THE POWER TO KNOW.

SAS® EVAAS® Statistical Models

S. Paul Wright, John T. White, William L. Sanders, June C. Rivers

March 25, 2010

Table of Contents

1. Introduction and Summary **1**
2 Test Data Used in SAS EVAAS **2**
2.1 Dealing with Missing Data 2
2.2 Scaling Issues 3
2.3 Converting Test Scores to Normal Curve Equivalents 3
3 The EVAAS Multivariate Response Model (MRM) **4**
3.1 The Linear Mixed Model 4
3.2 The EVAAS District and School MRMs 6
3.3 The EVAAS Layered Teacher MRM 7
4 Projecting Student's Future Test Scores **10**
5 The EVAAS Univariate Response Model (URM) **12**
References **14**

1 Introduction

SAS® EVAAS® provides analytical services, including value-added modeling and projection analyses, for the assessment of schooling effectiveness at the district, school and, when requested, at the classroom level. The results of these analyses, along with additional diagnostic information and querying capabilities, are made available via a secure web application. This report provides details on the statistical models used in SAS EVAAS.

It is important to keep in mind that there is not one, single EVAAS model used in all applications. Rather, there are multiple models implemented according to the objectives of the analyses and the characteristics and availability of the test data. Two general types of value-added models are employed. The multivariate response model (MRM) is a multivariate, longitudinal, linear mixed model. In traditional statistical terminology it is essentially a multivariate repeated-measures ANOVA model. With this approach, the entire set of observed test scores belonging to each student is fitted simultaneously. When the data have been scaled or transformed to allow comparable expectations of progress, evaluated over many schools and/or districts, regardless of entering levels of groups of students, then the MRM approach is preferred. This model is discussed in section 3.

When the data structures do not meet the requirements for a MRM analysis, a univariate response model (URM) is employed. This model is similar to traditional analysis of covariance (ANCOVA): student scores in a particular subject/grade/year serve as the response variable (dependent variable); these students' prior scores in multiple subjects/grades/years serve as covariates (predictor variables, independent variables); the categorical variable (class variable, factor) is an educational entity (district, school, teacher). The URM differs from traditional ANCOVA in that the categorical variable is treated as random rather than fixed. In this respect, the URM has much in common with certain hierarchical linear models (HLMs) that have been used for value-added analyses. To minimize selection bias and to minimize problems caused by errors of measurement in the predictor variables, the URM requires that each student must have at least three prior scores. However, all available prior achievement test scores for each student are used in the predictor variable set. This model is discussed in section 5.

In addition to value-added modeling SAS EVAAS provides projected scores for individual students on tests the students have not yet taken. These tests may include state-mandated tests (end-of-grade tests, end-of-course tests where available) as well as national tests such as college entrance exams (SAT, ACT). These projections can be used to predict a student's future success (or lack of success) and so may be used to guide counseling and intervention to increase students' likelihood of future success. The statistical model used for making projections is, like the URM, an ANCOVA model. This model is discussed in section 4.

2 Test Data Used in SAS EVAAS

SAS EVAAS analyses make use of scores on standardized tests such as those provided by major educational testing companies and those used by states to fulfill their NCLB obligations. For tests to be used in EVAAS analyses, the scales must meet three criteria:

- They must be highly correlated with the state's curricular standards.
- They must adequately measure the performance of both very low and very high achieving students.
- They must be reliable measures.

These criteria are met by almost all commercial and state-mandated standardized tests.

2.1 Dealing with Missing Data

A common problem with using test scores is missing data. There are many reasons for this. For example, a student could move into the school district this year, a record could be lost, or a student could be sick on test day. The importance of dealing properly with missing data is illustrated in the following example.

Assume that ten students are given a test in two different years with the results shown the Table 1. The goal is to measure academic growth (gain) from one year to the next. The right side of the table shows what happens when some of the scores are missing. Two simple approaches to take when data are missing are to calculate the mean of the differences or to calculate the differences of the means. When there are no missing data, these two simple methods provide the same answer (5.80 on the left in Table 1); however, when there are missing data, each method provides a different result (9.57 vs. 3.97 on the right in Table 1). A more sophisticated model is needed to address this problem.

The EVAAS multivariate response model (MRM, see §3) uses the correlation between current and previous scores in the non-missing data to estimate a mean for the previous and current score as if there were no missing data. It does this without explicitly imputing values for the missing scores. The difference between these two estimated means is an estimate of the average gain for this group of students. In this small example, the estimated difference is 5.7. This method is closer to the no-missing-data mean of 5.8 than the 3.97 obtained by the mean of the differences and the 9.57 obtained by the difference of the means. This method of estimation has

Student	4th Grade	5th Grade	Gain
1	51.9	74.8	22.9
2	37.9	46.5	8.6
3	55.9	61.3	5.4
4	52.7	47.0	-5.7
5	53.6	50.4	-3.2
6	23.0	35.9	12.9
7	78.6	77.8	-0.8
8	61.2	64.7	3.5
9	47.3	40.6	-6.7
10	37.8	58.9	21.1
Mean	49.99	55.79	5.80
	Diff.	5.80	

Student	4th Grade	5th Grade	Gain
1	51.9		
2	37.9		
3	55.9	61.3	5.4
4	52.7	47.0	-5.7
5	53.6	50.4	-3.2
6	23.0	35.9	12.9
7		77.8	
8		64.7	
9	47.3	40.6	-6.7
10	37.8	58.9	21.1
Mean	45.01	54.58	3.97
	Diff.	9.57	

Table 1: *Left* - Scores without missing data; *Right* - Scores with missing data

been shown, on average, to outperform both of the simple methods (Wright, 2004). Larger data sets, such as those used in actual EVAAS analyses, provide better correlation estimates, which in turn provide better estimates of means and gains.

2.2 Scaling Issues

As in the small example above, the EVAAS MRM estimates academic growth as a "gain," the difference between two estimated means. For such a difference to be meaningful, the two means (that is, the two tests whose means are being estimated) must measure academic achievement on a common scale. Some test companies supply vertically scaled tests as a way to meet this requirement. An attractive alternative to relying on vertically scaled tests is to convert scale scores to normal curve equivalents (NCEs), a process sometimes referred to as "normalization." First, NCEs are on a familiar scale: They are scaled to look like percentiles, but they avoid the disadvantages of percentiles (the percentile scale is not an equal-interval scale). Second, the NCEs used in SAS EVAAS analyses are based on a reference distribution of test scores that is directly relevant to each client. In most cases, the reference distribution is the distribution of scores on a state-mandated test for all students in the state in a chosen year (the base year). By definition, the mean NCE score in the base year is 50 for each grade and subject. This sets the growth standard (the difference in NCEs from one year/grade to the next, which represents a "normal" year's growth) to a value of zero. Keep in mind that a gain of zero on the NCE scale does *not* indicate no growth. Rather, it indicates that a student (or group of students in a district, school or classroom) has maintained the same position in the state distribution from one grade to the next. In NCE analyses, this is what is meant by "a year's worth of growth."

Score	Freq	Cum Freq	Percent	Cum Pct	Ptile Rank	Z	NCE
670	58	248	5.8	24.8	21.9	-0.776	33.66
675	82	330	8.2	33.0	28.9	-0.556	38.29
687	96	426	9.6	42.6	37.8	-0.311	43.45
692	110	536	11.0	53.6	48.1	-0.048	48.99
695	105	641	10.5	64.1	58.9	0.225	54.74
702	88	729	8.8	72.9	68.5	0.482	60.15
705	76	805	7.6	80.5	76.7	0.729	65.35

Table 2: Converting tabulated test scores to NCE values

An alternative to normalization (using NCEs) that is sometimes suggested is "standardization." Standardized scores are scores that have been rescaled to have a mean of zero and a standard deviation of one. The usual practice is to standardize each year's scores separately. Alternatively, one could standardize to a base year (just as NCE scores are normalized to a base year). Just as for NCEs, a gain of zero represents "a year's worth of growth" *but only for the average student*. If the distributions of the two scores making up the gain have different shapes, then for students who are not average, a year's worth of growth may be represented by a gain of greater or less than zero. (If the two distributions are both normal, then standardization produces the same results as normalization.) In contrast, with NCEs a gain of zero represents a year's worth of growth for every student, no matter where in the distribution their score falls. It is for this reason, along with the familiarity of the NCE scale, that EVAAS analyses use NCE scores.

2.3 Converting Test Scores to Normal Curve Equivalents

The left side of Table 2 shows an excerpt from a tabulated distribution of test scores. Many states provide such statewide distributions for their state-mandated tests. The tabulation shows, for each possible test score (in a particular subject/grade/year), how many students made that score ("Freq") and what percent ("Percent") that frequency was out of the entire student population (in Table 2 the total number of students is assumed to be 1000 — obviously not a statewide distribution but convenient for this example). Also tabulated are the cumulative frequency ("Cum Freq," the number of students who made that score or lower) and its associated percentage ("Cum Pct").

The next step is to convert each score to a percentile rank. Percentile ranks have been added to the tabulation on the right side of Table 2 ("Ptile Rank"). If a particular score has a percentile rank of 48, this is interpreted to mean that 48% of students in the population had a lower score and 52% had a higher score. In practice, a non-zero percentage of students will receive each specific score; for example, 11% of students received a score

of 692 in Table 2. The usual convention is to consider half of that 11% to be "below" and half "above." Adding 5.5% (half of 11%) to the 42.6% who scored *below* the score of 692 produces the percentile rank of 48.1 in Table 2.

NCEs are obtained from the percentile ranks by assuming that student achievement is normally distributed. Using a table of the standard normal distribution (found in many textbooks) or computer software (e.g., a spreadsheet), one can obtain, for any given percentile rank, the associated Z-score from a standard normal distribution. NCEs are Z-scores that have been rescaled to have a "percentile-like" scale. Specifically, NCEs are scaled so that they exactly match the percentile ranks at 1, 50, and 99. This is accomplished by multiplying each Z-score by approximately 21.063 (the standard deviation on the NCE scale) and adding 50 (the mean on the NCE scale).

3 EVAAS Multivariate Response Model (MRM)

SAS EVAAS employs three separate MRM analyses, for districts, schools and teachers. The district and school models are essentially the same. They perform well with the large numbers of students that are characteristic of districts and most schools. The teacher model uses a different approach that is more appropriate with the smaller numbers of students typically found in teachers' classrooms. All three models are a special case of a statistical model known as the linear mixed model. This model and its properties are described below. Subsequent subsections describe the application of the linear mixed model in the EVAAS MRMs.

3.1 The Linear Mixed Model

The linear mixed model is represented by the following equation in matrix notation:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\upsilon} + \boldsymbol{\epsilon}. \tag{1}$$

$\mathbf{y}$ in the EVAAS context is the $m \times 1$ observation vector containing test scores (usually NCEs) for all students in all academic subjects tested over all grades and years (usually up to five years).

$\mathbf{X}$ is a known $m \times p$ matrix which allows the inclusion of any fixed effects.

$\boldsymbol{\beta}$ is an unknown $p \times 1$ vector of fixed effects to be estimated from the data.

$\mathbf{Z}$ is a known $m \times q$ matrix which allows for the inclusion of random effects.

$\boldsymbol{\upsilon}$ is a non-observable $q \times 1$ vector of random effects whose realized values

are to be estimated from the data.

$\boldsymbol{\epsilon}$ is a non-observable $m \times 1$ random vector variable representing unaccountable random variation.

Both $\boldsymbol{\upsilon}$ and $\boldsymbol{\epsilon}$ have means of zero, that is, $\mathrm{E}(\boldsymbol{\upsilon} = \mathbf{0})$ and $\mathrm{E}(\boldsymbol{\epsilon} = \mathbf{0})$. Their joint variance is given by

$$\mathrm{Var}\begin{bmatrix} \boldsymbol{\upsilon} \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \tag{2}$$

where $\mathbf{R}$ is the $m \times m$ matrix that reflects the correlation among the student scores residual to the specific model being fitted to the data, and $\mathbf{G}$ is the $q \times q$ variance-covariance matrix that reflects the correlation among the random effects. If $(\boldsymbol{\upsilon}, \boldsymbol{\epsilon})$ are normally distributed, the joint density of $(\mathbf{y}, \boldsymbol{\upsilon})$ is maximized when $\boldsymbol{\beta}$ has value $\mathbf{b}$ and $\boldsymbol{\upsilon}$ has value $\mathbf{u}$ given by the solution to the following equations, known as Henderson's mixed model equations (Sanders et al., 1997):

$$\begin{bmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}. \tag{3}$$

Let a generalized inverse of the above coefficient matrix be denoted by

$$\begin{bmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} = \mathbf{C}. \tag{4}$$

If $\mathbf{G}$ and $\mathbf{R}$ are known, then some of the properties of a solution for these equations are:

1. $\mathbf{K}^T\mathbf{b}$ is best linear unbiased estimator (BLUE) of the set of estimable linear functions, $\mathbf{K}^T\boldsymbol{\beta}$, of the fixed effects.
2. $\mathbf{u}$ is the best linear unbiased predictor (BLUP) of $\boldsymbol{\upsilon}$.
   - $\mathrm{E}(\boldsymbol{\upsilon} \mid \mathbf{u}) = \mathbf{u}$
   - $\mathrm{Var}(\mathbf{u} - \boldsymbol{\upsilon}) = \mathbf{C}_{22}$
   - $\mathbf{u}$ is unique regardless of the rank of the coefficient matrix.
3.
   - $\mathbf{K}^T\mathbf{b} + \mathbf{M}^T\mathbf{u}$ is BLUP of $\mathbf{K}^T\boldsymbol{\beta} + \mathbf{M}^T\boldsymbol{\upsilon}$ provided $\mathbf{K}^T\boldsymbol{\beta}$ is estimable.
   - $\mathrm{Var}(\mathbf{K}^T(\mathbf{b} - \boldsymbol{\beta}) + \mathbf{M}^T(\mathbf{u} - \boldsymbol{\upsilon}) = (\mathbf{K}^T\,\mathbf{M}^T)\mathbf{C}(\mathbf{K}^T\,\mathbf{M}^T)^T$.
4. With $\mathbf{G}$ and $\mathbf{R}$ known, the solution is equivalent to generalized least squares, and if $\boldsymbol{\upsilon}$ and $\boldsymbol{\epsilon}$ are multivariate normal, then the solution is the maximum likelihood solution.

5. If $\mathbf{G}$ and $\mathbf{R}$ are not known, then as the estimated $\mathbf{G}$ and $\mathbf{R}$ approach the true $\mathbf{G}$ and $\mathbf{R}$, the solution approaches the maximum likelihood solution.

6. If $\boldsymbol{v}$ and $\boldsymbol{\epsilon}$ are not multivariate normal, then the solution to the mixed model equations still provides the maximum correlation between $\boldsymbol{v}$ and $\mathbf{u}$.

3.2 The EVAAS District and School MRMs

The EVAAS district and school MRMs do not contain random effects; consequently, in the linear mixed model, the $\mathbf{Z}\boldsymbol{v}$ term drops out. The $\mathbf{X}$ matrix is an incidence matrix (a matrix containing only zeros and ones) with a column representing each school (in the school model), subject, grade and year of data. The fixed-effects vector $\boldsymbol{\beta}$ contains the mean score for each school, subject, grade and year, with each element of $\boldsymbol{\beta}$ corresponding to a column of $\mathbf{X}$. Note that, since EVAAS models are generally run separately for each district, the district itself need not be represented in the model.

Unlike the case of the usual linear model used for regression and analysis of variance, the elements of $\boldsymbol{\epsilon}$ are *not* independent. Their interdependence is captured by the $\mathbf{R}$ matrix. Specifically, scores belonging to the same student are correlated. If the scores in $\mathbf{y}$ are ordered so that scores belonging to the same student are adjacent to one another, then the $\mathbf{R}$ matrix is block diagonal with a block, $\mathbf{R}_i$, for each student. Each student's $\mathbf{R}_i$ is a subset of the "generic" covariance matrix $\mathbf{R}_0$ that contains a row and column for each subject and grade. Covariances among subjects and grades are assumed to be the same for all years (technically, all cohorts), but otherwise the $\mathbf{R}_0$ matrix is unstructured. Each student's $\mathbf{R}_i$ contains only those rows and columns from $\mathbf{R}_0$ that match the subjects and grades for which the student has test scores. In this way, the EVAAS MRM is able to use all available scores from each student. Because so few students repeat grades, it is not possible to estimate the elements of a student's $\mathbf{R}_i$ representing the correlation between repeated scores in the same subject and grade (but in different years). These correlations are, in effect, set to zero. Computationally this is done by treating students-within-cohorts as the independent elements of the $\mathbf{y}$ vector. That is, a student's scores within a cohort are correlated, but when a student changes cohorts (e.g., by repeating a grade) the scores after the change are independent of the scores before the change.

Algebraically, the district MRM is

$$y_{ijkl} = \mu_{jkl} + \epsilon_{ijkl}. \tag{5}$$

y_{ijkl} represents the test score for the ith student in the jth subject in the kth grade during the lth year. μ_{jkl} is the estimated district mean score for

this particular subject, grade, and year. ϵ_{ijkl} is the random deviation of the ith student's score from the district mean.

The EVAAS school MRM is

$$y_{ijkls} = \mu_{jkls} + \epsilon_{ijkls}. \tag{6}$$

This is the same as the district model with the addition of the subscript s representing s^{th} school.

Solving the mixed model equations for the EVAAS district or school MRM produces a vector $\mathbf{b}$ that contains the estimated mean score for each school (in the school model), subject, grade and year. To obtain a value-added measure of average student growth, the estimated means are converted to gains by subtracting the estimated mean score for the previous year and grade from the estimated mean score for the current year and grade. In the school model, this is done for each school. Because students may change schools from one year to the next (in particular when transitioning from elementary to middle school, for example), the estimated mean score for the prior year/grade is a composite of the means from the feeder schools attended by the students in the prior year/grade. Feeder schools which contribute too few students (fewer than 5) to be considered a reliable sample from the feeder school are omitted from the composite.

In addition to reporting the estimated mean scores and mean gains produced by these models, EVAAS reports (1) cumulative gains across grades (for each subject and year), (2) three-year-average gains (for each subject and grade), and optionally (3) composite gains across subjects.

3.3 The EVAAS Layered Teacher MRM

In this model, to allow for the possibility of many teachers with relatively few students per teacher, teachers are entered as random effects via the $\mathbf{Z}$ matrix in the linear mixed model. The $\mathbf{X}$ matrix contains a column for each subject/grade/year, and the $\mathbf{b}$ vector contains an estimated district mean score for each subject/grade/year. The $\mathbf{Z}$ matrix contains a column for each subject/grade/year/teacher, and the $\mathbf{u}$ vector contains an estimated teacher effect for each subject/grade/year/teacher. To allow for comparisons among teachers at different schools within an entire district, school is not represented in the model. The $\mathbf{R}$ matrix is as described above for the district/school MRM. The $\mathbf{G}$ matrix contains teacher variance components, with a separate variance component for each subject/grade/year. To allow for the possibility that a teacher may be very effective in one subject and very ineffective in another, the $\mathbf{G}$ matrix is constrained to be a diagonal matrix. Consequently, the $\mathbf{G}$ matrix is a block diagonal matrix with a block for each subject/grade/year. Each block has the form $\sigma^2_{jkl}\mathbf{I}$ where σ^2_{jkl} is the teacher

variance component for the jth subject in the kth grade in the lth year, and $\mathbf{I}$ is an identity matrix.

Algebraically, the teacher model is

$$y_{ijkl} = \mu_{jkl} + (\sum_{k^* \leq k} \sum_{t=1}^{T_{ijk^*l^*}} w_{ijk^*l^*t} \times \tau_{ijk^*l^*t}) + \epsilon_{ijkl}. \tag{7}$$

y_{ijkl} is the test score for the ith student in the jth subject in the kth grade in the lth year. $\tau_{ijk^*l^*t}$ is the teacher effect of the tth teacher on the ith student in the jth subject in grade k^* in year l^*. The complexity of the parenthesized term containing the teacher effects is due to two factors. First, in any given subject/grade/year a student may have more than one teacher. The inner (rightmost) summation is over all the teachers of the ith student in a particular subject/grade/year. $T_{ijk^*l^*}$ is the number of those teachers. $w_{ijk^*l^*t}$ is the fraction of the ith student's instructional time claimed by the tth teacher. Second, the EVAAS teacher MRM allows teacher effects to accumulate over time. That is, how well a student does in the current subject/grade/year depends not only on the current teacher but also on the accumulated knowledge and skills acquired under previous teachers. The outer (leftmost) summation accumulates teacher effects not only for the current subject/grade/year (subscripts k and l) but also over previous grades and years (subscripts k^* and l^*) in the same subject. Because of this accumulation of teacher effects, the EVAAS teacher MRM is often called the "layered" model.

Table 3 illustrates how the $\mathbf{Z}$ matrix is encoded for three students who have three different scenarios of teachers during grades 3, 4 and 5 in three subjects, math (M), reading (R) and language arts (L). Teachers are identified by the letters A–F.

Tommy's teachers represent the conventional scenario: Tommy is taught by a single teacher in all three subjects each year (teachers A, C and E in grades 3, 4 and 5, respectively). Notice that in Tommy's $\mathbf{Z}$ matrix rows for grade 3, there are ones (representing the presence of a teacher effect) not only for third grade teacher C but also for second grade teacher A. This is how the "layering" in encoded. Similarly, in the the grade 4 rows, there are ones for grade 4 teacher E, grade 3 teacher C, and grade 2 teacher A.

Susan is taught by two different teachers in grade 2, teacher A for math and language arts, teacher B for reading. In grade 3, Susan had teacher C for reading and language arts. For some reason, in grade 3 no teacher claimed Susan for math even though Susan had a 3rd grade math test score. This score can still be included in the analysis by entering zeros into the Susan's $\mathbf{Z}$ matrix rows for 3rd grade math. In grade 4, on the other hand, Susan had no test score in language arts. This row is completely omitted from the $\mathbf{Z}$ matrix. There will always be a $\mathbf{Z}$ matrix row corresponding to each test score in the $\mathbf{y}$ vector. Since Susan has no entry in $\mathbf{y}$ for

			Teachers																	
			2nd Grade						3rd Grade						4th Grade					
			A			B			C			D			E			F		
Grade	Subject	Student	M	R	L	M	R	L	M	R	L	M	R	L	M	R	L	M	R	L
2	M	Tommy	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	R		0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	L		0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	M	Tommy	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	R		0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	L		0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
4	M	Tommy	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
	R		0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0
	L		0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
2	M	Susan	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	R		0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	L		0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	M	Susan	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	R		0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
	L		0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
4	M	Susan	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
	R		0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0
2	M	Eric	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	R		0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	L		0	0	.5	0	0	.5	0	0	0	0	0	0	0	0	0	0	0	0
3	M	Eric	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
	R		0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
	L		0	0	.5	0	0	.5	0	0	0	0	0	1	0	0	0	0	0	0
4	M	Eric	1	0	0	0	0	0	0	0	0	1	0	0	.8	0	0	.2	0	0
	R		0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0
	L		0	0	.5	0	0	.5	0	0	0	0	0	1	0	0	.5	0	0	.5

Table 3: Encoding the **Z** matrix

4th grade language arts, there can be no corresponding row in $\mathbf{Z}$.

Eric's scenario illustrates team teaching. In grade 2 language arts, Eric received an equal amount of instruction from both teachers A and B. The entries in the $\mathbf{Z}$ matrix indicate each teacher's contribution, 0.5 for each teacher. Similarly, in 4th grade language arts, Eric was taught equally by teacher E and F. In 4th grade math, however, while Eric was taught by both teachers E and F, they did not make an equal contribution. Teacher E claimed 80% responsibility and teacher F claimed 20%.

Recall that estimates from the MRM district and school models were estimated means scores. Value-added effectiveness measures were obtained by taking differences between estimated mean scores to obtain mean gains. In contrast, the EVAAS teacher MRM produces teacher "effects" (in the $\mathbf{u}$ vector of the linear mixed model). It also produces, in the fixed-effects vector $\mathbf{b}$, district-level mean scores (for each subject/grade/year). Because of the way the $\mathbf{X}$ and $\mathbf{Z}$ matrices are encoded, in particular because of the "layering" in $\mathbf{Z}$, teacher gains can be estimated by adding the teacher effect to the district mean gain. That is, the interpretation of a teacher effect in the EVAAS teacher MRM is as a gain expressed as a deviation from the average gain for the district as a whole.

Because teacher effects are treated as random effects in the EVAAS teacher

MRM, their estimates are obtained by shrinkage estimation, technically known as best linear unbiased prediction or as empirical Bayesian estimation. This means that *a priori* a teacher is considered to be "average" (with a teacher effect of zero) until there is sufficient student data to indicate otherwise. This method of estimation protects against false positives (teachers incorrectly evaluated as effective) and false negatives (teachers incorrectly evaluated as ineffective), particularly in the case of teachers with few students.

4 Projecting Student's Future Test Scores

In addition to providing value-added modeling, SAS EVAAS provides a variety of additional services including supplemental diagnostic reports and querying capabilities. These additional services also include projected scores for individual students on tests the students have not yet taken. These tests may include state-mandated tests (end-of-grade tests, end-of-course tests where available) as well as national tests such as college entrance exams (SAT, ACT). These projections can be used to predict a student's future success (or lack of success) and so may be used to guide counseling and intervention to increase students' likelihood of future success. In some states (Tennessee, Ohio, Pennsylvania) these projections have also been approved for use in the Growth Model Pilot Program of No Child Left Behind.

The statistical model that is used as the basis for the projections is, in traditional terminology, an analysis of covariance (ANCOVA) model. In this model, the score to be projected serves as the response variable (y, the dependent variable), the covariates (x's, predictor variables, explanatory variables, independent variables) are scores on tests the student has already taken, and the categorical variable (class variable, factor) is the school at which the student received instruction in the subject/grade/year of the response variable (y). Algebraically, the model can be represented as follows for the ith student.

$$y_i = \mu_y + \alpha_j + \beta_1(x_{i1} - \mu_1) + \beta_2(x_{i2} - \mu_2) + \ldots + \epsilon_i. \qquad (8)$$

The μ terms are means for the response and the predictor variables. α_j is the school effect for the jth school, the school attended by the ith student. The β terms are regression coefficients. Projections to the future are made by using this equation with estimates for the unknown parameters (μ's, β's, sometimes α_j). The parameter estimates (denoted with "hats," e.g., $\hat{\mu}$, $\hat{\beta}$) are obtained using the most current data for which response values are available. The resulting projection equation for the ith student is

$$\hat{y}_i = \hat{\mu}_y + \hat{\beta}_1(x_{i1} - \hat{\mu}_1) + \hat{\beta}_2(x_{i2} - \hat{\mu}_2) + \ldots \pm \hat{\alpha}_j. \qquad (9)$$

The reason for the '$\pm$' before the $\hat{\alpha}_j$ term is that, since the projection is to a future time, the school that the student will attend is unknown, so this term is usually omitted from the projections. This is equivalent to setting $\hat{\alpha}_j$ to zero, that is, to assuming the student encounters the "average schooling experience" in the future. In some instances, a customer may be willing to provide a list of feeder patterns from which it is possible to determine the most likely school that a student will attend at some projected future date. In this case, the $\hat{\alpha}_j$ term can be included in the projection.

Two difficulties must be addressed in order to implement the projections. First, not all students will have the same set of predictor variables due to missing test scores. Second, because of the school effect in the model, the regression coefficients must be "pooled-within-school" regression coefficients.

The strategy for dealing with missing predictors is to estimate the joint covariance matrix (call it $\mathbf{C}$) of the response and the predictors. Let $\mathbf{C}$ be partitioned into response (y) and predictor (x) partitions, that is,

$$\mathbf{C} = \begin{bmatrix} c_{yy} & \mathbf{c}_{yx} \\ \mathbf{c}_{xy} & \mathbf{C}_{xx} \end{bmatrix}. \tag{10}$$

This matrix is estimated using the EM algorithm for estimating covariance matrices in the presence of missing data provided by the MI procedure in SAS/STAT®. Only students who had a test score for the response variable in the most recent year and who had at least three predictor variables are included in the estimation. Given such a matrix, the estimated regression coefficients for the projection equation can be obtained as

$$\mathbf{b} = \mathbf{C}_{xx}^{-1}\mathbf{c}_{xy}. \tag{11}$$

This allows one to use whichever predictors a particular student has to get that student's projected y-value ($\hat{y}_i$). Specifically, the $\mathbf{C}_{xx}$ matrix used to obtain the regression coefficients *for a particular student* is that subset of the overall $\mathbf{C}$ matrix that corresponds to the set of predictors for which this student has scores.

Because of the school effect in the projection model, $\mathbf{b}$ contains pooled-within-school regression coefficients. Therefore, the $\mathbf{C}$ matrix must be a pooled-within-school covariance matrix. This is obtained by first subtracting school mean scores from each student's response score (y) and each predictor score (the x's). Note that the school whose mean is subtracted from y and from each x is the school at which the student obtained the response score, i.e., it is the school represented by the α_j term in the projection model. It is these mean-adjusted scores that go into the estimation of the $\mathbf{C}$ matrix. Only students at schools having at least 10 students are included.

The projection equation also requires estimated mean scores for the response and for each predictor (the $\hat{\mu}$ terms in the projection equation).

These are not simply the grand mean scores. It can be shown that in an ANCOVA, if one imposes the restriction that the estimated school effects should sum to zero (that is, the school effect for the "average school" is zero), then the appropriate means are the means of the school means. These are obtained by finding the average of y and each x at each school then finding the simple average of these averages over all schools.

In any case, because of missing predictor values, the estimated means for the predictors are likely to be biased. The EM algorithm used to obtain the $\mathbf{C}$ matrix also estimates a mean for each variable. Since mean-adjusted scores go into the EM algorithm, these means would all be zero if there were no missing data. However, because of missing predictor values, the estimated means will be non-zero, and these non-zero values reflect the amount of bias due to missing data. Consequently, the estimated means used to make the projections are obtained by adding the means from the EM algorithm to the previously calculated means of the school means.

Once the parameter estimates for the projection equation have been obtained, projections can be made for any student with any set of predictor values. However, to protect against bias due to measurement error in the predictors, projections are made only for students who have at least three available predictor scores. In addition to the projected score itself, the standard error of the projection is calculated. Given a projected score and its standard error, it is possible to calculate the probability that a student will reach some specified benchmark of interest. Examples are the probability of scoring at the proficient (or advanced) level on a future end-of-grade test, or the probability of scoring sufficiently well on a college entrance exam to gain admittance into a desired program. The probability is calculated as the area above the benchmark cutoff score using a normal distribution with its mean equal to the projected score and its standard deviation equal to the standard error of the projected score.

5 EVAAS Univariate Response Model (URM)

As presently implemented, the EVAAS MRM reports value-added effectiveness in terms of gains. This has two implications. First, the scores must be scaled so that taking the difference between two (mean) scores to obtain a (mean) gain is a meaningful mathematical operation. This is most often accomplished by converting scores to NCEs. Second, there must be an obvious "before" and "after" from which to form a difference. This works well with end-of-grade tests given in elementary and (usually) middle school where "before" is the previous grade and previous year. It does not work with high school (and sometimes middle school) end-of-course tests. To estimate value-added effectiveness in situations where the MRM is not feasible, SAS EVAAS has implemented a univariate response model (URM).

The URM statistical model is identical to the EVAAS projection model: it is an analysis of covariance (ANCOVA) model. This model is described in section 4. One difference from the projection model is that, since the URM may be used at the district, school or teacher level, the categorical variable in the ANCOVA may be the district, school or teacher. A second difference is that, in the teacher URM, a student may have more than one teacher. Finally, unlike the projection model, the primary objective is not to produce an equation for making projections, but to estimate the "effect" itself, the α_j term in the equation.

As was the case in the projection model, different students will have different sets of predictors. In the URM this is handled exactly as in the projection model. In fact, the first step in the URM is to obtain a "projection" for each student using whatever set of predictors that student has available. However, rather than projecting a future score using a student's present and past scores, in the URM one is "projecting" a student's present score using their past scores. Also, in the URM, unlike the projection model, the estimated parameters are pooled-within-district, pooled-within-school, or pooled-within teacher, as required. As a reminder, here is the projection equation. For reasons given below, the "projection" for the ith student is denoted C_i.

$$C_i = \hat{\mu}_y + \hat{\beta}_1(x_{i1} - \hat{\mu}_1) + \hat{\beta}_2(x_{i2} - \hat{\mu}_2) + \ldots. \tag{12}$$

The "projection" C_i is nothing more than a *composite* of all the student's past scores (thus the letter C). It is a one-number summary of the student's level of achievement prior to the current year. The different prior test scores making up this composite are given different weights (by the regression coefficients, the $\hat{\beta}$'s) in order to maximize its correlation with the response variable. Thus a different composite would be used when the response variable is math than when it is reading, for example. Note that the $\hat{\alpha}_j$ term is *not included* in the equation. Again, this is because C_i represents prior achievement, before the effect of the current district/school/teacher. As in the projection model, to avoid bias due to measurement error in the predictors, composites are obtained only for students who have at least three prior test scores.

The second step in the URM is to estimate district/school/teacher effects (α_j) using the following ANCOVA model.

$$y_i = \beta_0 + \beta_1 C_i + \alpha_j + \epsilon_i. \tag{13}$$

The effects (α_j) are considered to be random effects. Consequently the $\hat{\alpha}_j$'s are obtained by shrinkage estimation (empirical Bayes) as described in §3.3.

As noted above, in the teacher-level URM a student may have more than one teacher. This is encoded in the $\mathbf{Z}$ matrix of the linear mixed model as described in §3.3. The difference is that in the URM, the response vector

$(\mathbf{y})$ contains scores for only a single subject/grade/year; that is why it is called the *univariate* response model. Consequently, the columns of the $\mathbf{Z}$ matrix represent districts or schools or teachers only for the current subject/grade/year. There is no "layering." Instead, prior schooling is captured by the composite score C_i that appears as a predictor (or covariate) in the ANCOVA model.

References

Sanders, W. L., Saxton, A. M., and Horn, S. P. (1997). "The Tennessee Value-Added Assessment System: A Quantitative, Outcomes-Based Approach to Educational Assessment." In Millman, J. (ed.), *Grading Teachers, Grading Schools*, Thousand Oaks, CA: Sage Publications, pp. 137-162.

Wright, S. P. (2004). "Advantages of a Multivariate Longitudinal Approach to Educational Value-Added Assessment Without Imputation." Paper presented at National Evaluation Institute, on-line at `http://www.createconference.org/documents/archive/2004/Wright-NEI04.pdf`.

Wright, S. P., Sanders, W. L., and Rivers, J. C. (2006). "Measurement of Academic Growth of Individual Students toward Variable and Meaningful Academic Standards." In Lissitz, R. W. (ed.), *Longitudinal and Value Added Models of Student Performance*, Maple Grove, MN: JAM Press, pp. 385-406.

SAS Institute Inc. World Headquarters +1 919 677 8000
To contact your local SAS office, please visit: **www.sas.com/offices**

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.
Copyright © 2010, SAS Institute Inc. All rights reserved. S56935.0310