January 30, 2020

SCHOOL EFFECTS IN CHAIN-LINKED PREDICTIVE ANALYTICS MODELS

Research Paper, Education Analytics

ROBERT H. MEYER

We investigate issues involved in adding school effects to chain-linked predictive analytics models and systems.

Topics:

Introduction

The predictive analytics models considered in this memo included school effects for each outcome in each grade-specific equation (although some school effects, such as in the educational model, encompass school effects for multiple grades). We are interested in school effects for at least several reasons. First, schools may differentially contribute to student outcomes, a prime motivation for implementing value-added models. It is arguably important to include school effects in predictions of student outcomes if these effects vary across the different school attended or predicted to be attended by students. Second, it is important to identify which outcomes have the largest school effects, typically captured by the noise-corrected variance of effects. This information can be used to prioritize further investigation into the determinants of positive and negative school effects. Third, as discussed in a later section, information on the magnitude of school effects for each outcome could be helpful in identifying variables that are not measured on the same scale across schools. In other words, large school effects may be evidence that schools measure and report variables on different scales rather than evidence of differences in true school effectiveness. This issue could arise, for example, if some schools grade on the A-to-C scale and some schools grade on the A-to-F scale. If there are substantial differences in grading practices across schools and districts, predictions for students in different schools, if based off the same calibration/estimation sample, would differ due to non-comparability of data rather than as a genuine reflection of differences in statistically expected outcomes.

An additional objective of the analysis is to compare the approach here of including school effects in all predictive analytic models versus the approach of excluding school effects from these models but analyzing chain-linked predictions of student-level educational attainment (such as high school graduation) at the school level. We have referred to the latter approach as quasi-value added. One advantage of the approach considered in this memo is that it deconstructs overall school effects with respect to educational attainment into the separate (but possibly correlated) effects of the variables used to predict educational attainment, thus identifying the variables that are the strongest predictors of attainment. Including school effects in predictive analytic systems has the additional advantage of producing more accurate estimates of the uncertainty associated with predictions (prediction and forecast error).

We focus in this memo on three periods and models: (1) predicting educational attainment outcomes such as high graduation or AG-certified graduation given 9th grade outcomes and (2) predicting the set of 9th grade outcomes (the predictors in the attainment model) given all 8th grade outcomes and predictors and (3) predicting the set of 8th grade outcomes (the predictors in the models of 9th grade outcomes).[1]

Some cautions are in order. One, the models considered in this memo are predictive analytics models that are designed to maximize predictive accuracy and thus are not necessarily appropriate for estimating (approximately) causal school effects, the objective of value-added growth models. Nonetheless, the models take the same form as value-added models. Two, the educational attainment model considered in this memo is a long-span model that jumps directly from 9th grade outcomes to the educational attainment outcome measured three years later and thus does not address outcomes that occur during grades 10 through 12, including school mobility.

---

[1] This analysis will shed light on the quasi-value-added method used at EA, although in our earlier work we chain-linked two different models: educational attainment given 10th grade outcomes and 10th grade outcomes given 8th grade outcomes.

Long-span educational attainment model

The attainment model is defined by two equations, an equation to predict educational attainment and school enrollment in the grades beyond *9* (S_{Ai}). We define education attainment as a series of binary measures: high school graduation, AG-certified graduation, enrollment in college, and college persistence. These binary outcomes can be viewed as different levels of educational attainment measured as years of education. Let A_i represent the latent variable in a probit model for one of the binary attainment outcomes. There is a separate attainment equation for each of these outcomes. To simplify the predictive analytics system presented in this memo, we specify all other models as linear models. The attainment model is given by:

$$A_i = W_{9i}\theta_A + S_{Ai}\alpha_A + e_{Ai} \qquad (1)$$

$$S_{Ai} = W_{9i}\pi_A + S_{9i}\tau_A + u_{Ai} \qquad (2)$$

where:

i indexes students. The variables are currently not subscripted by a year/time index *t*, since the models are estimated (calibrated) using data from a given cohort of students. Time subscripts can be added as needed; for example, to address the stability of school effects across cohorts.

A_i = the latent variable in a probit model of a binary educational attainment outcome.

W_{9pi} = student variable *p* ($p = 1,...,P_9$) measured at the end of 9th grade. $W_{9i} = [W_{9pi}]$ = row vector of P_9 9th grade variables (with dimension P_9). These variables can vary over time and across schools. In general, in a predictive analytics model with chain linking the predictors in a model for a given grade become outcome variables in outcome models for a prior grade. Predictors include course level variables and grade point outcomes, attendance rates, and enrollments in programs such as ELL, special education, or gifted and talented, and economic disadvantage status. See Appendix A for a more comprehensive list of variables, grouped by type.

S_{Ai}, S_{9i} = row vectors of binary (zero/one) school enrollment indicators for the schools students attended grades 10-12 (with dimension K_A) and grade 9 (with dimension K_9), respectively. School indicators may or may not be included as predictors in the models, although in this memo we generally focus on the case where school indicators are included. If not included, the school effects are absorbed into the error term.

θ_{Ap}, π_{Ap} = the coefficients on the predictor W_{9pi} in the attainment and school enrollment equations for variable *p*, respectively. $\theta_A = [\theta_{Ap}]$ represents the column vector of coefficients (with dimension P_9) and $\pi_A = [\pi_{Ap}]$ represents the matrix of coefficients (with dimension $K_A x P_9$).

α_A = the column vectors of school effects in the attainment model after controlling for 9th grade predictors (with dimension K_A).
τ_A = the matrix of coefficients in predicting S_{Ai} given S_{9i}, controlling for 9th grade predictors (with dimension $K_A x K_9$).
e_{Ai}, u_{Ai} = the student-level residuals, after controlling for 9th grade predictors and school effects in the attainment and school enrollment equations (with dimensions *1* and $1xK_A$, respectively).

Note that the above model (and all of the models presented in this memo) adopt the assumptions of a recursive model and thus do not include school indicators from grades prior to grade 10 or predictors as control variables prior to grade 9. In other words, we assume that it is adequate to include in the model single lags of variables. The assumption with respect to school indicators is not relevant if the sample is limited to students enrolled in the same school in all high school grades. If the assumptions of a recursive model are valid, we can obtain valid predictions of future variables by chain-linking multiple models given the assumption that the predictive relationships between variables stays the same. We refer to this assumption as the "steady state" or "no regime change" assumption.

An important feature of this is that it includes predictors measured at the end of 9th grade (W_{9i}), but school enrollment indicators recorded after 9th grade (S_{Ai}). This model specification thus shares the same form as a value-added model, although the focus of a value-added model is to measure school effects (here, represented by α_A), whereas in a predictive analytics model we are interested in the predictive power obtained primarily from the student-level predictors. Nonetheless, predictions of educational attainment could include a part due to student-level predictors (S_{Ai}) and a part due to predictions of future school effects (α_A). Inclusion of the school component is more complex than the student-based component because the including the school component requires predictions of schools attended by students in future years and predictions of future school effects, which generally are not perfectly correlated over time. The need to predict future school enrollment arises if predictions are provided at the end of a school year (rather than in the fall) or if predictions are required for future school years. For the special case where data is presented to students in the fall the "future" school is known and thus need not be predicted using equation (2). In all other cases school enrollment must be predicted using equation (2).

As indicated in equation (2), the school enrollment model consists of a system of binary enrollment indicators S_{Ai} with two sets of predictors: student outcomes W_{9i} and prior grade school enrollment indicators S_{9i}. If the predictive power of the student outcomes is minimal, given the inclusion of the prior enrollment indicators (or if the outcome variables are simply dropped from the model), the coefficients on the prior enrollment variables τ_A have a very simple interpretation: they capture the probabilities of students moving from one school to another school from grades *9* to *10.* The set of probabilities define a school-to-school transition matrix of dimension $K_A x K_9$. All school enrollment models considered in this memo take this form: they embed a transition matrix that documents patterns of school mobility. We note that the school enrollment models are especially important for students enrolled in schools that have terminal grades: for example, elementary schools that offer grades K-5 and middle schools that offer grades 6-8. We include student outcomes as predictors in the school enrollment models partly because they may have predictive power, but also so that the student predictors in the attainment (and in lower grades, student outcome) models are identical.

The coefficients in the predictive analytics models play a special role in chain linking: we use these coefficients as scoring weights in making predictions of education attainment A_i given prior grade variables. We use the term scoring weights to emphasize the point that the coefficients obtained from the attainment model indicate the predictor power of each variable (suitably standardized to permit comparisons). Large coefficients indicate that a given variable is a strong predictor of educational attainment. As discussed above, they do not necessarily indicate the importance of a variable with respect to causality.

Although the attainment model parameters are defined in terms of a latent variable equation, the model is generally estimated as a binary probit model, where *1* = an attainment outcome such as high school graduation (within a given time period, such as four years) and *0* = non-graduation. The probability of the attainment outcome is given by:

$$P_{1i} = \Phi\left[W_{9i}\theta_A + S_{Ai}\alpha_A\right]$$

if we include school effects in the model and predictions and where the standard normalization, $\text{Var}(e_{Ai}) = \sigma_{eA}^2 = 1$, is imposed. Note that in the model that includes school effects, if treated as fixed effects, it is necessary to drop all student observations from the model for a given school if the proportions of students in either binary outcome are zero or one. In a fixed effects model, coefficients are estimated using only the variability of outcomes and predictors within schools. There is no variance in outcomes if all students are in one of the binary outcomes. One alternative estimation strategy is to treat the school effects as random rather than fixed.

Despite the fact that the attainment model is designed to predict a binary outcome, it is actually simpler when evaluating the magnitude of the parameters and in chain linking to use the scale of the latent variable A_i, rather than probability scale.

Short-span models of outcomes in 9th grade given 8th grade predictors and outcomes in 8th grade given 7th grade predictors

A primary difference between the attainment model and the outcome models considered in this section is that the attainment model consists of one equation whereas the latter models constitute a system of equations with residuals that are, in general, correlated.[2] The outcomes models considered in this section are short-span models in that they include predictors from the immediate prior grade. The issues addressed below would equally apply to long-span models, although they would be complicated by the fact that some students change schools over time, especially at transition grades.

For simplicity, we use linear regression models for all of the predictors in the educational attainment model, including the outcome variables and school enrollment variables.[3] As in the attainment model we adopt the assumptions of a recursive model and thus include outcome variables and school indicators only from prior grades.

Models for the vectors of 9th grade student outcomes and school enrollment variables are given by:

$$W_{9i} = W_{8i}\theta_9 + S_{9i}\alpha_9 + e_{9i} \quad (3)$$

$$S_{9i} = W_{8i}\pi_9 + S_{8i}\tau_9 + u_{9i} \quad (4)$$

The similar set of models for 8th grade variables are given by:

$$W_{8i} = W_{7i}\theta_8 + S_{8i}\alpha_8 + e_{8i} \quad (5)$$

$$S_{8i} = W_{7i}\pi_8 + S_{7i}\tau_8 + u_{8i} \quad (6)$$

where:

W_{9i}, W_{8i}, W_{7i} = the row vectors of student variables measured at the end of 9th grade, 8th grade, and 7th grade, respectively (with dimensions P_9, P_8, and P_7, respectively). At each grade level, the predictors are identical in all models. As above, predictors include course level variables and grade point outcomes, attendance rates, possibly enrollments in programs such as ELL,

---

[2] Note that we define education attainment as a series of binary measures: high school graduation, AG-certified graduation, enrollment in college, and college persistence. These binary outcomes can be viewed as different levels of educational attainment measured as years of education. There is a separate attainment equation for each of the attainment outcomes.

[3] It is possible and, in fact, preferable, to use binary and ordered probit models to model the discrete outcomes. We simplify the presentation of the model by using linear models for all outcome and enrollment variables. Linear models may perform reasonably well in a predictive analytics system, but it is possible to generate predictions of binary variables that lie outside of the range of zero and one, which obviously cannot be interpreted as proper probabilities.

special education, or gifted and talented, economic disadvantage status, and test scores in math and ELA. Note that test scores are available in 7th and 8th grade, but not in 9th grade.
S_{9i}, S_{8i}, S_{7i} = row vectors of binary (zero/one) school enrollment indicators for the schools attended in grades 9, 8, and 7, respectively (with dimensions K_9, K_8, K_7).
$\theta_9, \theta_8, \alpha_9, \alpha_8, \pi_9, \pi_8, \tau_9, \tau_8$ = conformable parameter matrices in the four models. Later in the memo we address the issue that the school effect parameters α_{9c}, α_{8c} are likely not perfectly stable over time and thus need to be forecasted. We add cohort *c* subscripts to these parameters to allow for the fact that they may vary over time and thus across cohorts. The same treatment could be applied to the school effect parameters in the school enrollment equations τ_9, τ_8, but it may be reasonable to assume that these parameters are relatively constant over time. We adopt this assumption in this memo.
$e_{9i}, e_{8i}, u_{9i}, u_{8i}$ = the student-level residuals in the four models.

The models defined above represent systems of equations since the variables to be predicted are vectors of student outcomes and school enrollment variables. Given the assumptions of a recursive model, the equations at different grade levels are not connected, but the student outcome equations within each grade level are connected through the correlation of the student-level residuals e_{gi} across the P_g equations (where *g* indicates grade). The same is true for the school enrollment residuals u_{gi}. However, the student outcome residuals and school enrollment residuals are not correlated within the same grade or across grades, which allows these systems to be estimated separately. Let the variance-covariance matrices of the student-level residuals and school enrollment residuals in grade *g* be equal to Σ_{eegg} and Σ_{uugg}, respectively. These error variance matrices are used below to compute the variance of prediction errors.

Models with restrictions imposed

Many, if not most, schools serve students in all high school grades, all middle school grades, all elementary school grades, or all elementary and middle school grades. Thus, it may be reasonable when computing predictions to impose the restrictions that students are enrolled in the same school over all the grades provided by the schools. In this case, the school enrollment models are needed only for transition grades. These restrictions are imposed as follows:

$$\begin{aligned} S_{A/9i} &= S_{Ai} = S_{9i} \\ S_{8/7i} &= S_{8i} = S_{7i} \end{aligned} \tag{7}$$

To further simplify models, we impose the restrictions that student outcomes W_{9i}, W_{8i}, W_{7i} do not determine school enrollments; that is, $\tau_A = \tau_9 = \tau_8 = 0$. As discussed above, with this restriction imposed the school enrollment models take the form of transition matrices. The models that result given these restrictions are given by:

$$A_i = W_{9i}\theta_A + S_{A/9i}\alpha_A + e_{Ai} \quad (1a)$$

$$W_{9i} = W_{8i}\theta_9 + S_{A/9i}\alpha_9 + e_{9i} \quad (3a)$$

$$S_{A/9i} = W_{8i}\pi_9 + S_{8/7i}\tau_9 + u_{A/9i} \quad (4a)$$

$$W_{8i} = W_{7i}\theta_8 + S_{8/7i}\alpha_8 + e_{8i} \quad (5a)$$

where the letter “a” has been added to each equation to indicate that the equation imposes the stated restrictions.

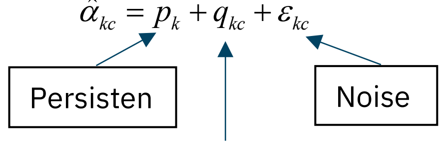
A model for forecasting school effects

In order to produce predictions of future outcomes, including educational attainment, one final set of models needed to be added to the system; that is, models to forecast future school effects for future cohorts. These models capture possible trends and volatility of school effects over time. In general, models of this type need to allow for a persistent (but possibly variable) component, a random transitory component, and the error in estimating the school effect. The persistent component of could take one of the following forms:

1. Static model: a persistent fixed component
2. Dynamic-linear time trend model: a persistent component subject to change
3. Dynamic-serial correlation model: a persistent component subject to change
4. Dynamic-persistent fixed component and serially correlated transitory component

In this memo we focus on the first form since it can be estimated with a relatively small number of years of data (say, 2-3 years). The static model, which can also be described as a variance components model or a stability model, takes the same form for all student outcomes. However, the static model is based on a single equation only for the educational attainment model. For all other grades, the number of student outcomes P_g is greater than one and thus the static model consists of a system of P_g equations. We first address the model based on a single outcome. We then consider the more general case of multiple outcomes. The static model for a single outcome is given by:

$$\hat{\alpha}_{kc} = p_k + q_{kc} + \varepsilon_{kc} \quad (8)$$



Transitory

where *k* indexes schools, c indexes cohorts, and the parameters are defined in the text boxes. Given this model, the optimal forecast of a future school effect $\tilde{\alpha}_{kc}$ for future cohort *c* is given by the shrinkage estimate:

$$\tilde{\alpha}_{kc} = \rho_k \hat{p}_k \qquad (9)$$

where $\hat{p}_k$ = an estimate of the persistent school effects and $\rho_k = \frac{\omega^2}{\omega^2 + \sigma_k^2}$ = the reliability of the estimated persistent school effect, constructed using the following terms: ω^2 = true (noise-corrected) variance of the persistent school effect and σ_k^2 = the variance of error in estimating the persistent school effect for school *k*.[4] The error of the forecast is given by:

$$\tilde{\varepsilon}_{kc} = \alpha_{kc} - \tilde{\alpha}_{kc} \qquad (10)$$

and equivalently, the true school effect is given by:

$$\alpha_{kc} = \tilde{\alpha}_{kc} + \tilde{\varepsilon}_{kc} \qquad (11)$$

Note that these two formulas take the same form for the case of multiple student outcomes – see below. The variance of the forecast error is equal to:

$$\mathrm{Var}(\tilde{\varepsilon}_{kc}) = \tilde{\sigma}_k^2 = \rho_k \sigma_k^2 \qquad (12)$$

Given the assumptions of the static persistence model, the variance of the forecast error is the same for all future cohorts and thus the variance is not subscripted by the cohort index. (See References for information on shrinkage estimation.)

The method for forecasting school effects readily extends to multiple student outcomes.[5] The optimal multivariate shrinkage estimator of the row vector α_{kc} (with dimension P_g) is given by:

$$\tilde{\alpha}_{kc} = \hat{p}_k \left[\Omega + \Sigma_{kk}\right]^{-1} \Omega \qquad (13)$$

where

[4] Note that there is a difference between computing shrinkage estimates for future cohorts versus cohorts with observed estimated school effects. Shrinkage estimates for the latter group can include both estimates of the persistent and transitory components.

[5] Later in the memo we address challenges that arise in estimating school effects for variables with low variance; for example, indicator variables of whether a student enrolled in a course in a specific subject.

$$W_k = \left[\Omega + \Sigma_{kk}\right]^{-1} \Omega \quad (14)$$

is the $(P_g x P_g)$ multivariate shrinkage matrix, Ω = true (noise-corrected) $(P_g x P_g)$ variance-covariance matrix of parameter vector (the matrix analog of the scalar parameter ω^2), Σ_{kk} = $(P_g x P_g)$ variance-covariance of error in estimating α_{kc} (the matrix analog of σ_k^2), and $\hat{p}_k$ = the estimate of the persistent school effects vector.[6] The associated variance-covariance of the forecast error of the parameter vector is given (with four equivalent formulas) by:

$$\tilde{\Sigma}_{kk} = W_k \Sigma_{kk} = (I - W_k)\Omega = \Omega\left(\Omega + \Sigma_{kk}\right)^{-1} \Sigma_{kk} = \Sigma_{kk}\left(\Omega + \Sigma_{kk}\right)^{-1} \Omega \quad (15)$$

The school-specific shrinkage estimates and forecast errors for all schools can be combined into single vectors $\tilde{\alpha}_{gc}$ and $\tilde{\varepsilon}_{gc}$, where we now add the grade subscript *g*.

Chain-linking predictive analytic models to produce long-span predictions

The predictive analytics models presented above are the building blocks for producing predictions of educational attainment over grade spans that exceed the number of years of longitudinal data observed for available student cohorts. These predictions are obtained by chain-linking models that sequentially predict outcomes and school enrollment that are, in turn, used to predict variables in subsequent grades. If the assumptions of a recursive model are valid and it is correct to assume that the predictive relationships between variables stay the same, the method of chain-linking produces valid predictions of future student outcomes. We refer to the latter assumption as the “steady state” or “no regime change” assumption.

As indicated earlier, predictions of educational attainment can be expressed in two different ways if educational attainment is modeled using a binary probit model: (1) compute predictions on the latent variable scale represented by A_i, using equation (1) or (2) compute predictions in terms of the predicted probability of the binary educational attainment outcome. To simplify the ensuing analysis, we focus on the first option. The predictive analytics system presented above supports computation of educational attainment predictions beginning in grade 7. Table 1 summarizes the data and model requirements need to compute predictions using the unrestricted models defined by equations (1) to (6). The models are identified by the letters A to F in the table. Table 2 summarizes the same requirements for the restricted models defined by equations (1a), (3a), (4a), and (5a). The restricted models effectively

[6] Note that the shrinkage formula in equation (14) is designed to shrink row vectors. The formula for shrinking a column vector is the transpose of that formula.

combine models A and B and models E and F. The restricted models are thus identified by the letters A/B, C, D, and E/F.

Table 1. Predictions of Educational Attainment Given Different Points in Time using the Unrestricted Models

Time of Data Availability	Known Data	Variables Predicted and Equation #	School Effects Forecasted for Cohort *c*
Chain-Linking Not Required			
A. Fall 10th grade	W_9, S_A	(1) A	$\tilde{\alpha}_{Ac}$
Chain-Linking Required			
B. Summer after 9th grade	W_9, S_9	(1) A (2) S_A	$\tilde{\alpha}_{Ac}$
C. Fall 9th grade	W_8, S_9	(1) A (2) S_A (3) W_9	$\tilde{\alpha}_{Ac}$ $\tilde{\alpha}_{9c}$
D. Summer after 8th grade	W_8, S_8	(1) A (2) S_A (3) W_9 (4) S_9	$\tilde{\alpha}_{Ac}$ $\tilde{\alpha}_{9c}$
E. Fall 8th grade	W_7, S_8	(1) A (2) S_A (3) W_9 (4) S_9 (5) W_8	$\tilde{\alpha}_{Ac}$ $\tilde{\alpha}_{9c}$ $\tilde{\alpha}_{8c}$
F. Summer after 7th grade	W_7, S_7	(1) A (2) S_A (3) W_9 (4) S_9 (5) W_8 (6) S_8	$\tilde{\alpha}_{Ac}$ $\tilde{\alpha}_{9c}$ $\tilde{\alpha}_{8c}$

Table 2. Predictions of Educational Attainment Given Different Points in Time using the Restricted Models

Time of Data Availability	Known Data	Variables Predicted and Equation #	School Effects Forecasted for Cohort *c*
Chain-Linking Not Required			
A/B. Summer after 9th grade or Fall 10th grade	W_9, $S_{A/9}$	(1a) A	$\tilde{\alpha}_{Ac}$
Chain-Linking Required			
C. Fall 9th grade	W_8, $S_{A/9}$	(1a) A (3a) W_9	$\tilde{\alpha}_{Ac}$ $\tilde{\alpha}_{9c}$
D. Summer after 8th grade	W_8, $S_{8/7}$	(1a) A (3a) W_9 (4a) $S_{A/9}$	$\tilde{\alpha}_{Ac}$ $\tilde{\alpha}_{9c}$

E/F. Summer after 7th grade or Fall 8th grade	W_7, $S_{8/7}$	(1) *A* (3a) W_9 (4a) $S_{A/9}$ (5a) W_8	$\tilde{\alpha}_{Ac}$ $\tilde{\alpha}_{9c}$ $\tilde{\alpha}_{8c}$

A/B: Predictions of educational attainment given data in Summer after 9th grade or Fall of 10th grade. As is evident in the table, chain-linking is required for all predictions other than for cases A and A/B; that is, predictions of educational attainment given data on 9th grade outcomes and school enrollment in high school. Since the two cases are essentially identical, we focus on case A/B. Given equations (1) and (11), the equations for predicting future educational attainment for future cohort *c* are given by the educational attainment equation:

$$A_{ic} = W_{9ic}\theta_A + S_{A/9ic}\alpha_{Ac} + e_{Aic} \tag{16}$$

and the school effect (shrinkage estimation) equation:

$$\alpha_{Ac} = \tilde{\alpha}_{Ac} + \tilde{\varepsilon}_{Ac} \tag{17}$$

Substituting the equation for the school effect into the attainment equation yields:

$$A_{ic} = [W_{9ic}\theta_A + S_{A/9ic}\tilde{\alpha}_{Ac}] + [S_{A/9ic}\tilde{\varepsilon}_{Ac} + e_{Aic}] \tag{18}$$

Predicted attainment $\tilde{A}_{ic}$ and the error in predicting/forecasting attainment $\tilde{\varepsilon}_{Aic}$ are given by the two terms in brackets in this equation:

$$\tilde{A}_{ic} = (W_{9ic}\theta_A + S_{A/9ic}\tilde{\alpha}_{Ac}) \tag{19}$$

$$\varepsilon_{Aic} = (S_{A/9ic}\tilde{\varepsilon}_{Ac} + e_{Aic}) \tag{20}$$

Predictions for new cohorts of students are obtained simply by plugging in values of predictors for these students, computing predictions using the estimated (calibrated) parameters from equation (1a) and adding the forecasted value of future school effects. This strategy is followed in all prediction models. The associated errors in prediction consists of three parts: estimation error in the coefficients of the prediction model, the student-level residual e_{Aic}, and error in forecasting the school effect $\tilde{\varepsilon}_{Ac}$. If the data sets used to calibrate the models are sufficiently large the first type of error will be quite small. We thus focus in this memo on the other two sources of error and have excluded the first type of error from the prediction equations.

The variance in prediction error due to the student-residual is equal to $\mathrm{Var}(e_{Ai}) = \sigma^2_{eAi}$, where the variance is subscripted by the student index *i* to allow for the possibility that this variance could in principle vary across students (the case of heteroscedastic error). It may be reasonable to suppose that this variance is constant. Indeed, since we have specified the educational attainment model as a binary probit model, we may accept the typical normalization in a probit model that $\mathrm{Var}(e_{Ai}) = \sigma^2_{eA} = 1$, as discussed above. The error in forecasting future school effects for school *k* is given by equation (12) in the case of a single school effect, as in the educational attainment model considered here, and by equation (15) in models where there is more than one outcome, as discussed below. Note that students enrolled in the same school share a common forecast error of the school effect. This needs to be considered when aggregating student data to the school level.

To obtain prediction of educational attainment given data from 9th grade or in prior grades it is necessary to chain-link models, as indicated in Tables 1 and 2. One of the major advantages of modeling student outcomes and school enrollment indicators using linear specifications (as predictors) and linear models (as outcomes) is that it is straightforward to chain-link the equations across grades. Chain-linking yields two statistics of interest: (1) predictions of future outcomes, including educational attainment, given predictors at a given grade, and (2) variances of prediction errors. The latter are used to compute errors in classification when students are assigned to risk categories based on predicted outcomes. Chain-linked equations are obtained by substituting for (or replacing) predictors in a model at given grade level with the equations that, in turn, predict these predictors (student outcomes and school enrollment indicators). Predictions over a long span are obtained by sequential substitution of multiple grade-specific equations. Below, we present the derived formulas for computing chain-linked predictions using the restricted models summarized in Table 2. Appendix B presents the comparable formulas for the unrestricted models summarized in Table 1.

C: Predictions of educational attainment given data in Fall of 9th grade. As indicated in Table 2, computing predictions of educational attainment given 9th grade data requires predictions of $A_{ic}, \alpha_{Ac}, W_{9ic}, \alpha_{9c}$. Equations for predicting the first two terms were given above in equations (16) and (17). Prediction equations for last two terms are given by:

$$W_{9ic} = W_{8ic}\theta_9 + S_{A/9i}\alpha_{9c} + e_{9ic} \tag{21}$$

$$\alpha_{9c} = \tilde{\alpha}_{9c} + \tilde{\varepsilon}_{9c} \tag{22}$$

The chain-linked equation is obtained, as above, by substituting equations (17), (21), and (22) into attainment equation (16). Equivalently, the chain-linked equation is obtained by substituting equations (21), and (22) into the prediction equation for system A/B – equation (18). This yields the following prediction equation:

$$A_{ic} = [W_{8ic}\theta_9\theta_A + S_{A/9ic}(\tilde{\alpha}_{Ac} + \tilde{\alpha}_{9c}\theta_A)] + [S_{A/9ic}(\tilde{\varepsilon}_{Ac} + \tilde{\varepsilon}_{9c}\theta_A) + e_{Aic} + e_{9ic}\theta_A] \quad (23)$$

Predicted attainment $\tilde{A}_{ic}$ and the error in predicting/forecasting attainment $\tilde{\varepsilon}_{Aic}$ are given by the two terms in brackets in the above equation:

$$\tilde{A}_{ic} = [W_{8ic}\theta_9\theta_A + S_{A/9ic}(\tilde{\alpha}_{Ac} + \tilde{\alpha}_{9c}\theta_A)] \quad (24)$$

$$\tilde{\varepsilon}_{Aic} = [S_{A/9ic}(\tilde{\varepsilon}_{Ac} + \tilde{\varepsilon}_{9c}\theta_A) + e_{Aic} + e_{9ic}\theta_A] \quad (25)$$

The key feature of this pair of equations is that the coefficient vector θ_A from the attainment equation multiplies all components from the equations for grade 9. As stated above, this coefficient vector acts as scoring weights in making predictions of education attainment A_i given prior grade variables. In particular, note that the coefficient matrix on W_{8i} is the product of θ_9 (a $P_8 x P_9$ matrix) and θ_A (a $P_9 x1$ vector), yielding a ($P_8 x1$) vector of scoring weights for 8th grade student outcomes. The variance of prediction/forecast errors $\tilde{\varepsilon}_{Aic}$ is constructed using values of the variances of each error component in (24), as defined above. Note that some of the error variances are also multiplied by the scoring weights θ_A.

D: Predictions of educational attainment given data in Summer after 8th grade. The difference between chain-link systems C and D is that the school to be attended in 9th grade is assumed to be unknown at the end of 8th grade and during the Summer before 9th grade. Since most students change schools between 8th and 9th grade, it is necessary to predict 9th grade school enrollment given 8th grade student outcomes and school enrollment using an equation like equation (4a):[7]

$$S_{A/9ic} = S_{8/7ic}\tau_9 + u_{A/9ic} \quad (26)$$

As above, the school enrollment equation is substituted for the unknown variable $S_{A/9i}$ in the chain-link equation for system C – equation (23), yielding the following prediction equation:

$$A_{ic} = [W_{8ic}\theta_9\theta_A + S_{8/7ic}\tau_9(\tilde{\alpha}_{Ac} + \tilde{\alpha}_{9c}\theta_A)] + [(S_{8/7ic}\tau_9 + u_{A/9ic})(\tilde{\varepsilon}_{Ac} + \tilde{\varepsilon}_{9c}\theta_A) + e_{Aic} + e_{9ic}\theta_A] \quad (27)$$

---

[7] In general, the need to predict future school enrollments depends on at what grade students transition from one school to another. For example, students must move to a new school at the end of the school year if they are 5th grade students enrolled in a K-5 school. The model considered in the memo illustrates the issues involved in modeling changes in school enrollments. Similar models could be implemented to address grades where school transitions are high. The unrestricted model allows for changes in school enrollment at any grade.

Predicted attainment $\tilde{A}_{ic}$ and the error in predicting/forecasting attainment $\tilde{\varepsilon}_{Aic}$ are given by the two terms in brackets in the above equation:

$$\tilde{A}_{ic} = [W_{8ic}\theta_9\theta_A + S_{8/7ic}\tau_9(\tilde{\alpha}_{Ac} + \tilde{\alpha}_{9c}\theta_A)] \quad (28)$$

$$\tilde{\varepsilon}_{ic} = [(S_{8/7ic}\tau_9 + u_{A/9ic})(\tilde{\varepsilon}_{Ac} + \tilde{\varepsilon}_{9c}\theta_A) + e_{Aic} + e_{9ic}\theta_A] \quad (29)$$

The equation for predicting attainment $\tilde{A}_{ic}$ is very similar to the prediction equation for system C, except that the high school enrollment vector is replaced by predicted high school enrollment $S_{8/7ic}\tau_9$, where τ_9 is a transition matrix that enables high school enrollment to be predicted given middle school enrollment. Predictions of school enrollment are thus transition probabilities that are in the (0,1) range. Moreover, students typically will have positive probabilities for multiple schools, although the probabilities will sum to one. The equation representing the prediction/forecast error $\tilde{\varepsilon}_{ic}$ has a new form and added complexity in that it includes error components that interact, rather than simply being additive. In Appendix C we address methods for computing the variance of errors that take this form.

E/F: Predictions of educational attainment given data in the Summer after 7th grade or Fall of 8th grade. To be added.

Evaluating the magnitude of school effects: general case

One possible challenge in estimating models with school effects is that at many grade levels, particularly in high school and the upper middle school grades, the number of outcome variables is quite large. (See Appendix A for a list of variables.) Some of these variables may have limited variance (and low means); for example, an indicator variable of whether a student is enrolled in a math course. There is strong justification for including variables of this type as predictors in models, but it could be problematic to estimate models with student-level predictors and school effects for these variables.

As a first step in addressing this issue we propose the following new approach for evaluating the magnitude of school effects; namely, construct composites of student outcomes using model coefficients. Above we noted that these coefficients act as scoring weights when using prior student outcomes to predict future outcomes. We consider two types of composites, an overall composite that includes all student outcome predictors for a given grade and composites that include subsets of these variables. As indicated in equations (1a) and (16), the coefficient vector θ_A is the appropriate vector of scoring weights for 9th grade student outcomes W_{9i}.[8] Similarly, as indicated in chain-linking equation (23), the matrix product of coefficients $\theta_9\theta_A$ is the appropriate vector of scoring weights for 8th grade student outcomes W_{8i}.

The overall composite based on 9th grade outcomes is thus obtained by multiplying equation (3a) by θ_A, which yields the following model of the composite variable $\tilde{A}_i(W_{9i}) \equiv W_{9i}\theta_A$:

$$\begin{aligned}\tilde{A}_i(W_{9i}) \equiv W_{9i}\theta_A &= W_{8i}\theta_9\theta_A + S_{9i}\alpha_9\theta_A + e_{9i}\theta_A \\ &= W_{8i}\theta_{9,A} + S_{9i}\alpha_{9,A} + e_{9,Ai}\end{aligned} \qquad (30)$$

where $\theta_{9,A} = \theta_9\theta_A$ is a coefficient vector (with dimension P_9) and $\alpha_{9,A} = \alpha_9\theta_A$ is the school effect (with dimension K_9) with respect to the composite variable. This equation can be estimated without imposing constraints on the parameters even though estimates of some of these parameters may be available from estimation of the predictive analytics models. The primary advantage of this approach is that it provides estimates of the magnitude of school effects for the overall composite, useful if it is infeasible to compute school effects for some of the student outcomes. The estimated noise-corrected variance of school effects for the predictive analytics composite can be compared with the noise-corrected variances from other models, such as value-added models of math and ELA test scores. In these models the noise-corrected variances as proportions of achievement in math and ELA tend to lie with the range of 0.02 to 0.10.

The above approach can also be applied to alternative composite variables based on subsets of student outcomes. As above, the composites are formed using elements of θ_A corresponding to the outcomes selected as scoring weights. Possible subsets of outcomes are listed below. This list references student outcomes listed in Appendix A. Additional information on course variables is provided in a later section:

1. All level indicators and grade point variables by subject.
2. All level indicators and grade point variables combined across subjects.
3. All diagnostic and program participation variables
4. Student attendance rate
5. Student test scores (not available in grade 9, but available in grades 3-8 and 11 in California.

[8] Given that educational attainment is a single outcome, rather than a vector of outcomes, the problems associated with estimating models for some 9th grade outcomes do not exist. Hence, the coefficient vector θ_A can be estimated in a model (equation 1a) that includes school effects – α_A. The objective of the models of composite variables is to estimate the variance of school effects for 9th grade outcomes.

As indicated above, composites can also be constructed using 8th grade variables, using the chain-linked scoring weights $\theta_9\theta_A$. One new challenge in this case is deciding what model should be used to estimate the matrix of coefficients θ_9, where each column of the matrix corresponds to the coefficients for a given 9th grade outcome. It may not be possible, as suggested above, to include school effects in all models of 9th grade outcomes. One option to insert into θ_9 coefficients from the models that included school effects if they can be estimated, but insert coefficients from the models that exclude school effects if they cannot be estimated. A second alternative is to use the estimate of $\theta_{9,A} = \theta_9\theta_A$ from equation (30), an equation that includes school effects as well as 8th grade student outcomes.

Evaluating the magnitude of school effects for subgroup variables and models of course enrollments and performance outcomes

Thus far, we have presented the attainment and 9th grade outcomes models in general terms without considering how best to model outcomes such as grade points by subject and course level that are only observed if a student was enrolled in a course in that subject and level. In particular, our preferred model categorizes courses in each subject area into three different difficulty levels (low, medium, and high) and allows the effects of course performance (measured by grade point on the standard 4.0 scale) to differ across course levels and subjects. The part of the attainment model that includes course enrollment and performance outcomes, represented by the latent variable A_{1i} (to distinguish it from the overall latent variable A_i) can be written as:

$$\begin{aligned} A_{1i} &= \sum_s \sum_{l=0}^{l=3} E_{9(sl)i}[\phi_{Asl} + G_{9(sl)i}\gamma_{Asl} + G(F)_{9(sl)i}\eta_{Asl}] + I(F)_{9i}\kappa_A \\ &= \sum_s \sum_{l=0}^{l=3} E_{9(sl)i}\phi_{Asl} + E_{9(sl)i}G_{9(sl)i}\gamma_{Asl} + E_{9(sl)i}G(F)_{9(sl)i}\eta_{Asl} + I(F)_{9i}\kappa_A \end{aligned} \tag{31}$$

where:

impacts are summed over subjects *s* and course levels *l=1,2,3* and where level *0* indicates that a student was not enrolled in any course in given subject *s*

$E_{9(sl)i}$ = binary (zero/one) indicator of enrollment by subject and level (*l=1,2,3*) and non-enrollment by subject

$G_{9(sl)i}$ = grade point on the 4.0 scale for the student enrolled in each course by subject and level. For simplicity, the grade point variable is entered as a linear variable (although see below for the effect of an F grade). This variable is set to zero if a student is not enrolled in each course by subject and level. For simplicity, we assume that a student is enrolled in no more than one full-year course in each subject and that all courses are full-year courses or combinations of one semester courses.

$G(F)_{9(sl)i}$ = a binary (zero/one) indicator of a student receiving an F grade in a given course by subject and level. This variable is highly related to the grade point variable $G_{9(sl)i}$ but it is included in the model to capture the possibility that the effects of grades are discontinuous at the F grade. This specification allows for the possibility that an F grade has a much larger negative effect than implied by the grade point variable entered in the model as a linear variable.

$I(F)_{9i}$ = a binary indicator of whether a student a student receiving more than one F grade across all courses. This variable allows for the possibility that there is an additional negative effect of receiving multiple F grades, above and beyond the effect given by the grade point and F indicator effects for a given course. The model could readily be extended to include additional indicators representing numbers of F grades greater than two or more.

The coefficients that correspond to the above variables are: $\phi_{Asl}, \gamma_{Asl}, \eta_{Asl},$ and κ_A.

The second line of equation (31) demonstrates that the equation can be represented as the sum of predictors multiplied by coefficients, the same form as equation (1). However, two of the predictors (the grade point and F grade predictors) involve interactions with the enrollment variables $E_{9(sl)i}$ (for subject *s* and level *l*). The interactions are: $E_{9(sl)i}G_{9(sl)i}$ and $E_{9(sl)i}G(F)_{9(sl)i}$.

Note: In the next iteration of this memo we will consider the merits of treating the interactions as single outcome variables in the 9th grade model versus splitting off the grade point and F grade variables and estimating separate models for these outcomes, but limited to the students enrolled in each course. One of advantages of the disadvantages of the treating interactions as single outcome variables is that it is difficult to interpret the variance of the school effects since they capture the combined effects of schools on enrollment and on course grades.

Are student outcomes measured on the same scale across schools and districts: Evaluation and implications for the validity of predictions across schools and districts

To be added.

Estimation of predictive analytics models and chain linking of estimated equations

A primary concern in estimating predictive analytics equations that are part of a system of chain-linked equations is whether and how to address possible endogeneities of student outcomes and possible model misspecifications. Endogeneities could arise in several ways.

If one or more of the predictors are measured with error, then it is well known that estimates of model parameters will be biased. Student test scores, for example, are always measured with error and information on test reliability and the precision of individual test scores, is routinely provided test vendors. The problem of measurement error is addressed in many models, including value-added models, by using errors-in-variables (EIV) estimators to control for pretest measurement error (Fuller, 1987). The EIV method is designed to produce consistent estimates of model parameters. Recent research at EA has demonstrated that in order to produce consistent estimates of student-level predictions (or, equivalently, student-level residuals) it is necessary to augment the EIV method by adding a second stage of measurement error correction. This method, which we refer to as double measurement error corrections (DMEC), is designed to produce consistent estimates of model parameters and student-level predictions.

The models presented in this memo assume that student outcomes over time can be modeled using a recursive structure in which models include only immediate prior lags of student outcomes, rather than multiple lags. This assumption can be tested, simply by including additional lags of outcomes in the models. Additionally, if the recursive assumption is retained, additional methods are available to control for measurement error, including using prior lags as instrumental variables (IV). This alternative method of controlling for measure error could be very valuable for outcomes such as course grades that may be measured with error, but for which there is no standard method of computing the magnitude of errors (as is the case with test scores). If the correct model includes double (or multiple) lags of predictors, then the prior lags do not qualify as valid instrumental variables. In this case, the strategy of using double lags as instrumental variables could be useful as a method of establishing bounds on predictions. It is plausible that the two sets of estimates – estimates using double lag variables as instrumental variables versus not using these variables is instruments – bound the true predictions. A key objective of producing predictions using both approaches is to test the robustness of estimates and predictions to alternative model assumptions

Another (complementary) method of assessing the validity of estimates and predictions produced using chain linking of equations is to compare predictions based on short-span versus long-span chain-linked equations. For example, if four years of longitudinal data is available, predictions of 8th grade outcomes could be produced using a long-span model that uses 5th grade outcomes as predictors. Predictions could then be compared with chain-linked equations of 8th on 7th grade outcomes, 7th on 6th grade outcomes, and 6th on 5th grade outcomes, with all equations estimated using data from the same cohort. If the coefficients obtained via chain linking do not equal the corresponding coefficients from the long-span model, this provides evidence that the coefficients in one or more of the short-span models are biased.

References

Fuller, W.A. (1987). *Measurement Error Models,* New York: John Wiley & Sons.

Longford, N.T. (1999). Multivariate shrinkage estimation of small area means and proportions, *Journal of the Royal Statistical Society*, 162(Part 2), 227-245.

Longford, N.T. (2005). Missing Data and Small-Area Estimation: Modern Analytical Equipment for the Survey Statistician, Springer.

Meyer, R.H. (2009). Shrinkage Estimation Formulas and References, memorandum.

Platek, R., Rao, J. N. K., Sarndal, C. E., Singh, M. P. (1987). *Small area statistics*. New York: John Wiley & Sons.

Rao, J.N.K. (2003). *Small Area Estimation*, New York: Wiley-Interscience.

Appendix A. A Multi-Stage Multilevel Causal Structure of Student Outcomes

This appendix presents a multi-stage path diagram that shows the types of student-level variables hypothesized to determine student outcomes. The predictive analytics model considered in this memo follows the typical strategy of including all variables in the model in order to maximize the explanatory power of the model.[9] In a future memo we will address how to explicitly model the multistage process that determines the predictors. This is an essential step if the objective of the analysis is to use the results to ask what would happen to students if they made different choices and had different experiences. The variables are presented in order of the hypothesized order of causal impact of the variables. Suggested variable names are included for each type of predictor.

For the purposes of this memo, the key difference between the predictors is that: (a) some are measured for all students (for example, enrollment variables and subject-specific test scores (typically measured at a given grade and (b) some are measured only for subsets of students (for example, course level grades and end-of-course test scores. In the memo, we propose a method for evaluating the impact of school effects (and other predictors) for the latter class of variables. In the diagram below, we denote these variables by $E \cdot G$, where E = a binary (zero/one) indicator of enrollment in a course level or program and G = the grade (on a standard 4.0 scale), an end-of-course test score, or program evaluation score.

[9] Note that it is necessary to employ a strategy for reducing the number of variables included in the model if there is severe multicollinearity among the full set of predictors (see, for example, Efron, xxxx). The models considered in this memo do not exhibit high multicollinearity.

Multi-Stage Multilevel Variables

Variable Name	Student-Level Variables
X	Student, family, and neighborhood resources and attributes
	Poverty status (free lunch eligibility) Parental education Student peer effects Note: additional variables could include those available in integrated data systems, for example information on health, family status, and homelessness.
S	School enrollment
D	Student diagnostics and program eligibility determinations Note: These diagnostics are measures of student status and are used to determine access to school-based inputs (see below)
	Gifted and talented Special education English language learner (ELL) Related: low performance outcomes
E	Inputs and Interventions
E1	Course enrollments
	By subject and difficulty level
E2	Programs and supports
	Tutoring Supplemental resources
F	Student inputs
	Student attendance
	Student Performance Outcomes: Cognitive. Non-cognitive, and Behavioral
$E \cdot G$	Outcomes for subsets of students
	Course grade points (on 4.0 scale) End-of-course test scores Program evaluation scores
Y	Outcomes for all students
	Subject-specific test scores (summative, interim, end-of-unit) Social emotional learning (SEL) Incidents and suspensions Interactions with the juvenile justice system Related: student attendance
Variable Name	**School-Level Variables**
Z	School and educator factors and quality
	Are advanced (e.g., AP) course offered? Are schools effective in hiring, developing, and retaining high quality educators? Is instruction aligned with (high quality) learning standards? Do educators have access to data on student progress during the school year? Are systems or protocols in place to ensure that resources are targeted to at-risk students?
S	Residual school effects not explained by observed variables

Appendix B. Chain-Linked Prediction Formulas for Unrestricted Prediction Models

To be added.

Appendix C. Methods for Computing the Variance of Errors with Interacted Error Components

To be added.