three high schools. This data may be modeled in a mixed-effects model as:

$$Y_{ijk} = \gamma_{000} + u_{00k} + u_{0jk} + e_{ijk}$$

where γ_{000} represents the average student achievement across all j middle schools and k high schools with student-level residuals e_{ijk} and middle school-level random effects u_{0jk} and high school-level random effects u_{00k}.

Notating the Cross-Classified Model

In the cross-classified example, student scores are modeled as:

$$Y_{i(j_1,j_2)} = \gamma_{000} + u_{00j_1} + u_{00j_2} + u_{00j_1xj_2} + e_{i(j_1,j_2)}$$

where γ_{000} represents the average student achievement for student i having attended middle school j_1 and high school j_2 with random effects u_{00j_1} for the middle schools, u_{00j_2} for the high schools, a random interaction effect $u_{00j_1xj_2}$ for belonging to both middle school j_1 and high school j_2 and student-level residuals $e_{i(j_1,j_2)}$. This parameterization follows the Rasbash and Browne (2001) notation.

Rethinking the Cross-Classified VAM

- We might rethink the value-added model as

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_n),\ \boldsymbol{b} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$$

  where $\boldsymbol{\theta}$ is the vector of variance components to be estimated. For the model we are describing $\boldsymbol{\theta}$ would have a total of nine elements; two variances and one covariance for each of the student, teacher and school classifications.
- Because the matrix $\boldsymbol{Z}$ can be huge (perhaps millions of rows and hundreds of thousands of columns) this representation would be interesting but completely impractical if it were not for the fact that $\boldsymbol{Z}$ is very, very sparse.

Cholesky Decomposition in Sparse Matrix Theory

The broad outline of this decomposition to express the variance-covariance of the random effects, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, in terms of a relative covariance factor, $\boldsymbol{\Lambda}(\boldsymbol{\theta})$, defined so that

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \sigma^2 \Lambda(\boldsymbol{\theta})\boldsymbol{\Lambda}'(\boldsymbol{\theta})$$

and optimize the profiled deviance, defined as

$$-2\tilde{\ell}(\boldsymbol{\theta}) = \log(|\boldsymbol{L}(\boldsymbol{\theta})|^2) + n\left[1 + \frac{2\pi r^2(\boldsymbol{\theta})}{n}\right]$$

where $\boldsymbol{L}(\boldsymbol{\theta})$ is the sparse, lower-triangular Cholesky factor satisfying

$$\boldsymbol{L}(\boldsymbol{\theta})\boldsymbol{L}'(\boldsymbol{\theta}) = \boldsymbol{\Lambda}(\boldsymbol{\theta})'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Lambda}(\boldsymbol{\theta}) + \boldsymbol{I}.$$

Graphical Representation of the Cholesky Decomposition

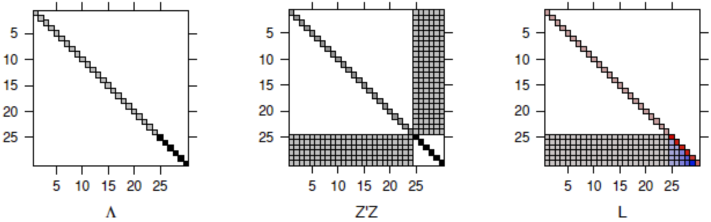


Fig. 2.4 Images of the relative covariance factor, Λ, the cross-product of the random-effects model matrix, $\mathbf{Z}^{\mathsf{T}}\mathbf{Z}$, and the sparse Cholesky factor, $\mathbf{L}$, for model `fm2`.

An Example with Only One Time Point

Our first model will be the "null" or unconditional model such that,

$$Y_{i(j_1,j_2)} = \gamma_{000} + u_{00j_1} + u_{00j_2} + e_{i(j_1,j_2)}$$

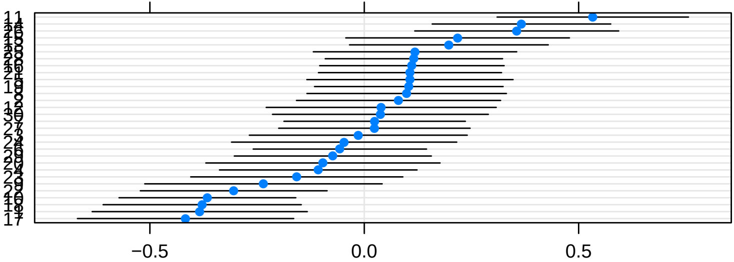
This is easily modeled in lme4 as:

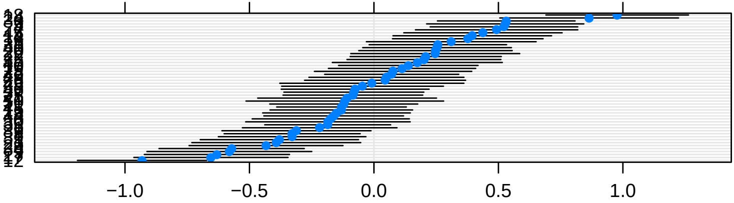
```
> m0 <- lmer(ACHIEV ~ 1 + (1 | PSCHOOL) + (1 | SSCHOOL),
+      REML = F, pupils)
> summary(m0)
```

```
Linear mixed model fit by maximum likelihood
Formula: ACHIEV ~ 1 + (1 | PSCHOOL) + (1 | SSCHOOL)
   Data: pupils
  AIC  BIC logLik deviance REMLdev
 2326 2345  -1159     2318    2321
Random effects:
 Groups   Name        Variance Std.Dev.
 PSCHOOL  (Intercept) 0.169348 0.41152
 SSCHOOL  (Intercept) 0.065401 0.25574
 Residual             0.513169 0.71636
Number of obs: 1000, groups: PSCHOOL, 50; SSCHOOL, 30

Fixed effects:
```

Examining the Distributions for Secondary and Primary





Adding Covariates

In addition to model m0, we can test for the effect of pupil ses level and pupil gender as fixed effects such that:

$$Y_{i(j_1,j_2)} = \gamma_{000} + \gamma_{100}\text{gender}_{i(j_1,j_2)} + \gamma_{200}\text{ses}_{i(j_1,j_2)} + u_{00j_1} + u_{00j_2} + e_{i(j_1,j_2)}$$

```
> m1 <- lmer(ACHIEV ~ PUPSEX + PUPSES + (1 | PSCHOOL) +
+      (1 | SSCHOOL), REML = F, pupils)
> summary(m1)
```

```
Linear mixed model fit by maximum likelihood
Formula: ACHIEV ~ PUPSEX + PUPSES + (1 | PSCHOOL) + (1 | SSCHOOL)
   Data: pupils
  AIC  BIC logLik deviance REMLdev
 2255 2285  -1122     2243    2258
Random effects:
 Groups   Name        Variance Std.Dev.
 PSCHOOL  (Intercept) 0.169010 0.41111
 SSCHOOL  (Intercept) 0.063606 0.25220
 Residual             0.474255 0.68866
Number of obs: 1000, groups: PSCHOOL, 50; SSCHOOL, 30

Fixed effects:
            Estimate Std. Error t value
(Intercept)  5.75548    0.10527   54.67
PUPSEXgirl   0.26131    0.04564    5.73
PUPSES       0.11409    0.01610    7.09

Correlation of Fixed Effects:
           (Intr) PUPSEX
```

Adding Random Effects

Furthermore, we may wish to model random effects for `gender` across all primary schools (and not secondary schools). This would be specified as:

$$
\begin{aligned}
Y_{i(j_1,j_2)} = & \gamma_{000} + \gamma_{100}\mathsf{gender}_{i(j_1,j_2)} + (\gamma_{200} + u_{10j_1})\mathsf{ses}_{i(j_1,j_2)} \\
& + u_{00j_1} + u_{00j_2} + e_{i(j_1,j_2)}
\end{aligned}
$$

```
> m2 <- lmer(ACHIEV ~ PUPSEX + PUPSES + (PUPSEX |
+     PSCHOOL) + (1 | SSCHOOL), REML = F, pupils)
> summary(m2)
```

```
Linear mixed model fit by maximum likelihood
Formula: ACHIEV ~ PUPSEX + PUPSES + (PUPSEX | PSCHOOL) + (1 | SSCHOOL)
   Data: pupils
  AIC  BIC logLik deviance REMLdev
 2257 2296  -1120     2241    2255
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 PSCHOOL  (Intercept) 0.144265 0.37982
          PUPSEXgirl  0.016480 0.12838  0.463
 SSCHOOL  (Intercept) 0.063742 0.25247
 Residual             0.470093 0.68563
Number of obs: 1000, groups: PSCHOOL, 50; SSCHOOL, 30

Fixed effects:
            Estimate Std. Error t value
(Intercept)  5.75892    0.10278   56.03
PUPSEXgirl   0.26551    0.04900    5.42
PUPSES       0.11301    0.01608    7.03

Correlation of Fixed Effects:
```

Time-Series Cross-Classified

Suppose that we have a situation where we are doing continuous progress monitoring on students three times a year across two years.

```
> xtabs(~time + cteach, vamdata)

    cteach
time 10 11 12 13 14 15 20 21 23 24
   0  2  2  2  2  2  2  0  0  0  0
   1  2  2  2  2  2  2  0  0  0  0
   2  2  2  2  2  2  2  0  0  0  0
   3  0  0  0  0  0  0  3  3  3  3
   4  0  0  0  0  0  0  3  3  3  3
   5  0  0  0  0  0  0  3  3  3  3
```

```
> vam1 <- lmer(achieve ~ time + (time | cteach) +
+      (1 | id), vamdata)
> summary(vam1)
```

```
Linear mixed model fit by REML
Formula: achieve ~ time + (time | cteach) + (1 | id)
   Data: vamdata
   AIC   BIC logLik deviance REMLdev
 289.7 305.6 -137.8    276.9   275.7
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 7.5470   2.7472
 cteach   (Intercept) 5.6589   2.3788
          time        1.1774   1.0851   -1.000
 Residual             1.0384   1.0190
Number of obs: 72, groups: id, 12; cteach, 10

Fixed effects:
            Estimate Std. Error t value
(Intercept)   2.1763     1.1172   1.948
time          1.2613     0.3535   3.568

Correlation of Fixed Effects:
     (Intr)
```

Random Effects for Teachers

```
> ranef(vam1)$cteach
   (Intercept)       time
10  -2.0531202  0.9365056
11   2.1450473 -0.9784371
12  -1.9189711  0.8753152
13   0.5772548 -0.2633077
14  -2.4467006  1.1160325
15   1.4752146 -0.6729011
20   2.0513364 -0.9356920
21  -2.9085858  1.3267158
23   3.5448827 -1.6169549
24  -0.4663579  0.2127235
```