A Generalized Value-Added Model with Conditional Random Effects and Multivariate Shrinkage

Robert H. Meyer, Director
Value-Added Research Center
Wisconsin Center for Education Research
University of Wisconsin-Madison

A. Equivalent Representations of a Linear Value-Added Growth Process

(1) Value-Added Student Achievement Growth Trajectory

$$Y_{it} = (\phi_i + \gamma_i t) + \sum_{s=1}^{s=t} S_{si}{}' \alpha_s + E_{ti} + v_{ti} \tag{1}$$

Contrast: nonlinear growth model with growth curvature parameter λ:

$$Y_{it} = \lambda^t \phi_i + \left(\frac{1-\lambda^t}{1-\lambda} \right) \gamma_i + \sum_{s=1}^{s=t} S_{si}{}' \alpha_s + E_{ti} + v_{ti} \tag{2}$$

Meyer and Christian (2008)

(2) Value-Added Equations in Reduced Form

$$
\begin{aligned}
Y_{0i} &= \phi_i + E_{0i} + v_{0i} \\
Y_{1i} &= S_{1i}{}'\alpha_1 + (\phi_i + \gamma_i) + E_{1i} + v_{1i} \\
Y_{2i} &= S_{1i}{}'\alpha_1 + S_{2i}{}'\alpha_2 + (\phi_i + 2\gamma_i) + E_{2i} + v_{2i}
\end{aligned}
\tag{3}
$$

(3) Value-Added Equations in Growth (Difference) Form

$$
\begin{aligned}
Y_{0i} &= \phi_i + E_{0i} + v_{0i} \\
Y_{1i} - Y_{0i} &= S_{1i}{}'\alpha_1 + \gamma_i + e_{1i} + v_{1i} - v_{0i} \\
Y_{2i} - Y_{1i} &= S_{2i}{}'\alpha_2 + \gamma_i + e_{2i} + v_{2i} - v_{1i}
\end{aligned}
\tag{4}
$$

(4) Value-Added Equations in Growth Form with School and Classroom Effects

$$
\begin{aligned}
Y_{0i} &= \phi_i + E_{0i} + v_{0i} \\
Y_{1i} - Y_{0i} &= S_{1i}{}'\alpha_1 + C_{1i}{}'\beta_1 + \gamma_i + e_{1i} + v_{1i} - v_{0i} \\
Y_{2i} - Y_{1i} &= S_{2i}{}'\alpha_2 + C_{2i}{}'\beta_2 + \gamma_i + e_{2i} + v_{2i} - v_{1i}
\end{aligned}
\tag{5}
$$

Total classroom effects:

$$
\begin{aligned}
(\alpha\beta)_{1jk} &= \alpha_{1k} + \beta_{1jk} \\
(\alpha\beta)_{2jk} &= \alpha_{2k} + \beta_{2jk}
\end{aligned}
\tag{6}
$$

B. Alternative T3 Model Specifications of Student Effects

1. Fixed Individual Effects (FE)
2. Random Individual Effects (RE)
3. Conditional Random Effects (CRE)
4. Conditional Random Effects with Multivariate Shrinkage (CRE-MS)

Focus on student effects in variance components structure:

γ_i = persistent student growth effect
ϕ_i = initial (year *0*) growth differences (apart from accumulated random effects

Issue: Is individual growth effect γ_i *correlated* with school and classroom enrollment variables?

Assumption:

- Strict exogeneity: $E(Y_{ti} - Y_{t-1i} \mid S_{1i}, S_{2i}, \gamma_i) = S_{ti}{}'\alpha_t + \gamma_i$.

(1) **Fixed Effects (FE)**

(Double difference)

School Value-Added Model

$$(Y_{2i} - Y_{1i}) - (Y_{1i} - Y_{0i}) = S_{2i}{}'\alpha_2 - S_{1i}{}'\alpha_1 + (e_{2i} - e_{1i}) + (v_{2i} - v_{1i}) - (v_{1i} - v_{0i}) \quad (7)$$

School and Classroom Value-Added Model

$$\begin{aligned}(Y_{2i} - Y_{1i}) - (Y_{1i} - Y_{0i}) &= S_{2i}{}'\alpha_2 - S_{1i}{}'\alpha_1 + C_{2i}{}'\beta_2 - C_{1i}{}'\beta_1 \\ &+ (e_{2i} - e_{1i}) + (v_{2i} - v_{1i}) - (v_{1i} - v_{0i})\end{aligned} \quad (8)$$

Comments on FE Model:

a. Eliminates student selection/assignment bias (given the assumption of strict exogeneity).
b. Year 1 and 2 school enrollments may be highly correlated; poor precision.
c. FE model may be the least attractive approach for estimating school effects; the elimination of bias may come at the expense of unacceptable imprecision.
d. Due to the common practice of mixing student class assignments from year to year, year 1 and 2 class enrollments are likely to be weakly correlated; better precision.
e. The strict exogeneity assumption could be violated if students are nonrandomly assigned to classrooms.

(2) **Random Effects (RE)**

Estimation strategy: hierarchical linear model (HLM), mixed model, or seemingly unrelated regression (SUR).

$$
\begin{aligned}
Y_{0i} &= \phi_i + E_{0i} + v_{0i} \\
Y_{1i} - Y_{0i} &= S_{1i}{}'\alpha_1 + \gamma_i + e_{1i} + v_{1i} - v_{0i} \\
Y_{2i} - Y_{1i} &= S_{2i}{}'\alpha_2 + \gamma_i + e_{2i} + v_{2i} - v_{1i}
\end{aligned} \tag{4}
$$

Sanders and Horn (1994)
Raudenbush and Bryk (2002)
Lockwood, McCaffrey, Mariano, and Setodji (2007).

Comments on RE Model:

a. Produces relatively precise school effect estimates.

b. If there is limited mobility across schools from year *1* to *2* then T3 random effect estimates will be very similar to T2 effect estimates and with similar precision. (SUR is identical to OLS if the regressor sets are identical in all equations.)

c. The weaknesses of the RE approach are:
   - The RE approach does not fully eliminate student selection/assignment bias, although this bias decreases as the length of the longitudinal panel increases (if there is sufficient mobility across schools).
   - The reported standard errors of school effect estimates do not include the error due to unknown selection bias.

d. The random effects approach may be attractive with respect to the tradeoff of bias and variance even if the individual growth effect γ_i is correlated with school and classroom enrollment variables.

e. A vector of student characteristics (say X_i) could also be added to the RE model.

(3) Conditional Random Effects (CRE)

Explicitly capture the correlation between growth effect γ_i and school (and classroom) enrollment variables via an auxiliary equation.

Growth selectivity with respect to school selection/assignment:

$$(\gamma_i \mid X_i, S_{1i}, S_{2i}) = X_i \pi + S_{1i}{}' g_1 + S_{2i}{}' g_2 + u_i \tag{9}$$

Initial (year 0) selectivity with respect to school selection/assignment:

$$(\phi_i \mid X_i, S_{1i}, S_{2i}) = X_i f_x + S_{1i}{}' f_1 + S_{2i}{}' f_2 + w_i \tag{10}$$

Conditional random effects (CRE) model: the individual effects γ_i and ϕ_i are random, *given* the conditioning variables X_i, S_{1i}, S_{2i}.

$$
\begin{aligned}
Y_{0i} &= X_i f_x + S_{1i}{}' f_1 + S_{2i}{}' f_2 + w_i + E_{0i} + v_{0i} \\
Y_{1i} - Y_{0i} &= X_i \pi + S_{1i}{}'(\alpha_1 + g_1) + S_{2i}{}' g_2 + u_i + e_{1i} + v_{1i} - v_{0i} \\
Y_{2i} - Y_{1i} &= X_i \pi + S_{1i}{}' g_1 + S_{2i}{}'(\alpha_2 + g_2) + u_i + e_{2i} + v_{2i} - v_{1i}
\end{aligned}
\qquad (11)
$$

$$
\begin{aligned}
\hat{\alpha}_1 &= (\alpha_1 + g_1) - \hat{g}_1 \\
\hat{\alpha}_2 &= (\alpha_2 + g_2) - \hat{g}_2
\end{aligned}
\qquad (12)
$$

Mundlak (1978)
Chamberlain (1984)
Meyer (2004)

Comments on CRE Model:

a. The strength of the CRE approach (as in the case of the FE approach) is that it eliminates student selection/assignment bias (given the assumption of strict exogeneity).
b. As in the case of the fixed effects model, year 1 and 2 school enrollments could be highly correlated at many grades, so the precision of the school effect estimates could be poor.
c. In contrast to the fixed effects model, the precision of effect estimates can be improved by imposing restrictions on the selection bias parameters. For example, it might be reasonable to impose this assumption that selection bias is equal in years *1* and *2*; that is, $g_1 = g_2 \equiv g$.

(4) Conditional Random Effects (CRE) with Multivariate Shrinkage (CRE-MS)

Idea: combine the conditional random effects approach with multivariate shrinkage to produce effect estimates that optimally trade off bias and variance in the sense of minimizing expected mean squared error (MSE).

Shrinkage

- Mixed models
- Best linear unbiased predictors (BLUP)
- James-Stein estimation
- Bayes and Empirical Bayes (EB) estimation
- Small area estimation
- Minimum mean squared error estimation (MSE)

Multivariate shrinkage: use information on all school-level effects $f_1, f_{2,} g_1, g_2, \alpha_1, \alpha_2$ to produce the shrinkage estimates.
Fuller and Harter (1987)

Definitions:

True parameter: $\theta_k' = [\, f_{1k} \quad f_{2k} \quad g_{1k} \quad g_{2k} \quad \alpha_{1k} \quad \alpha_{2k} \,]$

Unshrunken estimated parameter: $\hat{\theta}_k' = [\, \hat{f}_{1k} \quad \hat{f}_{2k} \quad \hat{g}_{1k} \quad \hat{g}_{2k} \quad \hat{\alpha}_{1k} \quad \hat{\alpha}_{2k} \,]$

Shrunken estimated parameter: $\theta_k^{*\prime} = [\, f_{1k}^* \quad f_{2k}^* \quad g_{1k}^* \quad g_{2k}^* \quad \alpha_{1k}^* \quad \alpha_{2k}^* \,]$

$\Omega =$ Variance-covariance matrix (over schools) of true effects θ_k.

$$\Omega = \begin{bmatrix} \omega_{f1f1} & \omega_{f1f2} & \omega_{f1g1} & \omega_{f1g2} & \omega_{f1\alpha1} & \omega_{f1\alpha2} \\ \omega_{f2f1} & \omega_{f2f2} & \omega_{f2g1} & \omega_{f2g2} & \omega_{f2\alpha1} & \omega_{f2\alpha2} \\ \omega_{g1f1} & \omega_{g1f2} & \omega_{g1g1} & \omega_{g1g2} & \omega_{g1f\alpha1} & \omega_{g1\alpha2} \\ \omega_{g2f1} & \omega_{g2f2} & \omega_{g2g1} & \omega_{g2g2} & \omega_{g2\alpha1} & \omega_{g2\alpha2} \\ \omega_{\alpha1f1} & \omega_{\alpha1f2} & \omega_{\alpha1g1} & \omega_{\alpha1g2} & \omega_{\alpha1\alpha1} & \omega_{\alpha1\alpha2} \\ \omega_{\alpha2f1} & \omega_{\alpha2f2} & \omega_{\alpha2g1} & \omega_{\alpha2g2} & \omega_{\alpha2\alpha1} & \omega_{\alpha2\alpha2} \end{bmatrix} \qquad (13)$$

Σ_{kk} = Variance-covariance matrix of estimation/sampling error of estimates of θ_k for school *k*.

$$\Sigma_{kk} = \begin{bmatrix} \sigma_{f1f1(kk)} & \sigma_{f1f2(kk)} & \sigma_{f1g1(kk)} & \sigma_{f1g2(kk)} & \sigma_{f1\alpha 1(kk)} & \sigma_{f1\alpha 2(kk)} \\ \sigma_{f2f1(kk)} & \sigma_{f2f2(kk)} & \sigma_{f2g1(kk)} & \sigma_{f2g2(kk)} & \sigma_{f2\alpha 1(kk)} & \sigma_{f2\alpha 2(kk)} \\ \sigma_{g1f1(kk)} & \sigma_{g1f2(kk)} & \sigma_{g1g1(kk)} & \sigma_{g1g2(kk)} & \sigma_{g1f\alpha 1(kk)} & \sigma_{g1\alpha 2(kk)} \\ \sigma_{g2f1(kk)} & \sigma_{g2f2(kk)} & \sigma_{g2g1(kk)} & \sigma_{g2g2(kk)} & \sigma_{g2\alpha 1(kk)} & \sigma_{g2\alpha 2(kk)} \\ \sigma_{\alpha 1f1(kk)} & \sigma_{\alpha 1f2(kk)} & \sigma_{\alpha 1g1(kk)} & \sigma_{\alpha 1g2(kk)} & \sigma_{\alpha 1\alpha 1(kk)} & \sigma_{\alpha 1\alpha 2(kk)} \\ \sigma_{\alpha 2f1(kk)} & \sigma_{\alpha 2f2(kk)} & \sigma_{\alpha 2g1(kk)} & \sigma_{\alpha 2g2(kk)} & \sigma_{\alpha 2\alpha 1(kk)} & \sigma_{\alpha 2\alpha 2(kk)} \end{bmatrix} \tag{14}$$

All value-added effects and other school-level effects are centered around zero; all estimates are shrunk toward zero.

Compute new estimates for each school:

$$\theta_k^* = w_k' \hat{\theta}_k$$

"Ensemble" estimation: Optimally incorporate (weight) information from the unshrunken estimated parameter vector, given knowledge of the multivariate distribution of the true parameter vector.

The matrix w_k that minimizes the expected mean squared error (EMSE) for each parameter is given by

$$w_k = [\Sigma_{kk} + \Omega]^{-1}\Omega. \tag{15}$$

$$EMSE = \Omega - \Omega'[\Sigma_{kk} + \Omega]^{-1}\Omega. \tag{16}$$

Special Cases

Univariate case: weight w_k equals the reliability of the effect estimate:

$$w_k = \frac{\omega}{\omega + \sigma_{kk}} \tag{17}$$

Bivariate case (for the single parameter θ_{1k}):

$$w_{1k} = \begin{bmatrix} w_{11k} \\ w_{12k} \end{bmatrix} = \begin{bmatrix} \dfrac{\omega_{11}(\omega_{22} + \sigma_{22kk}) - \omega_{12}(\omega_{12} + \sigma_{12kk})}{(\omega_{11} + \sigma_{11kk})(\omega_{22} + \sigma_{22kk}) - (\omega_{12} + \sigma_{12kk})^2} \\ \dfrac{\omega_{12}\sigma_{11kk} - \omega_{11}\sigma_{12kk}}{(\omega_{11} + \sigma_{11kk})(\omega_{22} + \sigma_{22kk}) - (\omega_{12} + \sigma_{12kk})^2} \end{bmatrix} \tag{18}$$

$\hat{\theta}_{2k}$ helps in the estimation of θ_{1k} if:

(a) $\omega_{12} > 0$ or
(b) $\sigma_{12kk} < 0$

$\hat{\theta}_{2k}$ acts as a noise suppressor for estimating θ_{1k} if the estimation errors for the two parameters are negatively correlated, even if the two parameters are uncorrelated in the population.

Comments on CRE-MS Model:

Consider the bivariate case: Use estimates of g_{1k} and $c_{11k} \equiv \alpha_{1k} + g_{1k}$ to construct the optimal estimator of α_{1k}.

$$\text{CRE: } \hat{\alpha}_{1k} = (\hat{c}_{11k} - \hat{g}_{1k}) \tag{19}$$

$$\begin{aligned} \text{CRE-MS } \alpha_{1k}^{*} &= w_{1k}\hat{\alpha}_{1k} + w_{2k}\hat{g}_{1k} \\ &= w_{1k}(\hat{c}_{11k} - \hat{g}_{1k}) + w_{2k}\hat{g}_{1k}\,. \\ &= w_{1k}\hat{c}_{11k} - (w_{1k} - w_{2k})\hat{g}_{1k} \end{aligned} \tag{20}$$

Subtracting the full amount of $\hat{g}_{1k}$ reduces bias, but it does so at the expense of inflating the noise level. The CRE-MS approach optimally trades off bias and variance.

Population Covariance Matrix Ω

The covariance matrix Ω is typically unknown and must be estimated. Defer discussion.

C. Applications

Scenario 1: Student mobility rate varies:

- 10%, (very low mobility)
- 30% (moderate mobility)
- 50% (high mobility)

Low mobility reduces the precision of separate estimates of α_t and g_t.

Scenario 2: Magnitude of student selectivity (as measured by the variances of g_{1k} and g_{2k}) varies.

Models:

- CRE
- CRE-MS1: One Cohort
- CRE-MS2: Two Cohorts

Table 1. The Effect of Student Mobility on Precision and Shrinkage Weights

	CRE	CRE-MS		CRE	CRE-MS		CRE	CRE-MS	
	One Cohort	One Cohort	Two Cohorts	One Cohort	One Cohort	Two Cohorts	One Cohort	One Cohort	Two Cohorts
Conditions									
Mobility Rate	0.5	0.5	0.5	0.3	0.3	0.3	0.1	0.1	0.1
Var(alpha1)=Var(alpha2)	9.79	9.79	9.79	9.79	9.79	9.79	9.79	9.79	9.79
Cov(alpha1,alpha2)	1.96	1.96	1.96	1.96	1.96	1.96	1.96	1.96	1.96
Var(g1)=Var(g2)	4.90	4.90	4.90	4.90	4.90	4.90	4.90	4.90	4.90
Cov(g1,g2)	4.41	4.41	4.41	4.41	4.41	4.41	4.41	4.41	4.41
Cov(alpha,g)	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
Weights for alpha2									
g1	-	-	-0.090	-	-	-0.052	-	-	-0.002
g2	-	-	0.037	-	-	0.026	-	-	0.025
alpha1	-	-	-0.053	-	-	-0.075	-	-	-0.108
alpha2	-	-	0.242	-	-	0.220	-	-	0.187
g1	-	0.056	0.082	-	0.141	0.146	-	0.237	0.222
g2	-	0.351	0.300	-	0.323	0.281	-	0.300	0.268
alpha1	-	-0.096	-0.045	-	-0.131	-0.067	-	-0.175	-0.100
alpha2	1	0.587	0.462	1	0.557	0.444	1	0.517	0.414
Standard Error									
g2	2.213	1.026	0.949	2.692	1.034	0.975	4.426	1.054	1.025
alpha2	3.129	1.841	1.615	3.807	1.905	1.713	6.259	1.989	1.858

Table 2. The Effect of Student Selectivity on Precision and Shrinkage Weights

	CRE	CRE-MS		CRE	CRE-MS		CRE	CRE-MS	
	One Cohort	One Cohort	Two Cohorts	One Cohort	One Cohort	Two Cohorts	One Cohort	One Cohort	Two Cohorts
Conditions									
Mobility Rate	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Var(alpha1)=Var(alpha2)	9.79	9.79	9.79	9.79	9.79	9.79	9.79	9.79	9.79
Cov(alpha1,alpha2)	1.96	1.96	1.96	1.96	1.96	1.96	1.96	1.96	1.96
Var(g1)=Var(g2)	4.90	4.90	4.90	0.98	0.98	0.98	0.00	0.00	0.00
Cov(g1,g2)	4.41	4.41	4.41	0.88	0.88	0.88	0.00	0.00	0.00
Cov(alpha,g)	2.08	2.08	2.08	0.93	0.93	0.93	0.00	0.00	0.00
Weights for alpha2									
g1	-	-	-0.090	-	-	-0.001	-	-	0.177
g2	-	-	0.037	-	-	0.136	-	-	0.317
alpha1	-	-	-0.053	-	-	-0.041	-	-	0.015
alpha2	-	-	0.242	-	-	0.254	-	-	0.309
g1	-	0.056	0.082	-	0.228	0.181	-	0.423	0.291
g2	-	0.351	0.300	-	0.548	0.418	-	0.749	0.533
alpha1	-	-0.096	-0.045	-	-0.042	-0.018	-	0.040	0.015
alpha2	1	0.587	0.462	1	0.645	0.492	1	0.728	0.526
Standard Error									
g2	2.213	1.026	0.949	2.213	0.678	0.643	2.213	0.000	0.000
alpha2	3.129	1.841	1.615	3.129	1.683	1.488	3.129	1.607	1.365

Conclusions

- The CRE-MS may be a valuable alternative or complement to the widely used random effects model.
- The CRE-MS method may produce important reductions in expected MSE relative to other estimation strategies.
- The CRE-MS method provides informative estimates of precision that take into account the error due to uncontrolled selection bias as well as traditional estimation error.
- As in all RE models, producing high-quality shrinkage estimates requires an accurate measure of the population covariance matrix Ω.