Senior Thesis in Mathematics

Multilevel Regression in Value Added Modeling for Teacher Assessment

Author:
Caroline Zaia

Advisor:
Dr. Jo Hardin

Submitted to Pomona College in Partial Fulfillment
of the Degree of Bachelor of Arts

April 2, 2015

Abstract

Value Added Modeling is one method for measuring teacher accountability. Using multilevel regression, a model is created to isolate teacher influence and control for student or classroom characteristics such as race, gender, class size, socioeconomic status, etc. Because of inherent dependencies between variables, basic linear regression is not sufficient to create an accurate model. For this reason, multilevel regression is used to create value added models. Multilevel regression consists of breaking the model into different levels and using the estimates from one level in the model of the next level. This allows the error structures and variance components to carry through correctly, culminating with a final composite model which measures teacher influence.

Contents

Chapter 1

Background

1.1 Introduction

Over the past 20 years, school, teacher, and student test-score performances have been increasingly utilized when considering the effectiveness of school systems in educating students. Additionally, there have been efforts not only to increase educators' accountability, such as the No Child Left Behind Act, but also to quantify exactly how effective teachers are in educating students. Recently, a select few school districts have become integral to the discussion regarding accountability by adopting a new method of evaluating teacher performance: Value Added Modeling (VAM). Tennessee was the first state to implement this method of evaluation, with the Tennessee Value-Added Assessment System, which began in 1992; other districts in Washington D.C., New York, and Chicago have also utilized VAM [9]. The method relies heavily on student test scores. The idea is that, based on the change in a student's test scores from one year to the next, the value a teacher adds to a specific student can be measured through a linear model, while controlling for other characteristics that affect a student's ability to learn. Districts that implemented VAM use it to assist with teacher evaluation and impact teacher salary, applications for tenure, and overall program evaluation [12].

There are two useful implications of using VAM to evaluate educators. First, it filters out the effect of non-academic factors on a student's education. The method incorporates information about factors that influence education, which can fall into three categories: student, classroom, and teacher characteristics. Student characteristics include a student's race, family income, gender, etc., and classroom characteristics include class size, and characteristics of the class as a whole, such as the percent of students who are a specific race or socio-economic status. Finally, teacher characteristics include the number of years a teacher has been at her current school and overall years of experience. This information is used to then isolate the effects of teachers and schools on student learning, which helps narrow the evaluation, focusing on accountability more than student improvement. Additionally, VAM can be a tool to help educate teachers. It is used to determine if a teacher is effective or ineffective. Common characteristics or teaching styles among effective teachers may then be exposed, and action can be taken to improve the skills of teachers who have been deemed less effective.

More specifically, VAM uses linear regression to create a model to measure teacher effectiveness. The purpose is to isolate the influence of a teacher by incorporating explanatory variables into the model in an attempt to remove the effect of extraneous influences, where the coefficient of interest will simply be teacher value added.

1.2 Reception of VAM

The adoption of VAM in teacher assessment is a controversial topic in the education world [4]. Proponents of the method believe that it increases accountability, as well as transparency. Education reformers support the use of VAM because it serves as a tangible measure of the skills of teachers, as related to how well students are performing. The *LA Times* received student information and test scores from about 6,000 elementary schools in the district and created a model to evaluate the teachers. The data came from test scores in math and English language arts from 2003-2010. The newspaper then published the results and created a searchable database, which can be accessed at `http://projects.latimes.com/value-added/`, where those interested could search a teacher or school by name and obtain the results of the analysis [10]. Proponents of VAM claimed that the works of the *Times* created more transparency in teacher evaluations and rightfully made that information public. Additionally, those in favor of VAM argue that it creates a more financially efficient and less subjective evaluation system. Instead of sending an evaluator into every teacher's classroom for a day or even half a day in order to assess that teacher's performance, the method allows computational assessment through data, which is much more financially efficient than employing evaluators to send to thousands of teachers' classrooms and allows for a more objective and standardized method of evaluation.

Opponents of the use of VAM claim that it relies too heavily on the use of standardized test scores to evaluate teachers [11]. While the debate regarding the use of standardized testing to assess students' knowledge and capabilities is a different topic entirely, it can be argued that standardized tests are not a sufficient tool to measure how much a student has learned or how much a teacher has taught. There are many areas in which a student can excel while still performing poorly on standardized tests [1]. Furthermore, in many cases, the teacher's career, pay, or even opportunity for tenure relies heavily on how students perform on tests. One could argue that a student does not have much incentive to try his or her best, however the tests are very influential on teachers' performance outcomes. This also creates an incentive for a teacher to "teach to the test," teaching students the specific information that is tested on instead of teaching students how to think, how to critically evaluate concepts and ideas, or even for the sole purpose of learning. Another critique is that, when using the VAM, teachers are ranked against each other. After each teacher receives a score, those scores are ranked against other teacher in the same school or district [4]. Rather than having a baseline score for what is considered a "good/effective" teacher, the ranking system simply compares teachers; a school where the vast majority of teachers are excellent will still have half the teachers perform below the median with this model.

1.3 Linear Regression

VAM uses linear regression to create a model to measure teacher effectiveness. The following is an explanation of basic linear regression, as explained by DeGroot and Schervish [3].

The goal of linear regression is to model the relationship between the dependent or response variable, Y, and the explanatory variables, $X_1, X_2, \ldots, X_i$. This is achieved by estimating the parameters $\beta_1, \beta_2, \ldots, \beta_i$ to predict Y. The method of least squares estimates the parameters by minimizing the sum of squared residuals.

The standard linear regression model:

$$Y_i = \beta_o + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_{p-1} X_{i,p-1} + \epsilon_i$$

$$\text{with } \epsilon_i \sim N(0, \sigma^2)$$

can be written in matrix notation as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where each term represents a vector or matrix of variables.

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \dots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{n,p-1} \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{p-1} \end{bmatrix}$$

Considering the equation

$$Q = \sum (Y_i - \beta_o - \beta_1 X_{i1} - \beta_2 X_{i2} - ... - \beta_{p-1} X_{i,p-1})^2$$

the estimated b_i will be those values of β_i that minimize Q.

Using matrix notation, we see that

$$Q = (Y - X\beta)^t (Y - x\beta)$$

and we need to minimize Q. This can be achieved by taking the derivative at each β and setting that derivative equal to zero.

$$\begin{aligned} \frac{\partial Q}{\partial \beta_0} &= 2\sum_{i=1}^{n}(Y_i - X_i\beta)(-1) = 0 \\ \frac{\partial Q}{\partial \beta_1} &= 2\sum_{i=1}^{n}(Y_i - X_i\beta)(-X_{i1}) = 0 \\ &\vdots \\ \frac{\partial Q}{\partial \beta_j} &= 2\sum_{i=1}^{n}(Y_i - X_i\beta)(-X_{ij}) = 0 \end{aligned}$$

which, in matrix notation, is the same as solving the equation

$$(Y - X\beta)^t(-X) = 0$$

$$Y^tX + X^tX\beta = 0$$

$$\implies X^tX\beta = Y^tX$$

$$\implies \mathbf{b} = [\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{Y}$$

where **b** is the vector of the estimates of β_i. Estimating this the coefficients in this manner is known as the method of least squares. Using the least squares estimates as a function of Y, we can also look at the variance and expectation of **b**.

$$\begin{aligned} var(\mathbf{b}) &= var[(X^tX)^{-1}X^tY] \\ &= (X^tX)^{-1}X^t var(Y) X(X^tX)^{-1} \\ &= (X^tX)^{-1}X^t \cdot \sigma^2 \cdot I \cdot X(X^tX)^{-1} \\ &= \sigma^2 \cdot (X^tX)^{-1} \end{aligned}$$

if $var(Y) = \sigma^2 \cdot I$. We can also see that

$$\begin{aligned} E(\mathbf{b}) &= E((X^tX)^{-1}X^tY) \\ &= (X^tX)^{-1}X^tE(Y) \\ &= (X^tX)^{-1}X^tX\beta \\ &= \beta \end{aligned}$$

The expectation of **b** shows that the estimate is unbiased since $E(\mathbf{b}) = \beta$.

Chapter 2

Multilevel Regression

2.1 Motivation for the Different Models

Throughout this section, I will create a Value Added Model using different methods of regression. For all the models, the assumption exists that students take multiple standardized test per year (for example, one per month), which allows for more observations. The regression equations will utilize the terms Y_k representing test scores, T_k and $Math_k$ which are indicator variables as defined in the following section, ϵ_k which is the normally distributed error term, and β_i, the parameter(s) for the explanatory variable(s). Finally, the error structures throughout the three different models will be important, and a key difference in determining the accuracy of the models.

2.2 Model 0

The first method, Model 0, is basic linear regression using method of least squares estimation. In order to simplify the derivations and motivation, I will restrict the many possible variables to only two. Additionally, the subscript k represents the observation. One observation includes a test score, whether that test was math or ELA, and whether the student had the teacher of interest. These three components together form a single observation, k. Model 0 is as follows:

$$Y_k = \beta_0 + \beta_1 T_k + \beta_2 Math_k + \beta_3 T_k Math_k + \epsilon_k$$

$$\text{where } \epsilon_k \sim N(0, \sigma^2)$$

$$T_k = \begin{cases} 1 & \text{if } k^{th} \text{ observation had teacher of interest} \\ 0 & \text{else} \end{cases}$$

$$Math_k = \begin{cases} 1 & \text{if } k^{th} \text{ score is from a math test} \\ 0 & \text{if } k^{th} \text{ score is from an ELA test} \end{cases}$$

In the model above, β_1 is the crucial component. This parameter is the measure of a teacher's influence, thus produces the value, or score, of interest. Eventually, the main question will be whether β_1 is 0. The error term in Model 0, ϵ_k, is normally distributed and represents the error in each observation, k.

The main problem with using the method of least squares for this type of dataset is that it operates under the assumption that all observations, or all test scores, are independent. This is not the case. We could have, for example, six test scores from one student and four from another. Each test score is treated as an independent observation, disregarding the fact that multiple test scores come from the same student. The lack of independence violates the assumption made in Section 1.3 that the $var(Y) = \sigma^2 \cdot I$. Inherent in the variance is the fact that $cov(Y_i, Y_j) = 0$, which is not necessarily true if the observations are not independent. The lack of independence causes the allusion that the sample size is very large (assuming each observation is independent), which will consequently underestimate the amount of variability in our model. The underestimation of the variance will then overestimate the test statistic and result in a lower p-value. The decreased p-value produces more false positives meaning the coefficients (βs) will seem positive when they may not be. Teachers may seem more effective than the data suggests. Therefore, we need an alternative method for creating the model.

2.3 Multilevel Regression - Method 1

Due to the inaccuracy of Model 0, we need a slightly different method of estimation. The model is the first method of multilevel regression, Model 1, is split into two different levels, and the estimates from the first level are incorporated into the equation in the second level. Further explanation of multilevel regression can be found in *Broadening Your Statistical Horizons: Generalized Linear Models and Multilevel Models* [6].

Level 1

$$Y_{ij} = \lambda_{0i} + \lambda_{1i} Math_{ij} + \epsilon_{ij}$$

where $\epsilon_j \sim N(0, \sigma^2)$ and

$$Math_{ij} = \begin{cases} 1 & \text{Score comes from a math test} \\ 0 & \text{Score comes from an ELA test} \end{cases}$$

Level 2

$$\lambda_{0i} = \beta_{00} + \beta_{01} T_i + \epsilon_{0i}$$

$$\lambda_{1i} = \beta_{10} + \beta_{11} T_i + \epsilon_{1i}$$

where $T_i = 1$ if Student i has the teacher in question and

$$\begin{bmatrix} \epsilon_{0i} \\ \epsilon_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_1^2 \end{bmatrix} \right)$$

In these equations, i represents the i^{th} student and j provides a way to index student i's tests. The subscript k is no longer used because we are no longer considering the test score, subject, and teacher together as one "observation." Instead, we are utilizing each test score from each student for whom we have data. ϵ_{0i} is the error term for students' math test scores, and ϵ_{1i} is the error term for students' ELA test scores.

For each student i, Level 1 will estimate a λ_0 and a λ_1; if there are n students, then there will be n pairs of λ estimations. In this level, the estimations are as follows:

$$\hat{\lambda}_{0i} = \text{average ELA test score for student } i$$
$$\hat{\lambda}_{0i} + \hat{\lambda}_{1i} = \text{average math test score for student } i$$

In Level 2, the estimates are used within the linear regression on the second set of equations. The focus then turns to the β estimates. To find the meaning of these estimations, consider the expectation of each equation in Level 2.

$$T_k = 1 \Rightarrow E(Y) = \beta_{00} + \beta_{01} + \beta_{10} + \beta_{11}$$
$$T_k = 0 \Rightarrow E(Y) = \beta_{00} + \beta_{10}$$

$\therefore \hat{\beta}_{01} + \hat{\beta}_{11} =$ Estimate of value added measure of teacher of interest to Math

Similarly, for ELA scores,

$$T_k = 1 \Rightarrow E(Y) = \beta_{00} + \beta_{01}$$
$$T_k = 0 \Rightarrow E(Y) = \beta_{00}$$

$\therefore \hat{\beta}_{01} =$ Estimate of value added measure of teacher of interest to ELA

After looking at the expectations, it is also helpful to compute the variance. However, we must first write Model 1 in composite form to make the computation of the variance more clear. The composite model is simply formed by replacing the λs in Level 1 with the equivalent equations from Level 2.

$$Y_{ij} = \beta_{00} + \beta_{10}T_i + \epsilon_{0i} + \beta_{10}Math_{ij} + \beta_{11}T_iMath_{ij} + \epsilon_{1i}Math_{ij} + \epsilon_{ij}$$

$$\begin{aligned}
var(Y_{ij}) &= var(\beta_{00} + \beta_{10}T_i + \beta_{10}Math_{ij} + \beta_{11}T_iMath_{ij} + \epsilon_{0i} + \epsilon_{1i}Math_{ij} + \epsilon_{ij}) \\
&= var(\epsilon_{0i} + \epsilon_{1i}Math_{ij} + \epsilon_{ij}) && \text{(After removing the constants)} \\
&= var(\epsilon_{0i} + \epsilon_{1i}Math_{ij}) + var(\epsilon_{ij}) && \text{(Because of independence)} \\
&= var(\epsilon_{0i}) + Math_{ij}^2 var(\epsilon_{1i}) + 2Math_{ij}cov(\epsilon_{0i}, \epsilon_{1i}) + var(\epsilon_{ij}) \\
&= \sigma_0^2 + Math_{ij}\sigma_1^2 + \sigma^2 && \text{(Since } cov(\epsilon_{0i}, \epsilon_{1i}) = 0\text{)}
\end{aligned}$$

As shown above, Model 1 addresses the lack of independence that existed in Model 0 by running the regression for each individual student, ensuring that test scores from the same student are all used simultaneously. While this method of multilevel regression is more accurate than Model 0, it does have some drawbacks. First, this method does not account for correlation between test scores of different subjects from the same student. Although it does account for correlation between same subject test scores from the same student (as was the goal in fixing the problem of independence found in Model 0), Model 1 does not address the correlation of a student's test scores in different subjects. In Level 2, there are two separate equations used, one for Math scores and one for ELA scores. Thus, there is no connection between the two equations, i.e., no connection between the two different subjects, even if the scores came from the same student. Another drawback of this method is that it gives equal weight to each student, regardless of the number of test scores available from that student. A student who has 8 observations (test scores) is given equal weight as a student who only has 4 observations.

2.4 Multilevel Regression - Method 2

In an attempt to address the drawbacks of Model 1, the second method of multilevel regression, Model 2, uses similar equations from the two separate levels of Model 1, but combines the equations into a composite model to allow correlation between the components of Model 1.

Composite Model

$$Y_{ij} = [\beta_{00} + \beta_{10}T_i + \beta_{10}Math_{ij} + \beta_{11}T_iMath_{ij}] + [\epsilon_{0i} + \epsilon_{1i}Math_{ij} + \epsilon_{ij}]$$

with $\epsilon_{ij} \sim N(0, \sigma^2)$ and

$$\begin{bmatrix} \epsilon_{0i} \\ \epsilon_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \\ \sigma_{01} & \sigma_1^2 \end{bmatrix} \right)$$

The two levels in this method are very similar to those in Method 1, with the variables and parameters equivalently defined. This method uses Maximum Likelihood Estimation to estimate the 8 parameters, the βs and the σs. The term in the bottom left corner of the variance components matrix, σ_{01}, is key. Model 2 includes a covariance term between math and ELA, which accounts for the correlation between tests in Math and English Language Arts from the same student. Since Model 2 accounts for this correlation, the distribution matrix of the error terms ϵ_{0i} and ϵ_{1i} must include the covariance term σ_{01}. This error structure does make the coefficients harder to estimate.

Recall that ϵ_{0i} represents the person to person variability associated with differences in the intercept in the math scores equation, meaning the deviation of student i from the mean test scores after accounting for the teacher. ϵ_{1i} signifies differences in slope, meaning the deviation of student i from the mean differences in test scores after accounting for the teacher. The intercept and slope are negatively correlated, which implies that an improved model is one that gives multivariate structure to the error terms.

We can also look at the expectation and variance of the model.

$$E(Y_{ij}) = \beta_{00} + \beta_{10}T_i + \beta_{10}Math_{ij} + \beta_{11}T_iMath_{ij}$$

The expectation exhibits the important parameters and what they signify, as shown in Section 2.2.

$$\begin{aligned}
var(Y_{ij}) &= var(\beta_{00} + \beta_{10}T_i + \beta_{10}Math_{ij} + \beta_{11}T_iMath_{ij} + \epsilon_{0i} + \epsilon_{1i}Math_{ij} + \epsilon_{ij}) \\
&= var(\epsilon_{0i} + \epsilon_{1i}Math_{ij} + \epsilon_{ij}) && \text{(After removing the constants)} \\
&= var(\epsilon_{0i} + \epsilon_{1i}Math_{ij}) + var(\epsilon_{ij}) && \text{(Because of independence)} \\
&= var(\epsilon_{0i}) + Math_{ij}^2 var(\epsilon_{1i}) + 2Math_{ij}cov(\epsilon_{0i}, \epsilon_{1i}) + var(\epsilon_{ij}) \\
&= \sigma_0^2 + Math_{ij}\sigma_1^2 + 2Math_{ij}\sigma_{01} + \sigma^2
\end{aligned}$$

The additional allowance for correlation between students' scores in Math and ELA in Method 2 leads to the correlation in error structure, seen both here with the existence of σ_{01} in the variance as well as in the variance components matrix on the previous page.

2.5 Restricted Maximum Likelihood Estimation

The method of least squares was explained in a previous section, however, that method of estimation is not sufficient when estimating parameters from the two-level model, especially when estimating the variance components. This is where restricted maximum likelihood estimation is useful.

To begin, restricted maximum likelihood estimation (REML) takes the regression residuals from the fixed effects section of the model and considers what the statistical model is for those residuals. Then, we perform maximum likelihood estimation on the residuals; the results are estimates of the variance components, which are the estimates of interest. REML adjusts the estimates of the variance for the fact that we are using residuals. Conceptually, the variance is supposed to measure how far data points are from the population line. Instead, the variance components measure how far the points are from the sample line, which leads to the underestimation of parameters. This is why REML is used. REML produces unbiased estimates of variance components and is more useful in this situation since maximum likelihood estimates are biased. The way the estimates are adjusted towards unbiasedness is by using a denominator of $n-p$ for the error variance, where p is the number of fixed effects parameters, as stated by Oehlert and shown below [8].

MLE of σ^2:

$$\hat{\sigma}^2 = \frac{\sum(X_i - \bar{X})^2}{n}$$

where

$$\sigma^2 = E(X-\mu)^2 \approx \frac{1}{N}\sum_{i=1}^{N}(X_i-\mu)^2$$

We can also compute the following expectation

$$\begin{aligned}
E(\Sigma(X_i - \bar{X})^2) &= E[\Sigma(X_i - \mu + \mu - \bar{X})^2] \\
&= E(\Sigma[(X_i-\mu)^2 + (\mu-\bar{X})^2 + 2(X_i-\mu)(\mu-\bar{X})]) \\
&= \Sigma E(X_i-\mu)^2 + n\cdot E(\mu-\bar{X})^2 + 2E[(\mu-\bar{X})\Sigma(X_i-\mu)] \\
&= \Sigma E(X_i-\mu)^2 + n\cdot E(\mu-\bar{X})^2 - 2E(\bar{X}-\mu)^2 \\
&= \Sigma\sigma^2 + n\cdot\frac{\sigma^2}{n} - 2\sigma^2 \\
E(\Sigma(X_i - \bar{X})^2) &= (n-1)\sigma^2
\end{aligned}$$

and we know

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$$

The above expectation shows that the MLE is biased (it underestimates σ^2 on average). Due to this underestimation, we use

$$\tilde{\sigma}^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$$

$$\Rightarrow E(\tilde{\sigma}^2) = \sigma^2$$

The point is that MLE's are often biased because they maximize the likelihood based on parameter estimates instead of true parameters. As seen above, $E(\sum(X_i - \bar{X})^2) < E(\sum(X_i - \mu)^2)$. Thus, we use REML because it adjusts the estimate of the variance, making it unbiased.

Chapter 3

Application

3.1 New York Model

Thus far, a value added model containing only two of the possible variables representing characteristics that influence education has been presented. The following is an actual model used in school districts in New York, as discussed in the technical report published by the Wisconsin Center for Educational Research [2].

$$Y_{1i} = \zeta + \lambda y_{0i} + \lambda^{alt} y_{0i}^{alt} + \beta' X_i + \gamma' Z_i + \alpha' J_i + \epsilon_i \tag{3.1}$$

Y_{1i}: Measured student achievement

ζ: intercept parameter

y_{0i}, y_{0i}^{alt}: true prior achievement (same and other subject)

λ, λ^{alt}: slope parameters

β', γ': slope parameters

X_i: vector of student characteristics of student i

Z_i: vector of student i's classroom characteristics

J_i: vector of teacher indicators

α: vector of teach value-added effects

ϵ_i: error in predicting student achievement, given the explanatory variables

3.2 Multilevel Regression

Level 1

The first level of regression addresses student level variables. In this level, the parameters λ and β will be estimated, as the coefficients of pretest scores and demographic variables, respectively. The equation

$$Y_{1i} = \lambda Y_{0i} + \lambda^{alt} Y_{0i}^{alt} + \beta X_i + \alpha C_i + \epsilon_i$$

can be written in vector form as

$$Y_t = Y_{t-1}\lambda + W\delta + \epsilon$$

where Y_t represents post test scores, Y_{t-1} represents same-subject and alternate-subject pre-test scores, λ is a vector of coefficients of those pre-test scores, W is comprised of the student and classroom variables, δ is a vector of those variables' coefficients, and ϵ is the vector of error terms. As seen in Section 1.3, the method of least squares would result in biased estimates in the form of

$$\begin{bmatrix} \hat{\lambda} \\ \hat{\delta} \end{bmatrix} = \begin{bmatrix} Y'_{t-1}Y_{t-1} & Y'_{t-1}W \\ W'Y_{t-1} & W'W \end{bmatrix}^{-1} \times \begin{bmatrix} Y'_{t-1}Y_t \\ W'Y_t \end{bmatrix}$$

which is the same for as

$$\mathbf{b} = [\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{Y}$$

Due to accuracy errors stemming from measurement error, the following is the measurement error corrected regression, as explained in Fuller's *Measurement Error Models* [5].The adjusted estimates of the coefficients in the vectors λ and δ are

$$
\begin{bmatrix} \hat{\lambda} \\ \hat{\delta} \end{bmatrix} = \begin{bmatrix} Y'_{t-1}Y_{t-1} - \sum_{i}^{N} sem_{it-1} & Y'_{t-1}W \\ W'Y_{t-1} & W'W \end{bmatrix}^{-1} \times \begin{bmatrix} Y'_{t-1}Y_t \\ W'Y_t \end{bmatrix} \tag{3.2}
$$

where sem_{it-1} is the variance-covariance matrix of measurement errors of test scores for student i at time $t-1$.

When considering the explanatory variables, the variable's effects are considered random if the levels of the variable included in the model can be thought of as a sample drawn from a larger population of potential levels that could have been selected [7]. The extra variability is accounted for by estimating the coefficients using the adjusted estimates of λ and δ, which include sem_{it-1}.

Level 2

In the second level, the regression will produce estimates of γ, as the parameter for classroom level variables in equation 3.1. Using the estimations from Level 1, we can create an estimate for the variable q_{1i} by letting $q_{1i} = Y_{1i} - \hat{\lambda}Y_{0i} - \hat{\lambda}^{alt}Y_{0i}^{alt} - \hat{\beta}X_i$ as estimated in equation 3.2. Then, q_{1i} becomes the response variable for this second level of regression.

$$
q_{1i} = \zeta + \gamma Z_i + w_i
$$

where $w_i = \alpha J_i + \epsilon_i$ but is simply treated as the error term in Level 2 modeling. The second level of regression will result in estimates for γ, again, incorporating the estimates from the first level.

Level 3

After controlling for student level and classroom level variables, the third stage will take that information into account while estimating teacher influence, α. We know

$$q_{1i} = Y_{1i} - \lambda Y_{0i} - \lambda^{alt} Y_{0i}^{alt} - \beta X_i \text{ and } q_{1i} = \zeta + \gamma Z_i + w_i$$

which results in our level 3 model:

$$w_i = \alpha J_i + \epsilon_i$$

where $w_i = Y_{1i} - \zeta - \lambda Y_{0i} - \lambda^{alt} Y_{0i}^{alt} - \beta X_i - \gamma Z_i$ and α is the measure of interest. All of the necessary parameters and pieces are estimated in the previous two levels of regression. In Level 3, we use those estimates to find w_i, which produces the α value of interest.

While it was not explicitly stated in the technical report that explained the formation of this model, it is clear from the idea and usage of the multilevel regression that an important component is the variance matrix and how it describes the dependence and correlation between the estimations of parameters.

Chapter 4

Conclusion

As shown above, method of least squares estimation is not an adequate way to create a value added model for teacher assessment. Thus, multilevel regression is used. Model 1 of multilevel regression attempts to address the issues regarding independence faced by using basic linear regression. However, Model 2 does a better job of dealing with the dependence in measured variables. While controversy does exist around Value Added Modeling, concerning whether teacher pay or job security should be based on student test scores, and furthermore, whether standardized test scores are truly an adequate metric for what or how much a student has learned, multilevel regression is an appropriate method for creating a value added model.

Bibliography

[1] Eva L. Baker, Paul E. Barton, Linda Darling-Hammond, Edward Haertel, Helen F. Ladd, Robert L. Linn, Diane Ravitch, Richard Rothstein, Richard J. Shavelson, and Lorrie A. Shepard.

[2] Value Added Research Center. Nyc teacher data initiative: Technical report on the nyc value-added model 2010. *Wisconsin Center for Education Research*, 2010.

[3] Morris DeGroot and Mark Schervish. *Probability and Statistics*. Pearson Education, 2012.

[4] S. Dillon. Formula to grade teachers' skill gains acceptance, and critics. *The New York Times*, 2010.

[5] Wayne A. Fuller. *Measurement Error Models*. John Wiley & Sons, Inc., 1987.

[6] J. Legler and P. Roback. *Broadening Your Statistical Horizons: Generalized Linear Models and Multilevel Models*. Taylor & Francis Group, 2013.

[7] Ramon Littell, Walter Stroup, and Rudolf Freund. *SAS for Linear Models*. SAS Publishing, 2002.

[8] Gary Oehlert. A few words about reml. *Carnegie Mellon University*, 2012.

[9] William Sanders and Sandra Horn. Research findings from the tennessee value-added assessment system (tvaas) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, page 248, 1998.

[10] LA Times staff. Los angeles teacher ratings. *Los Angeles Times*, 2011.

[11] Valerie Strauss. The fundamental flaws of ‘value added teacher evaluation. *The Washington Post*, 2012.

[12] M. Winerip. Evaluating new york teachers, perhaps the numbers do lie. *The New York Times*, 2011.