Missouri Growth Model Technical Documentation

Aug 7, 2013

1 Introduction

This document describes the estimation procedure employed by the Missouri Growth Model to generate growth measures for Local Education Agencies (LEAs) and schools. These measures are reported on the MSIP 5 APR and reflect systematic differences in academic achievement gains compared to baseline predictions.

It is important to note that these measures are just one gauge of effectiveness. They are not designed to be a measure of progress toward the state's 2020 performance targets, for example. Instead, they indicate how achievement gains among similarly circumstanced students in similarly circumstanced LEAs or schools differ as a function of the *particular* LEAs or schools where students were enrolled when they took the MAP exams. In this way, estimates generated by the Missouri Growth Model are relative.

2 Data

The Missouri Growth Model is estimated using individual student test results from the Missouri Assessment Program (MAP) exams given annually to public school students in the state of Missouri. Currently, the Missouri Growth Model uses data from the mathematics and English language arts exams administered to all students in grades three through eight.

At the current time, a three-year rolling panel is used as the analytic data sample. For example, following the 2012 academic year, exam scores from 2012, 2011, and 2010 were included as outcome variables in the model estimation. The use of multiple years of data improves the stability of the growth estimates. Of course, the tradeoff in including multiple years of data in the model estimation is that real improvements in school and LEA quality take longer to appear in the effect estimates. The three-year panel strikes a balance between the goal of improving the stability of effect estimates and the desire to help LEAs and schools demonstrate improvements more quickly.

2.1 Standardizing MAP Scale Scores

Growth measures in MSIP 5 are designed to provide estimates of schooling effectiveness for units (LEAs or schools) as a whole. It is therefore important that the measures have a meaningful interpretation at the unit-level. Moreover, the generalized predictive relationship between a student's exam score in a given year and his or her prior-year exam score cannot be estimated appropriately in cases when apparent gains may be confounded by differences in scaling from one grade to the next. Due to these considerations, MAP scale scores are standardized by year and grade prior to being submitted to the model.

Standardization is accomplished by converting MAP scale scores to z-scores. Z-score standardization is commonly performed on data that exist on different scales. A z-score of zero represents the mean for a given subject, year, and grade. The following example explains how a z-score is calculated:

STEP	EXPLANATION
1. Find the mean scale score for the given assessment. Each combination of grade level, content area, and school year is treated as a different assessment in this context.	The mean ($\bar{x}$) is the sum of the scale scores for all students with a valid score, divided by the number of students with a valid score (N).
2. Find the standard deviation of the scale score for the same assessment.	The formula for standard deviation is $s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$, where x_i is the scale score for a given student.
3. Take the student's scale score and subtract the mean. Then divide by the standard deviation. The result is the z-score.	If the mean is 640; the standard deviation is 38; and the student's actual scale score is 700; then: z =(700-640) / 38 z =60/38 z = 1.5789

Table 1: Calculation of z-scores

2.2 Method of Pairing Scores

The model uses test score pairs for estimation. A score pair is formed by matching an exam score for a student tested in year *t* (the outcome score) to a prior exam score for the same student in the same subject and previous grade from year *t-1* (a predictor score.) As a result, scores from fourth grade students are the first scores that can appear as outcome scores in the model. Scores from students who take the exam twice at the same grade level, due to being retained in grade, do not generate a valid score pair for the year the retest occurred.

The example below shows how an individual student's exam scores are arrayed as pairs:

Year *t*	**Grade Level in Year *t***	**Standardized MAP Scale Score for Year *t* *(Outcome Scores)***	**Grade Level in Year *t-1***	**Standardized MAP Scale Score for Year *t-1 (Predictor Scores)***
2012	8	1.30	7	1.10
2011	7	1.10	6	0.80
2010	6	0.80	5	0.60
2009	5	0.60		

Table 2: Arrangement of Data as Score Pairs

2.3 Treatment of Missing Data

A prior-year same-subject exam score (predictor score) is required for an outcome score to be included in model estimation. Specifically, if a student is missing the mathematics MAP score in year *t-1* when the outcome score in the model is the mathematics MAP score in year *t*, then that student's score is dropped from the analysis. The same rules are used to construct the English language arts estimation sample, i.e., both the year *t* and year *t-1* English language arts scores must be available to include the student's score pair in the analysis. This method was chosen because the absence of a lagged same-subject score can be seen as conceptually problematic in a gains model.

The model also uses prior year exam scores from the "other subject" to predict current year scores. For example, when a mathematics MAP score is the outcome score, a prior year English language arts score for the same student from the previous grade also is used as a predictor score. In cases where the lagged off-subject score is unavailable, the lagged off-subject score is set to zero, the standardized mean. This maximizes the amount of data included in the estimation and accounts for students with poor attendance during the week of examinations (a group that is likely to be non-random).

This data strategy sets a student's missing, lagged off-subject score equal to the statewide exam average. However, students with missing exam scores may systematically over or underperform relative to students that truly scored at the statewide average on the previous year off-subject exam (and for whom these data are available). To control for this possibility, an indicator variable signifying the presence of a missing score is also included in the model. Moreover, the model includes an interaction term to give more weight to the same-subject lagged MAP score for the observations where the lagged off-subject MAP score is missing, as it is now the sole source of empirical information about prior test performance. The full model estimation strategy is discussed in the next section.

3 Model Specification

3.1 First-Stage Predictive Model

The estimation procedure used to measure growth consists of two steps. In the first step, individual students' MAP scores, standardized by year, subject, and grade, are regressed on student and unit-level characteristics. The following equation is estimated using Ordinary Least Squares (OLS).

$$Y_{ijt}(x) = \beta_0 + \beta_1 Y_{ij(t-1)}(m) + \beta_2 Y_{ij(t-1)}(ela) + \beta_3 Missing + \beta_4 Missing \times Y_{ij(t-1)}(x) + \beta_5 M_{ijt} + \beta_6 S_{ijt} + \beta_7 Grade + \beta_8 Year + e_{ijt} \quad (1)$$

where

$Y_{ijt}(x) =$ A test score in subject x (m=math or ela=English language arts) for student i at unit j in year t.

The unit component is flexibly defined and can be applied at the LEA level, school level, etc. This flexibility is one of the benefits of the model. Models are currently being estimated at the LEA and school levels only.

$Missing$ = A binary indicator variable where the indicator is set to one if the lagged off-subject MAP score is missing and is set to zero otherwise.

$Missing \times Y_{ij(t-1)}(x)$ = An interaction term between the *Missing* indicator variable and the lagged same-subject MAP score.

M_{ijt} = A binary indicator variable set to one if the student was in the building where tested for less than the full academic school year and zero otherwise.

S_{ijt} = A vector of variables controlling for unit-specific characteristics.

The unit characteristics are also calculated from the MAP score records and measure average lagged mathematics and English language arts MAP scores, the percentage of students with missing lagged off-subject MAP scores (e.g., the percent missing lagged English language arts scores in the mathematics model), and the percentage of tested students that were in the building in which they were tested for less than a full school year. Note that the average lagged exam scores are based on the prior scores of students who took the MAP test at the unit in year *t*, and not on the year *t-1* scores of students that were actually in the unit at that time (although there may be substantial overlap between the two sets).

Grade = A set of binary indicator variables where the indicator is set to one if the student is in the relevant grade when the exam was taken, while all others are set to zero.

Year = A set of binary indicator variables where the indicator for the year when the test was taken is set to one, while all others are set to zero.

These two sets of indicator variables account for differences in the testing data that are observed across grades and over time and that are correlated to current-year MAP scores.

e_{ijt} = The OLS error term from the regression.

The model presented in equation (1) is then estimated using statewide exam score data. The OLS parameter estimates (regression coefficients) for this first-stage regression are available from http://dese.mo.gov/mogrowthmodel/documents/output.pdf. These estimates define the independent linear relationship between the predictor variables presented above and the outcome exam scores. Given these relationships, the model can then be used to predict each student's outcome scores given the values of his or her predictor variables. For example, consider a student with the data record for one year presented in Table 3.

Variable	**Value**
Current-Year Math Score (z-score units)	0.226
Prior-Year Math Score (z-score units)	0.127
Prior-Year English Language Arts Score (z-score units)	0.675
Missing Off-Subject (ELA) Prior-Year Score Indicator	0
Mobility Indicator	1
LEA Average Lagged Math Score	0.213
LEA Average Lagged English Language Arts Score	0.011
LEA Percent Mobile	5.12
LEA Percent of Students with Missing Off-Subject Scores	3.86
Grade 4 Indicator	0
Grade 5 Indicator	1
Grade 6 Indicator	0
Grade 7 Indicator	0
2010 School Year Indicator	0
2011 School Year Indicator	0

Table 3: Student Exam Score Prediction Sample Data

This record describes a grade-5 student who took the MAP mathematics exam in 2012 (the grade 5 indicator is set to 1, while the 2010 and 2011 school year indicators are set to 0. Note that this student also could have a data record included in the model estimation where the 4th grade MAP score is the outcome score and the 3rd grade scores are predictors). Moreover, the student was not present in the school in which the exam was taken for the entire year (the mobility indicator is set to one) but did take the MAP exam in an LEA with above average lagged exam scores and a low overall percentage of mobile students. The student also has lagged exam scores available in both subjects (note that the missing off-subject prior-year exam indicator is set to 0). Given these values and the coefficients for the 2012 LEA math model, the following calculation is used to determine the student's predicted 2012 exam score:

$$\hat{Y}_{ijt}(m) = \hat{\beta}_0 + \hat{\beta}_1 Y_{ij(t-1)}(m) + \hat{\beta}_2 Y_{ij(t-1)}(ela) + \hat{\beta}_3 Missing + \hat{\beta}_4 Missing \times Y_{ij(t-1)}(x) + \hat{\beta}_5 M_{ijt} + \hat{\beta}_6 S_{ijt} + \hat{\beta}_7 Grade + \hat{\beta}_8 Year \quad (2)$$

$$= 0.014 + (0.625)0.127 + (0.220)0.675 + (-0.077)0 + (0.043)(0 \times 0.127) + (-0.114)1 + (0.222)0.213 + (-0.068)0.011 + (0.004)5.12 + (-0.002)3.86 + (0.003)0 + (0.002)1 + (0.001)0 + (0.000)0 + (-0.000)0 + (0.001)0$$

= 0.189.

Hence, this student would be predicted to score 0.189 standard deviations above the mean on the 2012 MAP mathematics grade-5 exam.

Once the predicted scores are calculated, they are subtracted from the observed scores to generate residuals, which reflect the unexplained growth in student scores. For the above student, this value is $\hat{e}_{ijt} = 0.226 - 0.189 = 0.037$. In other words, the student scored higher than predicted by the model and would figure positively into the LEA effect estimate.

3.2 Extensions of the First-Stage Predictive Model

For the 2013 academic year, students taking the state's Algebra 1 end-of-course assessment were exempted from having to take the grade-level Mathematics assessment. In order to incorporate performance for these students, two additional models are fit to accommodate Algebra 1 scores for 7th and 8th graders who were not administered the corresponding grade level Mathematics exams.[1] These models are of the following form:

$$Y_{ijt}(x) = \beta_0 + \beta_1 Y_{ij(t-1)}(m) + \beta_2 Y_{ij(t-1)}(ela) + \beta_3 Missing + \beta_4 Missing \times Y_{ij(t-1)}(x) + \beta_5 M_{ijt} + \beta_6 S_{ijt} + e_{ijt} \quad (3)$$

All variables use the same definitions as those provided in Section 3.1.

Algebra 1 scores are standardized by year and grade so that a z-score of 0 represents the mean Algebra 1 score of either 7th graders (in the case of the 7th grade model) or 8th graders (in the case of the 8th grade model). This approach facilitates an interpretation of the student-level residuals as being relative to the performance of other students who likewise took the Algebra 1 assessment early.[2]

3.3 Second-Stage Effect Model

Once the residuals from the first-stage regressions ($\hat{e}_{ijt}$) are calculated and captured for each student, they are used as the dependent variable in a second-stage regression:

$$\hat{e}_{ijt} = \theta \cdot (Unit\ Indicator\ Variables) + u_{ijt} \quad (4)$$

The residuals, $\hat{e}_{ijt}$, are the part of outcome test scores not predicted from students' prior year scores and unit characteristics. The second-stage regression then captures how much of the variation in the residuals can be explained by the units under study, be it LEAs or schools. (For purposes of exposition, the assumption is that the units are schools throughout the rest of the model description.) Thinking of the model in terms of the baseline prediction in stage 1, and noting that the dependent variable in the second stage is the student-level deviation from the baseline prediction, the second-stage regression can be used to identify schools where the students systematically perform above or below their predicted values.[3] Equation (4) is estimated twice to produce two separate sets of school effect

---

[1] Students administered both the Algebra 1 and grade-level assessments in 2013 are considered duplicates. Their 2013 scores will not be used in the model estimation.

[2] To ensure that student performance on the EOC and MAP exams within grades has an equal impact on district ratings, the distribution of EOC residuals is adjusted by a scaling factor, within each grade, so that it has the same variance as the distribution of MAP residuals.

[3] Also note that the second-stage regression is estimated without an intercept. This is beneficial, as it allows an effect and, more importantly, a corresponding standard error to be estimated for every school under consideration.

estimates – one calculated using all student residuals associated with each school ($\hat{\theta}_{j,uncentered}^{s1}$) and one calculated using only the student residuals from super-subgroup students ($\hat{\theta}_{j,uncentered}^{s2}$). In both cases, the standard errors for the second-stage regression are calculated to be robust in the presence of heteroskedasticity and are clustered at the student-level to account for the fact that a single student can appear up to three times in the data, once for each of his/her exam score pairs included in the model. This effectively lowers the number of independent observations used in the estimation procedure.

Once the effects of all schools are estimated, they are centered appropriately. For MSIP 5, Standard 1, this is accomplished by calculating the average effect for all schools and then subtracting that average from each school effect. Specifically,

$$\hat{\theta}_j^{s1} = \hat{\theta}_{j,uncentered}^{s1} - \bar{\bar{\theta}}_{uncentered}^{s1} \tag{5}$$

where $\bar{\bar{\theta}}_{uncentered}^{s1}$ is the average of the uncentered effects for all schools in the state. As a result of this centering, the mean value for $\hat{\theta}_j^{s1}$ will be zero. For MSIP 5, Standard 2, the comparison group is the average residual for all non-super-subgroup students in the state. Hence, the centered effect estimate in this case is given by:

$$\hat{\theta}_j^{s2} = \hat{\theta}_{j,uncentered}^{s2} - \bar{\hat{e}}_{non-ssg}. \tag{6}$$[4]

3.4 Shrinkage and Conversion to NCE Units

After the estimates are centered, shrinkage techniques are then applied to them to help account for the fact that individual school effects are measured with differing amounts of noise.[5] This variation in the reliability of estimates is the result of a variety of factors including sample size differences across schools and variability in exam score measurement error across students. The shrinkage estimate for each school is a weighted average of that school's centered effect estimate, $\hat{\theta}_j$, and the overall average school effect, $\bar{\bar{\theta}}$. Schools with noisy estimates have relatively more weight placed on the overall average, while schools with less noisy estimates have relatively more weight placed on the effect estimate. The weight applied to the estimate for each school *j* is given by the following formula.[6]

$$r_j = \frac{\hat{\sigma}_\theta^2}{\hat{\sigma}_\theta^2 + \hat{\sigma}_j^2} \tag{7}$$

In (7), $\hat{\sigma}_\theta^2$ is an estimate the overall variance of the school effects (minus estimation error) and is calculated as the variance of the estimated school effects, $\hat{\sigma}_{\hat{\theta}}^2$, minus the adjusted mean of the

[4] In the calculation the centered effect estimates, it is assumed that the mean value for the reference group is equal to the true population value, so that the standard errors for the uncentered estimates are equal to the standard errors for the centered effect estimates.

[5] All of the procedures described in this section are performed separately on the estimates for MSIP 5 Standard 1 and MSIP 5 Standard 2. To simplify exposition, the superscripts on the effect estimates are suppressed and a single, general effect estimate ($\hat{\theta}$) is presented for illustration.

[6] This school-specific weight, r_j, is known as the reliability ratio, and it is used to calculate the shrunken effect estimate in the following manner: $\hat{\theta}_{j,shrunken} = r_j\hat{\theta}_j + (1 - r_j)\bar{\bar{\theta}}$.

estimated variance of each individual school's effect estimate, $\hat{\sigma}_j^2$, where $\hat{\sigma}_j^2$ is the square of each effect estimate's standard error.[7]

The shrunken effect estimates and the corresponding upper and lower bounds of their 95% confidence intervals are converted to normal curve equivalent (NCE) units via the following formula

$$\hat{\theta}_{j,shrunken,NCE} = 50 + 21.06\,\hat{\theta}_j. \quad (8)$$

Additionally, the shrunken effects can be tested for statistical significance using the shrunken standard errors associated with the effect estimate for each school.[8] For both MSIP 5 Standard 1 and 2, the test statistic is calculated via the following formula:

$$t_j = \frac{\hat{\theta}_{j,shrunken} - 0}{\hat{\sigma}_{j,shrunken}} = \frac{\hat{\theta}_{j,shrunken}}{\hat{\sigma}_{j,shrunken}}. \quad (9)$$

In both cases, the null hypothesis compares the shrunken effect estimate to zero. However, it is important to remember that the comparison group (the zero) differs by standard. For Standard 1, this value is simply the average statewide school effect. For Standard 2, the centering is in comparison to the average residual for all non-super-subgroup students in the state.

Given the high number of student observations in each model (nearly one million in the Standard 1 specification) and the convergence property of the t-distribution, these test statistics are then compared to the standard normal distribution to determine statistical significance.[9] For Standard 1, significant positive effects indicate the school performed above the state average in a statistically distinguishable way, while significant negative effects indicate the school performed below the state average. School effects that are not statistically significant cannot be differentiated from the mean with available data. For Standard 2, significant positive effects indicate that the super-subgroup students in the school, on average, outperformed the non-super-subgroup students in the state in a statistically distinguishable way; conversely, significant negative effects indicate that the opposite is true. Insignificant effects indicate that the test score growth of super-subgroup students in the school cannot be statistically differentiated from the statewide test score growth of non-super-subgroup students.

---

[7] Specifically, the adjusted mean is calculated as $\frac{1}{n-1}\sum_{j=1}^{n}\hat{\sigma}_j^2$. This procedure is based on Aaronson et al. (2007), who use the same calculation to estimate the estimation-error variance of teacher fixed effects in their study.

[8] The shrunken standard errors are simply the unshrunken standard errors multiplied by the reliability ratio, i.e. $\hat{\sigma}_{j,shrunken} = r_j\hat{\sigma}_j$.

[9] All statistical tests are conducted at the 0.05 significance level.

References

Aaronson, Daniel, Lisa Barrow and William Sander. 2007. Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics* 25(1), 95-135.

Harris, Douglas N., *Value-Added Measures in Education: What Every Educator Needs to Know*, Harvard Education Press, 2011.

McCaffrey, Daniel F., et al., *Evaluating Value-Added Models for Teacher Accountability*, RAND Corporation (MG-158), 2006.

McCaffrey, Daniel F., J.R. Lockwood and Ann Hass. "What Variables Belong in Your Value-Added Model? "Presented at the Annual Meeting of the American Educational Research Association. New Orleans, LA. 2011.