ESTIMATES OF SCHOOL PRODUCTIVITY AND IMPLICATIONS FOR POLICY

A Thesis presented to the Faculty of the Graduate School
University of Missouri-Columbia

In partial Fulfillment
Of the Requirements for the Degree

Master of Arts

by
XIAO PENG

Doctor Michael Podgursky and Jianguo Sun, Thesis Advisors

AUGUST 2007

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled:

ESTIMATES OF SCHOOL PRODUCTIVITY

AND IMPLICATIONS FOR POLICY

presented by Xiao (Art) Peng,

a candidate for the degree of Master of Arts,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Michael Podgursky

Professor Jianguo (Tony) Sun

Professor Xiaoguang (Shawn) Ni

Professor Lori Thombs

………………………….Thanks, Mom and Dad.

ACKNOWLEDGEMENTS

I would like to thank my economics advisor, Dr. Michael Podgursky, for his teaching in American educational research and for his advice on my thesis paper. Dr. Podgursky introduced me to the educational study of labor economics, and raised my interest in examining the difference in difference estimation for educational policy reform. He pointed to the literature for the value-added studies and my research plan. Dr. Podgursky generously invested the tremendous resources for the economics portion of my research paper. In addition, I would like to thank my statistics advisor, Dr. Jianguo Sun, for his instruction to longitudinal data analysis and his seminar course work. His seminar works in applying the statistical methods for the various areas in analyzing the longitudinal data.

I am also indebted to Dr. Shawn Ni for his incredibly generous support throughout my master thesis paper. Dr. Ni offered the relevant material, suggestions and guidelines in working on the complicated statistics methodology.

Other thanks are given to Dr. Mark Ehlert and Margie Gurwit for their consulting services and programming support. Dr. Ehlert helped prepare the sample data sets, and provided the knowledgeable discussion about the education background of the school district; Margie assisted in the programming and graph designing.

Final thanks are given to the school district[1] for the contribution of the administrative student data. Without its generous assistance, I would not have been able to complete my thesis paper. Any mistakes are the fault of the author.

[1] The name of the school district is ignored in this paper because the provider of the student data does not allow me to release the source to the public.

TABLE OF CONTENTS

LIST OF TABLES AND PLOTS

ESTIMATES OF SCHOOL PRODUCTIVITY AND IMPLICATIONS FOR POLICY

Xiao Peng

Doctor Michael Podgursky and Jianguo Sun, Thesis Advisors

ABSTRACT

School productivity was not perfectly estimated because of the sampling error and the measurement error. The traditional Ordinary Least Square (OLS) leaves the estimation of school productivity questionable. Moreover, Hierarchical Linear Model (HLM) encounters a large proportion of the variance unexplained in the level-1 equation. In the paper, I will first introduce the Kalman Filter (KF) algorithm together with the Bayesian random draw mechanism to simulate the accurate school effects, and then compare the simulated results with the estimates generated from OLS and HLM. The comparison of the school effects will conclude that the Kalman Filter is more reliable and accurate for the educators and school administrators to supervise the allocation of the school resources for school improvement.

Introduction

The current focus of American K-12 education is to raise students' academic performance. When we look back, American K-12 education has endured many changes, and the investment per student has increased in the past four decades. Within the school, data show that class size gets smaller. Most of the states have experienced the increasing teacher-pupil ratio since 1960s. In addition, the spending per student in K-12 education increased remarkably. However, student performance was stagnating, and concern about education quality was intensely argued by the school teachers, principals, and students' parents.

The simple production theory uses student performance as a measure of output changes when putting educational inputs into a school. Will we increase spending in schools? Did we invest enough money to achieve the desired changes for student achievement? The more proper argument is that the spending in K-12 education is inefficient and misused by schools. In other words, the current educational policies are ineffective to raise the school quality, especially in increasing student test scores. In order to establish an effective educational policy for students' academic performance, the economists should begin looking for the accurate estimate of the school effects within a school district and across multiple districts since schools affect student performance.

When thinking of schooling as the investment for the human capital, we understand that the inputs the students have today will influence their future productivity. Generally speaking, education benefits students and economic growth. Two facts come out from the empirical study of labor economics. The first fact is that wage is strongly correlated with the student's educational level. Another one is that China's economy grew at scorching pace when China had a tremendous number of Masters and PhD students in its local labor market.

Where do our students go to obtain their best education? How do we know which institutions provide the best education? This is a common issue within classrooms. Parents care about the quality of the school. Even though students go to schools near their living places, classroom arrangement happens all the time. Parents frequently influence, pressure, or request school administrators for a specific teacher or classroom assignment. The school principals, however, only have limited capability in identifying the school personnel's performance. For example, principals will not know the quality of the new incoming hiree until he/she is hired for teaching. In addition, parents have much less knowledge about teacher and school capacity than these educators. Therefore, this paper will find a way to present an estimate of school productivity in order to purify the school effects and to raise the effectiveness of the education policy in working for the improvement of the education quality. In addition, the paper opens a quantitative method in the future study of the teacher effects.

To carefully look at the reasons why K-12 students did not have much change on their performance, this paper pays much more attention on the school productivity which creates the differences across the schools within one medium-size school district. The estimates of the school productivity will present the effect of the overall school on student performance by years. The result is not to lead the parents to pick a school for their children. The primary purpose of this paper is to use the sophisticated statistical method to evaluate the school productivity. This paper is supposed to explain the estimation method to the school administrators and educators using the less technical terms.

When observing the school as a whole, the estimate of school effects will have more accuracy in the contribution to students' performance by using the Kalman Filter and Bayesian methods. From the statistics point of view, these sophisticated statistical methods take into account the measurement error on the test scores, the small sample issue on the test takers within schools, and the recursive estimation on smoothing school productivity. The simulation results generated from a large number, at least from the economics point of view, are more reliable than other estimates generated from traditional Ordinary Least Square (OLS), and much dependable than Hierarchical Linear Model (HLM) analysis. The school principals and educators shall be made aware by the implications of the paper of how to self-examine education inputs, such as teacher turnover, curriculum design, and school characteristics. When looking at the estimation results, these people have the ability to identify school performance when holding student demographic factors constant. We will see the differences of school effects crossing all possible years between schools. The higher quality school raises the students test score above the z-score mean, but the lower quality school pushes down the test performance.

This paper enables to extension to the study of teacher effects across schools. As far back as the "Coleman Report", which states the family effect as the significant source of contribution to student performance, the current research finds that school matters to students. Many researches went beyond the school level data, and studied the teacher effects within and across schools. Rivkin, Hanushek and Kain (2005) find that the large variation between schools matters to student achievement. Aaronson, Barrow and Sander

(2003) also reached the same conclusion concerning the educational and statistical significance of teachers. They focused on the relationship between teacher effects and student mathematical achievement at 9^{th} grade. However, evaluating teacher effects is constrained by the lack of data and the uncertain usage of the sophisticated statistical mechanisms. In addition, the unobservables of school and student status are driving much of the dispersion in teacher quality (Rivkin, Hanushek and Kain, 2005). The lack of data and noise of the student and teacher individual data cause this research progress slowly. In this paper, the school data allows me to estimate the school productivity. When well establishing the iteration and prior setting at school level, the paper will be the baseline for estimating the teacher effects.

This paper focuses on school productivity, and reports these estimates by schools by years within a school district. There are three questions raised and addressed in this paper.

1. What estimates of school effects are precise enough?
2. Is the methodology transparent enough to the policy-makers and educators?
3. What does the paper suggest to raise student performance through educational policy?

This paper examines the school productivity on student academic performance for elementary school students at 3^{rd}, 4^{th} and 5^{th} grades. The model in use is the value-added model following Todd and Wolpin (2003), which is a general cumulative model for student achievement. The idea is to have the student prior one-year-test score as the key control variable and to estimate the current one year school productivity based on the student prior performance. In my paper, the current and prior tests are not the same type but they belong to the same subject because of the lack of test scores taken in the continuous grade span. This paper, for example, studies MAP Math test for the 4th grade test takers regressed on TerraNova Math for the 3rd grade identical test takers. The student test data shows that the students only take MAP Math test in 4th, 8th, and 10th grade levels in any years before 2006. Moreover, the students only take TerraNova Math at 3rd, 5th, 6th, 7th, and 8th grade level, but I only have the data available in both 2004 and 2005. I also have other test combinations, which depend on the available tests the students could have in the data set. However, it is not likely to have the same test for estimating the school productivity crossing all the school years.

As many papers mentioned, the noise within the student and school data would be the one-time event to the students, the large sample variation and the measurement error of the test scores. This paper is working on simulating the school effects from the well-known distribution at large times and on filtering out the test score noise and small sample issue. The traditional OLS regression will cause inaccurate estimates because the data noise is more likely to miscompute the magnitude of the school productivity. Moreover, by applying multilevel analysis, HLM still leaves a large proportion of the variance unexplained. To reduce the inaccurate estimate, we import the computation method before each simulation loop in adjusting the measurement error for the test score(s). The comparison across OLS, two-level HLM and KF simulation results will provide the dispersion of the noise to the accurate school effects.

By doing so, the paper attempts to address two major estimation issues. First, the current research fails to correct the measurement error within test scores. Moreover, most studies of school productivity encountered the sampling variation and one-time factors issues (Kane and Staiger, 2001) that can contaminate the student test measures. The other concern about contaminating the estimates is that students have extra learning hours outside of school.[1] These student test scores will benefit from extra study done by studying effort outside of school. Second, the current methods do not have the strong confidence on isolating the school productivity from other potential factors that influence test scores. The correlation between schools and these factors that are kept in the linear model is further likely to counteract the school productivity estimated by test score measures.

There is another significant issue to this type of research. The administrative student data is not easily accessed by economic researchers. My data is provided from one medium-size K-12 school district with 19 elementary schools. The longitudinal data allows me to control for the student cognitive ability, sex, race and Free/Reduced Lunch participation in the test score regression and simulation. The current test score depends on the prior one-year test score and the students' endowed capacity in learning. In addition, the school effects absorb the achievement gain when holding other students' characteristics constant[2].

Once we have the confident estimate from the economics analysis, we could play around with these schools to raise student achievement. For instance, the high-performing school will have the better educational inputs than these low-performing ones. When carefully identifying these differences across schools, the policy-makers will set up a clear method to provide the effective educational inputs to the poor schools in order to

[1] This issue is obvious when estimating the language/reading test score. This paper does not provide any analysis for any other test subjects except Mathematics.
[2] The selection about the test score or percentile rank in determining the student achievement gain is a debatable issue. I use test score measures because the student test data do not provide the percentile rank for the most tests.

raise student achievement gain. The implication of the future American K-12 educational policy will not progress the student performance when the policy cannot in practice directly focus on the estimation of school productivity and the factors causing the differences in test scores across schools.

The paper is organized in the following way. The next section briefly summarizes the existing literatures. The motivation emphasizes the importance of quantitative analysis in education. The description of the data explains the student data structure paired by test grades. The methodology comparisons were established in the model and theory portion. The follow-up sections report the results of school fixed effects in predicting mathematics test score for elementary school students, and open the discussion for the policy implication. The conclusion is the last section.

Literature Review

School productivity can be measured as the effect to student learning within a fixed period of time. The study of school effects helps identify the effective schools and facilitate school improvement. Raudenbush and Willms (1995) define two types of the school effects. One can be viewed as an effect through the observation of parents picking out a school for a child. Another one can be viewed as an effect favoring the school administrators that evaluate the performance of the school. Their studies mainly examine the latter of the school effects.

In their quasi-experiments of school effects, they estimate both school effects and their variance by looking at the school practice, school context, and student background. The model allows for the interaction of effects between school practice and context crossing all students in a school. The decomposition of the variations between and within schools for the school effects provides an indication for either parents in choosing the optimal school or school administrators in estimating the school system.

Raudenbush and Willms apply the above variance decomposition techniques to the survey data from Scotland to estimate the school effects between and within schools. They conclude that determining those effective schools with these unbiased school effects requires randomly assigning students with similar backgrounds to the different treatments.

Another paper from Raudenbush and Bryk (1986) applies the hierarchical linear model (HLM) to estimate school effects for a longitudinal study. As their paper quoted from other early papers, "Methodologists have warned that the use of traditional linear models to study multilevel phenomena can produce misleading results (p. 1)". The study utilizes the slope-as-outcomes approach. The random-coefficient regression model estimates the variability of the regression coefficients for both intercepts and slopes.

Assume the normal distribution of the dependent variable(s) and the level-one coefficient(s). Researchers frequently employ the hierarchical linear model in social and psychological researches. HLM enables the study of longitudinal data with multiple levels. The Raudenbush and Willms' paper provides three motives for employing HLM in school effects research. HLM is natural in analyzing the multiple levels data. Its analysis separates the parameter and sampling variances. The covariance matrix of the coefficients explains the relationships among the within-group coefficients.

Here is an example of a two-level HLM study. In a school district, we are interested in the individual students' test score regressed on their characteristics, such as gender, race, and social-economic status. The level-one equation estimates the within-school effects of the individual students' characteristics on test scores. Furthermore, HLM allows that the coefficients in the level-one equation vary across schools. Therefore, the between-school level shows that each coefficient in the within-school level is the dependent variable regressed on the school-level independent variables.

In reality, Raudenbush and Bryk's (1986) paper studies the Coleman et al.'s assertion that Catholic schools are more effective than public schools. They use the High School and Beyond Data (HSB) survey in running two-level HLM regressions. The model studies the SES-achievement relationship across all schools. The dependent variable is the standardized mathematics achievement score; the independent variable is the student social-economic status (SES). They report the parameter and sampling variances for the level-one parameter(s) with random effects. The ratio of the parameter variance to the observed variance indicates the reliability of the parameter. "Stated somewhat differently, the traditional measure for the adequacy of a regression model, R^2, cannot exceed [the reliability of the parameters] (p. 8)". The reliability of the level-one parameter(s) shows evidence for the necessity of random effects within the parameter(s) if the reliability is large enough.

In addition to HLM analysis for school effects research, most of the recent researchers apply the value-added assessment method to estimate school performance. Some of them even further go to estimate the teacher effects by using value-added-like analysis.

Ballou (2005, p. 272) wrote that "The central idea of value-added assessment is straightforward: educators are to be evaluated based on the progress of their students, or the difference between incoming and outgoing levels of achievements”. The Tennessee Value-Added Assessment System (TVAAS) provides an explicit example for value-added analysis. TVAAS studies the outcomes of the students in two adjacent grades. A particular student, however, scores higher than the district average at grade t, and scores much higher at grade $t+1$. The progress contributed by the school effects (or even further at the teacher level) is measured by the additional gain at both grade levels between the students’ test score and the district average.

Nevertheless, the estimation of the school/teacher effects requires being unbiased and precise. In practice, the teachers are not freely assigned to the students. When controlling for the students’ demographic and social-economic status, there is little work on correcting the bias issue. Ballou’s paper reports the strong correlations between the original and modified TVAAS teacher effects. In addition, within a year the one-time noise causes the imperfect test score with the uncertain measurement error. Ballou (2005, p. 276) observed: “The actual progress of a student will deviate from measured progress due to test error”. With only three-year teachers in the sample, the percentage of teachers significantly different from the average is close to 60% for the grade 7-8 at the 10% significant level from Ballou’s study. In addition to the significant identification for teachers, Ballou points out that the imprecision will produce the instability for the teacher effects, together with school effect, even if the true effects are stable over time.

Ladd and Walsh (2000) write a summary paper for the implementation of the value-added approach to measure school performance, especially in student achievement. The student test score is applied to measure the learning of students, and the value-added study is implemented as the policy tool for increasing the school performance. Instead of studying the levels of the student performance, the researchers focus on the achievement gains from one year to the next within one school.

The basic model in measuring the school effectiveness in Ladd and Walsh's paper is to decompose the school characteristics. The educational production function has the following variables:

1. Lagged test score;
2. Family background characteristics;
3. School resources;
4. Effectiveness of school's staff and administration.

As the paper mentions, the lagged test score picks up the effect of the prior schools and family characteristics. The family background, especially the income level, is difficult to collect from the census data. Instead, the student social economic status (SES) is an alternative estimate separating the receivers and non-receivers of the free reduced lunch. Other student characteristics are too unclear to be included in the analysis. For example, the race variable explains "the relative effectiveness of school serving minorities (Ladd & Walsh, 2000, p. 8)", but vaguely interprets the educational logic in student achievement. In addition, there is no significant school resources data available for the individual schools. "In general, the state maintains data on resources and spending only at the level of the school district (Ladd & Walsh, 2000, p. 9)". Therefore, the current studies associated with the value-added approach did not collect all the above variables in the analysis for the evaluation of school performance. For instance, the model does not study the effectiveness of the school when the school resources are dropped from the educational production function. In other words, the estimated school effect in the incomplete model is explained as “all the effects on student learning with each school (p. 11)”.

Regardless the effectiveness estimated from the full production model, the measurement error causes the imprecise estimate of the school effects in any value-added studies. “Measurement error can arise for many reasons, including the limited ability of statistical agencies to collect accurate information, and the deviation between the variables specified on economic theory and those collected in practice (Angrist & Krueger, 2001, p. 3)”. However, we consider the case that two students who learn the same amount of the knowledge are likely to have different test scores. The test score does not reflect the students’ knowledge and testing ability. Any events occurring on the test day will affect the test score. Other possible one-time factors sensitive to the test score are “a dog barking in the playground on the day of the test, a severe flu season, one particularly disruptive student in a class or favorable “chemistry” between a group of students and their teacher (Kane & Staiger, 2001, p. 1)”. Therefore, the test score itself has a random error adjusted by the unobservable factors.

Ladd and Walsh’s paper suggests that the measurement error issue can be solved by using the instrumental variable. For example the test score at lower-level grade is the instrument for the prior test score in the regression to predict the current test score. The comparison between the regression with or without the correction of the measurement error shows the significant changes in the estimated school effects to the upper and lower tail students. In their paper, the measurement error biases downwards to the schools serving the low-performing students and upwards to the one serving the high-performing students. Therefore, the policy implication with the inaccurate estimate will not reduce the gap between the low- and high-performing students. The true purpose of the educational accountability will not work on the school improvement properly.

In the more recent year, Todd and Wolpin (2003) present a cumulative model showing the school performance and student achievement relationship. The conceptual framework of the model considers the input from parent(s) and school(s), and is a function of the student age. The entire history of both parents' and schools' contribution affects the students' current test scores. Moreover, the model allows the student endowed capacity to pick up the effect of learning. However, "the lack of data on input histories and endowed capacity has led researchers to adopt what has been called value-added approach to estimating achievement production functions (Todd and Wolpin, 2003, p. F19)". If assuming the inputs are age-invariant and the student endowed capacity is constant, the simplified model only includes the current school input, the lagged test score, and the student's fixed effect(s).

When considering the causal effects in education, the current studies decompose the school resources within the educational production function. In general, the educational resources are teachers-pupil ratio, teacher education, teacher experience, teacher salary, expenditure per pupil, administrative inputs, and facilities. In collecting the estimated effects for these school resources, the studies show that there is no systematically consistent conclusion for any of those resources. Hanushek (1997) presents a table listing the 377 studies of the key resources affecting student performance. There is no evidence for the consistency of either statistically significant (or insignificant) positive or negative effects on the student test scores.

Webbink (2005) concludes that the existence of the inconsistent estimates is due to the endogenous bias. The endogeneity bias occurs when the intervention variable is correlated with the unobserved factor(s) in the error term. For example, the motivated parents tutor their child when the class size is small in one school. The effect of the class size is biased for the student achievement gain since the parents' motivation cannot be observed in the equation.

Webbink points out that one of the methods dealing with endogeneity is to randomly assign the students in both controlled and treatment classrooms. The most well-known examples are the project Student/Teacher Achievement Ratio (STAR) in Tennessee and the experiments with vouchers in New York, Dayton Ohio and Washington. Although the randomized experiment solves the endogeneity bias, researchers will not always exploit the random assignment because the innate structure of the education system and the cost and time-consuming issues. Another method constructs the instrumental variables (IV) to produce the unbiased estimate from the regression. IV is considered to be a variable correlated with the intervention but independent from the unobserved factors and the test score. Other methods in the recent studies are institutional rules and natural variation, which create the exogenous variation in the analysis.

Later on, the paper provides the estimates with the correction of the endogeneity bias. Webbink constructs a table presenting the stronger consistency of the estimates when dealing with the endogeneity bias crossing the board for the school resources analysis. He concludes that 1) Endogeneity bias misleads the estimates; 2) The estimates of school resources tend to be consistent when exploiting exogenous variation; and 3) Achievement gain for the students varies with the school resources and incentives.

Regardless the endogeneity bias, the school accountability system based on test scores is contaminated by the sampling variance. The small number of students in the elementary grade levels produces a large variability in the performance by years. Kane and Staiger (March 2001) employ the variance component and empirical Bayesian "filtered" prediction to decompose the sampling variance and the variance between (and among) schools. Controlling for the student characteristics including the year dummies, the school-specific effect is the true school performance along with a noise component. The true school effect can be separated by the persistent component occurring over years and the non-persistent component, like these one-time factors occurring on the day of the test. The current persistent component is a linear function of the prior one with an error term. The authors employ Optimum Minimum Distance (OMD) method to choose the sample moment for the theoretical moment, and compute the goodness-of-fit as a measure of the objective function.

One finding from Kane and Staiger's paper is that the math score is the better indicator than reading for the school performance since the non-persistent component is larger for the reading test. Parents are more likely to provide the efforts for students' reading than to teach them math during the off-school time. Moreover, the comparison between test score levels and achievement gain shows that the gain score is a better approach for the value-added study. Lastly, the filtering approach collects the history information for the school estimates. Kane and Staiger mention that "the filtered estimates are far more accurate than more naïve estimates of school performance (p. 5)".

Motivation

As Sherman Dorn (1998) wrote that “statistical accountability systems are important because numbers are visible power in public debate (p. 2)”. The policy-makers and educators seek the accurate inference from the educational studies to improve the educational quality. However, the educational accountability system based on test scores does not provide convincing evidence for school improvement. The past and recent educational programs for increasing the students’ test score are cost-inefficient and discourage the incentives to the school personnel. Moreover, spending associated with school resources does not significantly matter to the school performance because of the implementation of the imprecise accountability for the school evaluation. Therefore, the studies of school productivity require either to correct the estimation error from the past educational studies or to establish an advanced approach to evaluate school performance.

We will think about the school-student level analysis by applying the recursive prediction method. Our goal is to seek the high-performing school(s) that students score higher within the school(s) than other students do in other schools controlling for the student learning ability. The indicator of the students endowed capacity is the CogAT test score. Both 2^{nd} and 4^{th} graders before the academic year 2005 in this median size school district take this IQ test, but since the year 2005 the 5^{th} graders will take the test instead of the 4^{th} graders. Moreover, the study of the CogAT test shows that there is no significant difference between two CogAT test scores for the individual students.

With the school performance accountable, the scenario is that the high-performing school raises the student standardized test score above the mean or at least higher than the students in the low-performing school holding the students’ characteristics and the endowed capacity constant. Moreover, the school effects vary slightly across years. There is no huge change for the school ranking since the test score does not dramatically fluctuate by years. Each individual school shares the same fixed effect with the school district, but separately, schools have their own specific effects to the student learning within a year. Therefore, conditional on the student characteristics and the learning ability, we will know the increase of the standardized score contributed by schools.

The students took the test at both incoming and outgoing years, respectively. After adjusting the measurement error in test scores, we expect to see that identical students gain more when learning at the more productive school and less at the less productive one. Within the school, there is no classroom effect, and assume that the students benefit or lose the same amount of test score from the school. Nevertheless, the best and poorest students might be the exception because the best will learn nothing or the poorest cannot learn anything in schools. I suggest that equalizing the school productivity, at least in the district level, will shrink the gap across students for whichever school they attend. The identification of the school productivity is necessary for the school principals and administrators to go further onto rewarding and sanctioning the school personnel, and to shift the school resources for the equity of the education.

Data

The median size school district provides the student test scores and demographic data from their administration data base. The student administrative data contain the student characteristics, including schools attended, mobility status, retained information, sex, race, Free/Reduced lunch, student gifted status, and English learner conditions. The test data consists of Missouri Assessment Program (MAP) Test, TerraNova (TN) Test, Cognitive Abilities Test (CogAT), GATE, Scholastic Reading Inventory (SRI), Explore, PLAN, and Scholastic Aptitude Test (SAT). However, the different tests will only be taken in the assigned grade levels, and some tests switched the grade for test takers recently. The test scores are more plentiful in both 2004 and 2005. The K-12 students in the data set studied in 31 different schools within one school district (3 special schools, 19 elementary schools, 3 middle schools and 6 high schools).

I use student number, year and the grade level to link between the students' characteristics and test scores. The link is nearly perfect. I have the students who were retained at the same grade taking the test more than one time, but the last test score will be kept only for the analysis. Moreover, the race and gender in the most recent year's data are changed for uncertain reason(s). I will assign the most recent value showing in the gender and race variables to update the data in prior years.

I do not have any tests taken by students in more than two continuous years. In the analysis, I will use the standardized test scores (z-scores) from the different tests within the same subject to assign the student's current and prior test scores for the value-added study. The CogAT score is divided by 100 to be consistent with other z-scores. Therefore, the sample contains the observations who had all the current, prior and CogAT test scores. A few students who did not have prior score were also included with the test dummy set to one. There are 5017 student records in all years, roughly 3000 observations for the students at 5th grade regressed on 4th grade and 2000 for ones at 4th grade regressed on 3rd grade.

Methodology and Theory

Basic Value-Added Model

The model is well used by the value-added school effects analysis, even in study of teacher effects. The school resources are excluded for the estimate of the overall school productivity because of the data limitation. Instead, the school dummy variables by years are the key components of this type of analysis. The value-added model sets up as follows:

$$S_t^{ij} \sim Z_t^j + S_{t-1}^{ij} + A_t^i + L_t^i + C^i + \varepsilon$$

where S_t^{ij} and S_{t-1}^{ij} are the current and prior test scores for the student i at school j in year t, respectively; Z_t^j is the school specific fixed effect at school j in year t; A_t^i is the endowed capacity for the student i; L_t^i is Free/Reduced Lunch (F/RL) dummy; and C^i is the i^{th} student demographic status. The error term has the property of normality, $N \sim (0, \sigma^2 \mathrm{I})$. The model fits the value-added analysis, using the student achievement gain to estimate the school productivity between schools.

Following the cumulative educational production function in Todd and Wolpin's paper in 2003, I employ the CogAT test score as the indicator of the student endowed capacity. Todd and Wolpin's paper only presents the student current test score as the function of the student prior test score and other student's characteristics. However, I see the most recent value-added model includes the student previous one year test score and his/her cognitive test score as the kernel components to predict the current test performance. We will see that the one year prior test and student ability score explain the most variation of the student current performance in the basic OLS regression model.

Hierarchical Linear Model

Another famous modeling structure is the two-level HLM for the school-student level data. The first student level sets up just like the basic OLS regression model. At the school level, the model allows the random effects to adjust the measurement error occurring at the test scores. Moreover, the school-level variables are included to estimate the effects between schools.

The within-school equation will have the coefficients for test scores, Free/Reduced Lunch, and the dummies of gender, race and school for the individual students:

Level-1 Model (with the last school dummy dropped):

$$\begin{aligned} \mathrm{Y}_{\mathrm{ijt}} = \mathrm{B}_0 &+ \mathrm{B}_1\text{*(CogAT)} + \mathrm{B}_2\text{*(Score}_{\mathrm{t-1}}) + \mathrm{B}_3\text{*(F/RL)} + \mathrm{B}_4\text{*(Female)} \\ &+ \mathrm{B}_5\text{*(Scodum)} + \mathrm{B}_6\text{*(White)} + \mathrm{B}_7\text{*(Black)} + \mathrm{B}_8\text{*(Asian)} \\ &+ \cdots + \mathrm{B}_{\mathrm{ijt}}\text{*(SCH}_{\mathrm{jt}}) + \cdots + \mathrm{R}_{\mathrm{ijt}} \end{aligned}$$

where $\mathrm{Y}_{\mathrm{ijt}}$ is the current math score for the student i in school j in year t, B_{j0} is the mean math achievement gain for the school j, $\mathrm{B}_1 \cdots \mathrm{B}_8$ are the test score related coefficients, $\mathrm{B}_{\mathrm{ijt}}$ is the school dummy for student i in school j in year t, and $\mathrm{R}_{\mathrm{ijt}}$ is the error of estimate for student i in school j in year t. Within the student control variables, the receiver of the lunch is coded as 1, and otherwise F/RL = 0; gender is coded to be 1 for female, and 0 = male; the score dummy is 1 when the student does not have the prior test score, and 0 for otherwise; the race variable is set to be 1 if the student has the race in the corresponding race variable, and 0 for others; the school dummy is 1 when the student i attended in school j, and 0 in any other cases.

Level-2 Model (for the level-one parameters with the random effect and with the school-level regressor only):

$$\mathrm{B}_1 = \mathrm{G}_{10} + \mathrm{U}_1,$$

$$\mathrm{B}_2 = \mathrm{G}_{20} + \mathrm{U}_2,$$

$$\mathrm{B}_3 = \mathrm{G}_{30} + \mathrm{G}_{31}{*}(\%\mathrm{F/RL}),$$

$$\mathrm{B}_4 = \mathrm{G}_{40} + \mathrm{G}_{41}{*}(\%\mathrm{Female}),$$

where G_{10} and G_{20} are the ground mean for the CogAT and prior Math test score across all schools, respectively; U_1 and U_2 are the random effects of school j on both the CogAT and prior test score associated with the current math score; $\%\mathrm{FRL}$ and $\%\mathrm{FEMALE}$ are computed by schools and by years. The full model in the mixed form will be provided in the Appendix I.

Kalman Filter Algorithm Method

Kalman Filter algorithm consists of the basic value-added model and a dynamic equation of school effect. This method was well-established by Hamilton (1994). The first stage of Kalman Filter algorithm is the measurement equation:

$$\underset{(I_t x1)}{S_t^j} = \underset{(I_t xJ)}{A_t^j}\ \underset{(Jx1)}{z_t^j} + \underset{(I_t xk)}{M_t}\ \underset{(kx1)}{\beta} + \underset{(I_t x1)}{\varepsilon_t},$$

where I_t is the number of the students in year t, and J is the number of the schools. S_t^j is the test score by students at school j in year t, and M_t is the matrix of the prior test score, CogAT score and students' characteristics. A_t^j represents the school dummy matrix in which the element, a_{ijt}, is one when i^{th} student is in the j^{th} school in year t and zero in any other cases. β is the covariance vector assumed to be the same across all the years and schools. z_t^j is the school productivity measurement at school j in year t. The residual, ε_t, is the Gaussian White noise with mean zero and variance-covariance $\underset{(I_t x I_t)}{\sigma^2 \mathrm{I}_t}$.

The second stage is the state equation:

$$\underset{(Jx1)}{z_t^j} = \eta \underset{(Jx1)}{z_{t-1}^j} + \underset{(Jx1)}{\upsilon_t}$$

where η is the coefficient of the lag operator. If $|\eta| < 1$, there is a covariance-stationary process for the school productivity. If $|\eta| \geq 1$, the school productivity accumulates rather than dies out over time. The school productivity is a random work when $\eta = 1$. In addition, η will be the same crossing all years. The residual, υ_t, is the Gaussian White noise with mean zero and variance-covariance $\underset{(JxJ)}{\Omega}$. The school productivity is the dynamic consequences of events over time.

The Kalman Filter method updates and smoothes the school productivity based on the initial parameters, such as β, σ^2, η, and Ω. We use the Bayesian method to compute the posterior distribution of these parameters based on the estimated school productivity from the current iteration. The posterior distributions are computed by

$$\pi(\theta_1, \theta_2 \mid y) = \pi(\theta_1)\pi(\theta_2) f(y \mid \theta_1, \theta_2),$$

$$\pi(\theta_1 \mid y, \theta_2) = \pi(\theta_1) f(y \mid \theta_1, \theta_2),$$

$$\pi(\theta_2 \mid \theta_1, y) = \pi(\theta_2) f(y \mid \theta_1, \theta_2),$$

where $\pi(\theta_1)$ and $\pi(\theta_2)$ are independent. The posterior distribution of the school productivity can be derived from its prior distribution. This method was well proved by many empirical analyses. In my case, there are four sets of parameters, $\theta = (\beta, \sigma^2, \eta, \Omega)$, and the distribution algorithm is provided in the Appendix II.[3] Once we know the distribution of the unknown parameters, the random draw constructs a large number of the data for the individual parameters. Based on the large number theory, the mean of the drawn values converges to the true value of the real parameter. Within the Bayesian method, I use Gibbs sampler algorithm to compute the parameters by the consequence that provides the quickest computation.

Here is the summary of the Kalman Filter and Bayesian iterations. First of all, I set up the initial values for the covariance parameters, such as β, σ^2, η and Ω, and the parameters in the prior distributions, such as $\sigma_0^2, \Xi, m, \upsilon_0^2, \tilde{\eta}, \kappa_0, q_0$ and Λ_0. Using these initial values, I update the school productivity of schools by years. Next, to filter out the noise, I smooth the school productivity back to the starting year $t = 1$ from $T - 1$. The idea is to use all the previous information to project the current value; therefore I start smoothing back at $T - 1$ and end at $t = 1$. In the Kalman Filter algorithm, the procedure starts with the initial value in year one conditional on year zero. The attempts are to obtain the predicted value based on the estimates of the previous year. The filter updates the value to the last possible year, T, in the data. Therefore, the value of these estimates is overall conditional on the assigned parameters and initial values. Smoothing the school productivity backwards using the estimates from the updated values generates one set of

[3] Dr. Shawn Ni taught and helped build up the algorithm from the book "Time Series Analysis" written by Hamilton (1994).

school effects by years. I kept the set of school productivity in the further Bayesian distribution calculation and Gibbs sampler algorithm. At the third step, with the posterior distribution setting, I compute all the assigned parameters, and update these parameters for the next iteration. The last iteration step is to compute the mean of the iterated values from the large number of iterations based on the Kalman Filter smoothing and on the random draws of the posterior distributions. The mean of the estimates of the last thousands iterations will be the true value for school productivity because the beginning iterations have the large diversion from the true estimates. I keep the iterated values when the generated values have the small and stable fluctuation. The mean of other parameters in the measurement model are also provided in the analysis of this paper.

There is a measurement error issue involved within the test scores. The true scores are observed with an error. Before each loop for simulating the school effects and other unknown parameters, a Bayesian treatment of measurement errors is applied to correct the errors for the prior year test and CogAT scores (See Appendix III for the full computation method). Zellner (1971) describes the n-mean problem related to the "errors-in-the variables". Consider n independent observations drawn from n normal distributions with different means, $X'\,(x_1, x_2, \cdots, x_n)$, and the same variance, σ_τ^2. With the σ_τ^2 known, we could make inferences about the elements of X. Applying Zellner's method (1971), the true score, x_i, is a simple linear regression on the observed score, x_i^*, with the unobserved factors, τ. The true test score is defined as follows:

$$x_i^* = x_i + \tau,$$

where x_i^* and x_i are the test score with and without measurement error, respectively; $i = 1$ or 2 for the prior and CogAT test scores, respectively; and $\tau \sim N(0, \sigma_\tau^2)$. Using the simulated parameters in the previous loop, I will compute the true scores conditional on the overall available information. The true score will have the following posterior distribution when $i=1$, for instance:

$$\pi(x_1 \mid \mathrm{D}, \mathrm{Z}, \mathrm{b}, \sigma^2, \sigma_\tau^2) \propto \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\sum_{j=1}^{I_t}\left[\frac{(u_{jt}-x_{j1t}b_1)^2}{\sigma^2}+\frac{(x_{j1t}-x_{j1t}^*)^2}{\sigma_\tau^2}\right]\right\},$$

where $u_{jt}=y_{jt}-\sum_{k=2}^{m}b_k x_{jkt}$. I will compute the true CogAT score, and then calculate the prior math score. The two true scores will be replaced in the simulation data set before the next iteration occurs.

Other computation issues will be provided in the Appendix IV. SAS allows to generate random numbers from a specific distribution, excluding Inv-gamma and Inv-wishart related to the random draws in the KF iteration. The alternative method to generate θ~Inv-gamma(α,β) is to compute the inverse value randomly generated from Gamma(α,$\beta=1$) and times the scalar. The Wishart is a matrix-variate random variable and ultimately requires the draws from the standard normal distribution. For θ~Inv-wishart($T+q_0,\Lambda$), I apply the cholesky decomposition to obtain the lower triangle matrix multiplied by the random vector, $z_{ii}\sim N(0,I)$. The estimated Λ will be the inverse of the sum of the above multiplications with at least $T+q_0$ times. In order to shorten the computation time for each loop, I use an updating formula to simplify the matrix inverse with several steps computed before the KF iteration. In addition, the initial setup and the theory behind it are attached in the Appendix V and IV, respectively.

Results

Traditional OLS and HLM analysis

Different from other value-added analysis for school productivity, my model includes the student endowed capacity measured as the CogAT test score. With the CogAT score, the overall model is able to explain 71.9% of the variation, but only 64.5% variation for the regression without the CogAT score. The summary (Table 1) of OLS regressions presents the estimates of the covariance variables and the number of the significant school effects. The Free/Reduced Lunch, gender, and race variables are consistent in the column 1, 7 and 8. A student who receives Free/Reduced Lunch and female students will score significantly worse on the Math test. White and black students will score much less than Asian but not statistically significant. There is a positive effect on the prior Math and CogAT test. One standard deviation increase in the prior Math and CogAT test scores will raise 0.4 and 0.46 standard deviation, respectively, to the current math score. The highest proportion of the school effects that are significant is 62.3% in the column 8 with one school in one year dropped due to the colinearity of the model.

Table 2 presents the detail estimation of the school effects in the unconstrained, constrained and error-in-variance adjustment OLS regression results. I construct a constrained regression by summing all school effects equal to zero. In Column 2, the change across years for each school is huge, and 14 out of 54 schools by years are significant at $p < 0.05$.

With the measurement error in mind, I rerun the regression by assigning the reliability, equal to 0.9, for the prior and CogAT test scores. I have 33 out of 53 schools by years (the last school dummy dropped from the regression) that show statistical significance at $p < 0.05$. However, the downside is that I end up with the big negative intercept and all positive school effects.

In the multilevel analysis, the summary (Table 3) presents the fixed effects of covariance variables in both the level-one and two HLM analysis, and the random effect associated with the level-one variables. The effects of the covariance variables are as similar as the OLS regression results. The proportion of the significant school effects are more than the proportion in the OLS regression. The nonlinear feature of two-level HLM reports the reliability of the random effects for the level-one variables. The specific level-two variables, like %F/R Lunch and % Female within schools, capture the effects between schools.

I assign the random effects to both the prior and CogAT test scores, and provide the %F/R Lunch and %Female for the school-level analysis. The reliability for both the prior and CogAT test scores is 0.602 and 0.626, respectively. Table 4 shows HLM results for both test scores. I have two random effects at level-2 in the model. Both variances are closed to each other in the variance-covariance matrix. However, the negative covariance means that the higher than average values of the CogAT score tend to be paired with lower than average values of prior math score. The proportion of either component is the percentage of level-2 variance in each slope. I have 24.89% of level-2 variance is explained by the variance of the CogAT score, and 30.16% by prior math score. There is still a lot of variance in the outcomes unexplained by the model due to the small proportion of the overall level-2 variance components. The percentage of level-2 variance components is only 5.78% with statistically significant variability for both the prior and CogAT test scores.

Kalman Filter and Bayesian Simulation

The prior and initial values for the unknown parameters are important to the results simulated from the KF algorithm. The school effects vary with the subjective understanding for the model. Assume the school productivity is invariant across years and the measurement error for test scores is less than 0.04 standard deviation. In addition, to check the stability of the simulation numbers, I plot the simulated school effects for each school by years. Figure 1 shows that the simulated values for school effects have the normality properties after the large number of the random draws. The starting 3000 simulation are deleted before computing the mean and variance for the school effects. The effects of the school by years are smoother than the results from OLS and HLM regressions. Figure 2 compares the results generated from the OLS, HLM and KF methods. When thinking that KF generates the true value for each school, OLS results understate the school effects by having lots of larger negative values, and HLM always overstates with all positive effects. For now, the variance of the measurement equation is smaller than the level-1 equation in HLM. The predicted current test score is correlated with the true current test score at 85%. In addition, the standard error for school effects is much smaller than the one in OLS and HLM.

The comparison between the regression and simulation results is the most interesting topic to discuss. We strongly believe that Bayesian method is more accurate to estimate the school effects than any other method. The current result shows from the 3000 simulations that school productivity is worse off to the students in our data. Most of the school effects are negative, and vary from -0.2 to +0.3 standard deviation. The parameters of the control variables are the similar as the results from OLS, HLM and the iterations. As the number of the simulations increases, the school productivity will be stable within a small range, and the simulated numbers will have the normal distribution for each school in each year.

School effects estimated from the Kalman Filter and Bayesian methods allow for the adjustment of the measurement error on the test scores, the large number generation on the small sample estimation, and the filter of the smooth school productivity. However, the estimation of school effects is not precise until the set of the initial and prior values best fits the school data. It is rarely likely to have one set of initial and prior values be better than any others. Knowledge of the school district does not guarantee this setting issue. The estimate results depend on the attitude the researchers to the overall knowledge of the school district.

Furthermore, any tiny change of the individual initial or prior values will alter the estimate results. Some initial or prior changes may cause that the simulation fails to converge after generating a large number of times. For example, when assigning $\hat{P}_{1|0} = 2$, the scatter points for school effects keep going upwards, and do not turn backwards. Moreover, when m_0 is larger than 1500 and σ_{00}^2 is smaller than 0.001, the school effects keep going downwards with the tiny changes. In this case, the covariance variables are completely inconsistent with the results in OLS and two-level HLM.

The flexibility of choosing proper initial and prior values allow me to lower the variance of the residual in the measurement equation. With the much smaller σ^2, some of the covariance variables will not show the expected sign. The lunch participants is assumed to have the negative effects to the predicted math test score in the common sense. When estimated $\sigma^2 = 0.152$ holding other values constant, the estimate of Free/Reduced Lunch is 0.0012 with 0.0014 standard deviation. In addition, the correlation between the outcomes and predicted outcomes does not change due to the small variance of the residuals. The researcher concludes that the unobservable factor(s) does(do) exist in the residual term. The current model with only the prior and CogAT scores as the key predictors needs to be modified further. The precise estimates of school effects will not be achieved without other factor(s) taken into account.

Implication for policy

The empirical study shows that the smart students are easy to teach and get progressed, but the poor ones are not. The estimates without including the students' initial ability will understate the school productivity. Moreover, the extra learning hours after school contribute to the achievement gain to make the estimates imprecise. Assume the model reduces these unobservable factors that raise the achievement gain for students. The estimates will be useful to narrow the gaps across schools. Furthermore, the unequal education across schools and classrooms will disappear when considering inputting the same school resources.

The implication for American K-12 educational policy is to enforce the high and equal productivity to the schools at the district level, and to manipulate the education inputs for the schools to raise the student performance. The redistribution of the school resources will benefit overall K-12 students; at least, those students in each school at the district level have the same achievement progress. In addition, school administrators could well accomplish “No Child Left Behind”. In the future, there are no conceptual good or poor schools based on the student achievement point of view.

Assume that poor initial ability is not the reason that these poor students always score lower. The policy might reward the schools who raise the poor students’ test score comparative to the regular students. The extra payment for these poor students will be the future goal to raise the overall student performance. These poor students who hid behind the average score will contribute to the further improvement of school productivity. The general picture is to raise the American student academic performance and the student learning capability and to establish equal and productive schools for students to accumulate their human capital. In this way, we will see the truth that there is no child left behind.

Conclusion

The result will be closer to the true value of school productivity. However, sample selection, model comparison and the prior value setup remain issues. The attempt of this paper is to use alternative statistical mechanisms to establish the advanced measurement model. The comparison between the estimates from other methodologies will shed light on reducing noise data and constructing the sophisticated school productivity model.

I suggest that American K-12 educational policy needs to focus on the allocation of the educational inputs. Investment in the current poor achievement schools will narrow the performance gaps between schools within the school district. In the future, the measurement of school productivity could be used to balance the difference across states. The assumption is that the one standard deviation increase of the student achievement at different schools will require the same educational inputs for the poor and smart students. Although the test score is different for both poor and smart students, the same school productivity ensures that the changes of student performance are kept constant. The next step is to think about the way to especially raise the poor student performance by inputting extra educational inputs.

The estimate of school productivity is to measure the school performance after a student starts his/her learning. School productivity should be easily evaluated and manipulated by the school educators and principals. The school that raises the student achievement gain more than others will take the responsibility to assist other schools in having the improvement of school productivity. I hope that the pure school productivity will not be affected by the different initial ability and extra education the students could obtain based on the family status.

Bibliography

Aaronson, Daniel, Lisa Barrow and William Sander (2003). “Teachers and Student Achievement in the Chicago Public High Schools.” Working Paper Series WP-02-28, Federal Reserve Bank of Chicago.

Angrist, J. D. and Krueger, A. B. (2001). “Instrumental Variables and the Search for Identification: from Supply and Demand to Natural Experiments.” Journal of Economic Perspectives 15(4): 69-87.

Bryk, Anthony S. and Stephen W. Raudenbush (1992). “The Logic of Hierarchical Linear Models: in Hierarchical Linear Models: Applications and Data Analysis Methods.” SAGE Publications: Newbury Park, London & New Delhi, pp. 9-30.

Ballou, Dale (2005). “Value-Added Assessment: Lessons from Tennessee.” In R. Lissitz (Ed.), Value-added Models in Education: Theory and Applications (pp. 272-303). Maple Grove, MN: JAM Press, 2005.

Hanushek, Eric A. (Summer, 1997). “Assessing the Effects of School Resources on Student Performance: An Update.” Educational Evaluation and Policy Analysis, Vol 19, No. 2. pp. 141-164.

Hamilton, James. Time Series Analysis. Princeton University Press: Princeton, New Jersey, 1994.

Kane, Thomas and Douglas Staiger (March, 2001). “Improving School Accountability Measures.” NBER Working Paper 8156.

Ladd, Helen F. and Randall P. Walsh (2001). “Implementing Value-Added Measurers of School Effectiveness: Getting the Incentives Right.” Forthcoming in Economics of Education Review.

Rivkin, Steven G., Eric A. Hanushek and John F. Kain (March, 2005). “Teachers, Schools, and Academic Achievement.” Econometrica, Vol. 73, No. 2, 417-458.

Raudenbush, Stephen W. and Anthony S. Bryk (Jan., 1986). “A Hierarchical Model for Studying School Effects.” Sociology of Education, Vol. 59, No. 1. pp. 1-17.

Raudenbush, Stephen W. and J. Douglas Willms (Winter, 1995). “The Estimation of School Effects.” Journal of Educational and Behavioral Statistics, Vol. 20, No. 4. pp. 307-335.

Sherman, Dorn (1998). “The Political Legacy of School Accountability Systems.” Education Policy Analysis Archives, Vol 6.1 ISSN 1068-2341.

Todd, Petra E., and Kenneth I. Wolpin (2003). “On the Specification and Estimation of the Production Function for Cognitive Achievement”, The Economics Journal II3: F3-F33.

Webbink, H. Dinand. "Causal Effects in Education". Journal of Economic Surveys, Vol. 19, No. 4, pp. 535-560, September 2005 Available at SSRN: http://ssrn.com/abstract=785436

Zellner, Arnold. An Introduction to Bayesian Inference in Econometrics. New York: Wiley, 1971.

Appendix I – HLM Full Model in the Mixed Form

The two-level HLM analysis for school productivity study in the mixed form:

$$(\text{Score}_t)_{ij} = \gamma_0 + \gamma_1\text{*(CogAT)}_{ij} + \gamma_2\text{*(Score}_{t-1})_{ij} + \gamma_3\text{*(F/RL)}_{ij} + \gamma_4\text{*(Female)}_{ij}$$

$$+ \gamma_5\text{*(Scodum)}_{ij} + \gamma_6\text{*(White)}_{ij} + \gamma_7\text{*(Black)}_{ij} + \gamma_8\text{*(Asian)}_{ij}$$

$$+ \cdots + \gamma_{ijt}\text{*(SCH)}_{jt} + \cdots + \gamma_{31}\text{*(\%F/RL)}_j\text{*(F/RL)}_{ij} + \gamma_{41}\text{*(\%Female)}_j\text{*(Female)}_{ij}$$

$$+ u_1\text{*(CogAT)}_{ij} + u_2\text{*(Score}_{t-1})_{ij} + r_{ij}$$

where the school dummies that are equal to 0 are excluded. The mixed model uses the level-two parameters, γ_*, instead of the level-one parameters, B_*.

Appendix II – Kalman Filter Algorithm & Bayesian Computation

The model estimates school effects of the students' Math performance over the course of the academic year. The sample period is $t = 1, \cdots, T$.

In year t, each of the observed students, $i = 1, \cdots, I_t$, is assigned to one of the J schools. The score vector is $S_t = (S_{1t}, \cdots, S_{It})'$. The school effect is $z_t = (z_{1t}, \cdots, z_{Jt})'$. The covariate matrix is $x_t = (x_{1t}, \cdots, x_{It})'$, with k covariates for each student per year (scores of the previous year, race, gender, lunch status, etc). Let A_t be an $I_t xJ$ matrix. The i^{th} row and jth column of A_t is one if the i^{th} student is in the j^{th} school in year t and is 0 otherwise. Let k – vector β be the parameter of the covariate effect. The measurement model is:

$$S_t = A_t z_t + x_t \beta + \varepsilon_t . \tag{1}$$

The state equation is:

$$z_t = \eta z_{t-1} + \upsilon_t .$$

The stacked error terms are $\varepsilon_t \sim N(0, \sigma^2 \mathrm{I})$ and $\upsilon_t \sim N(0, \Omega)$. If the covariance matrix Ω shows strong cross-school correlation, then it may be indicative of district effect. The model is designed to deal with the problem of large noise in the observations (i.e., large Σ). The structure of the state variable, if correctly reflecting the nature of data-generating process, helps extract the signal from the noisy data. The state z_t can be viewed as a generalization of fixed effect (if $\Omega = 0$ and $\eta = 1$) and random effect (if $\eta = 0$).

We define

$$\begin{aligned}
\hat{y}_{t|t-1} &= \hat{E}(y_t \mid t-1), \\
\hat{z}_{t|t-1} &= \hat{E}(z_t \mid t-1), \\
\hat{z}_{t|t} &= \hat{E}(z_t \mid t), \\
\hat{\Sigma}_{t|t-1} &= E[(y_t - y_{t|t-1})(y_t - y_{t|t-1})' \mid t], \\
\hat{P}_{t|t-1} &= E[(z_t - z_{t|t-1})(z_t - z_{t|t-1})' \mid t-1],
\end{aligned}$$

where " $\hat{E}(\cdot \mid t)$ " corresponds to the linear projection on $y_t, y_{t-1}, \cdots$ and " $E(\cdot \mid t)$ " corresponds to expectation conditional on $y_t, y_{t-1}, \cdots$. For jointly normal-distributed variables, " $\hat{E}(\cdot \mid t)$ " and " $E(\cdot \mid t)$ " are equivalent.

Let the prior be

$$\pi(\theta) = \pi(\beta)\pi(\sigma^2)\pi(\eta)\pi(\Omega). \tag{2}$$

We assume $\pi(\beta) \sim N(\beta_0, \Xi_0), \pi(\sigma^2) \sim IG(m_0, \upsilon_0^2), \pi(\eta) \sim N(\eta_0, \kappa_0^2),$ and $\pi(\Omega) \sim IW(q_0, \Lambda_0).$

The posterior is

$$\pi(\theta, Z \mid D) \propto \pi(\theta, Z) f(D \mid \theta, Z). \tag{3}$$

The initial state and $(z_0, P_{1|0})$ are given.

Let

$$Y = \begin{pmatrix} s_1 - (A_1 z_1) \\ \vdots \\ s_T - (A_T z_T) \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ \vdots \\ x_T \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{pmatrix}. \tag{4}$$

Then we rewrite (1) as

$$Y = X\beta + \varepsilon. \tag{5}$$

Denote the total number of students -- $N = \Sigma_{t=1}^{T} I_t$, and $y_t = s_t - A_t z_t$. The posterior is

$$\pi(\beta,\sigma^2 \mid D, Z)$$
$$\propto \frac{1}{\sigma^N}\exp\{-\frac{1}{2\sigma^2}(Y-X\beta)'(Y-X\beta)\}\exp\{-\frac{1}{2}(\beta-\beta_0)'\Xi_0^{-1}(\beta-\beta_0)\}\frac{1}{\sigma^{2(m+1)}}\exp(-\frac{\upsilon_0^2}{2\sigma^2}).$$

Let $\beta_M=(X'X)^{-1}X'Y=(\sum_{t=1}^T x_t'x_t)^{-1}(\sum_{t=1}^T x_t'y_t).$ The conditional likelihood is

$$\pi(\beta,\sigma^2 \mid D, Z)$$
$$\propto \frac{1}{\sigma^N}\exp\{-\frac{1}{2\sigma^2}(\beta-\beta_M)'(\beta-\beta_M)\}\exp\{-\frac{1}{2}(\beta-\beta_0)'\Xi_0^{-1}(\beta-\beta_0)\}\frac{1}{\sigma^{2(m_0+1)}}\exp(-\frac{\upsilon_0^2}{2\sigma^2})$$
$$\propto \frac{1}{\sigma^{N+2(m_0+1)}}\exp\{-\frac{S^2+\upsilon_0^2+(\beta-\beta_M)'X'X(\beta-\beta_M)}{2\sigma^2}\}\exp\{-\frac{1}{2}(\beta-\beta_0)'\Xi_0^{-1}(\beta-\beta_0)\},$$

where $S^2=(Y-X\beta_M)'(Y-X\beta_M)=\Sigma_{t=1}^T\Sigma_{i=1}^{I_t}(y_{it}-x_{it}\beta_M)^2.$

Regarding the parameters in the state equation, given Z, the posterior is

$$\pi(\eta,\Omega_t \mid Z)$$
$$\propto |\Omega|^{-\frac{T}{2}}\exp\{-\frac{\sum_{t=1}^T(z_t-\eta z_{t-1})\Omega^{-1}(z_t-\eta z_{t-1})}{2}\}\exp\{-\frac{(\eta-\eta_0)^2}{2\kappa_0^2}\}|\Omega|^{-\frac{q_0+J+1}{2}}\exp\{-\frac{1}{2}\mathrm{tr}(\Omega^{-1}\Lambda_0)\}.$$

We are interested in estimating parameters $\theta=(\beta,\sigma^2,\eta,\Omega)$, and the latent state $z_t=(z_{1t},\cdots,z_{Jt})'$ given the observed data $D=\{(s_t,x_t),t=1,\cdots,T\}.$ The posterior "$\pi(\theta,Z \mid D)$" can be derived in the following steps.

The posterior "$\pi(\theta,Z \mid D)$" can be derived in the following Markov Chain Monte Carlo (MCMC) Gibbs sampling algorithm: suppose before the kth iteration we have $\theta^{(k-1)},\hat{z}_0,\hat{P}_{1|0},D.$

(A) (Kalman Filter) Given $\theta^{(k-1)},\hat{z}_0,\hat{P}_{1|0},D,$ simulate $Z^{(k)} \mid \theta^{(k-1)},\hat{z}_{1|0},\hat{P}_{1|0},D,$ where $\hat{z}_{1|0}=\eta\hat{z}_0.$ There are two steps in part (A).

(A1) (Updating) Obtain $z_t \mid \theta^{(k-1)},\hat{z}_{1|0},\hat{P}_{1|0},D^t.$ $D^t=\{(s_t,x_t),t=1,\cdots,T\}.$ For $t=1,\cdots,T$, given $\theta^{(k-1)},z_{t|t-1},P_{t|t-1},D^t,$ compute

$$
\begin{aligned}
\hat{\Sigma}_{t|t-1} &= (\sigma^2)^{(k-1)} I_t + A_t \hat{P}_{t|t-1} A_t', \\
\hat{z}_{t|t} &= \hat{z}_{t|t-1} + \hat{P}_{t|t-1} A_t' \hat{\Sigma}_{t|t-1}^{-1} (S_t - A_t \hat{z}_{t|t-1} - x_t \beta^{(k-1)}), \\
\hat{P}_{t|t} &= \hat{P}_{t|t-1} - \hat{P}_{t|t-1} A_t' \hat{\Sigma}_{t|t-1}^{-1} A_t \hat{P}_{t|t-1}, \\
\hat{z}_{t+1|t} &= \eta^{(k-1)} \hat{z}_{t|t}, \\
\hat{P}_{t+1|t} &= (\eta^{(k-1)})^2 \hat{P}_{t|t} + \Omega^{(k-1)}.
\end{aligned}
$$

To sum up, the filter updates $(\hat{z}_{t|t}, \hat{P}_{t|t}, \hat{\Sigma}_{t|t-1})$. The autoregressive structure of the state equation gives the prediction $(\hat{z}_{t+1|t}, \hat{P}_{t+1|t})$, and completes the recursive loop. When we move to $t = T$, all the conditional moments are functions of $\theta^{(k-1)}$, initial state $\hat{z}_{1|0}$ and $\hat{P}_{1|0}$.

(A2) (Smoothing) Obtain $z_t \mid \theta^{(k-1)}, \hat{z}_0, \hat{P}_{1|0}, D.$ From (A1) we draw z_T from $N(\hat{z}_{T|T}, \hat{P}_{T|T})$. For $t = T-1, \cdots, 1$, given $\theta^{(k-1)}, z_{t+1},$ and $D,$ compute

$$
\begin{aligned}
\tilde{z}_t &= \hat{z}_{t|t} + \eta^{(k-1)} \hat{P}_{t|t} \hat{P}_{t+1|t}^{-1} (z_{t+1} - \hat{z}_{t+1|t}), \\
\tilde{P}_t &= \hat{P}_{t|t} - (\eta^{(k-1)})^2 \hat{P}_{t|t} \hat{P}_{t+1|t} \hat{P}_{t|t},
\end{aligned}
$$

where $T = t+1$.

Draw $z_t^{(k)} \mid (\theta^{(k-1)}, \hat{z}_0, \hat{P}_{1|0}, D) \sim N(\tilde{z}_t, \tilde{P}_t)$. The ending result is the drawing data set of $Z^{(k)} = (z_1^{(k)}, \cdots, z_T^{(k)})'$.

(B) Draw from the conditional posteriors "$\pi(\theta^{(k)} \mid Z^{(k)}, D)$". The part (B) is done via Gibbs sampling algorithm with the vector parameter θ drawn by sub-vectors, each from the standard distributions.

(B1) Draw $\beta^{(k)} \mid ((\sigma^2)^{(k-1)}, Y^{(k)}, X)$ from the conditional posterior multivariate normal distribution

$$
\beta^{(k)} \mid ((\sigma^2)^{(k-1)}, Y^{(k)}, X) \sim N_{(k)}(\mu, \Xi),
$$

where

$$
\begin{aligned}
\mu &= \Xi((\sigma^{-2})^{(k-1)} (X'X) \beta_M^{(k)} + \Xi_0^{-1} \beta_0); \\
\Xi &= \{(\sigma^{-2})^{(k-1)} (X'X) + \Xi_0^{-1}\}^{-1}.
\end{aligned}
$$

(B2) Draw $(\sigma^2)^{(k)} \mid (\beta^{(k)}, Y^{(k)}, X)$ from the conditional posterior Inv-Gamma distribution

$$\frac{1}{\sigma^{N+2(m_0+1)}}\exp\{-\frac{S^2+\upsilon_0^2+(\beta^{(k)}-\beta_M^{(k)})'X'X(\beta^{(k)}-\beta_M^{(k)})}{2\sigma^2}\}$$
$$\propto IG(m_0+\frac{N}{2},\upsilon_0^2+\frac{S^2+(\beta^{(k)}-\beta_M^{(k)})'X'X(\beta^{(k)}-\beta_M^{(k)})}{2})$$
$$=IG(m_0+\frac{N}{2},\upsilon_0^2+\frac{(Y^{(k)}-X\beta^{(k)})'(Y^{(k)}-X\beta^{(k)})}{2})$$

Note that $(Y^{(k)}-X\beta^{(k)})'(Y^{(k)}-X\beta^{(k)})=\Sigma_{t=1}^{T}\Sigma_{i=1}^{I_t}(y_{it}-x_{it}\beta^{(k)})^2$.

(B3) Draw $\pi(\eta^{(k)}\,|\,\Omega^{(k-1)},Z^{(k)})$ from the conditional posterior normal distribution

$$\pi(\eta\,|\,\Omega^{(k-1)},Z^{(k)})\propto\omega^{-1}\exp\{-\frac{(\eta-\tilde{\eta})^2}{2\omega^2}\}=N(\tilde{\eta},\omega^2),$$

where

$$\omega=(\sum_{t=1}^{T}(z_{t-1}^{(k)})'(\Omega^{(k-1)})^{-1}z_{t-1}^{(k)}+\frac{1}{\kappa_0^2})^{-1};\ \tilde{\eta}=\omega\{\sum_{t=1}^{T}(z_{t-1}^{(k)})'(\Omega^{(k-1)})^{-1}z_{t-1}^{(k)}+\frac{\eta_0}{\kappa_0^2}\}.$$

(B4) Draw $\pi(\Omega^{(k)}\,|\,\eta^{(k)},Z^{(k)})$ from Inverse-Wishart distribution

$$\pi(\Omega\,|\,\eta^{(k)},Z^{(k)})$$
$$\propto|\Omega|^{-\frac{T+q_0+J+1}{2}}\exp\{-\frac{1}{2}\mathrm{tr}\Omega^{-1}[\sum_{t=1}^{T}(z_t^{(k)}-\eta^{(k)}z_{t-1}^{(k)})(z_t^{(k)}-\eta^{(k)}z_{t-1}^{(k)})'+\Lambda_0]\}=IW(T+q_0,\Lambda),$$

where $\Lambda=\sum_{t=1}^{T}(z_t^{(k)}-\eta^{(k)}z_{t-1}^{(k)})(z_t^{(k)}-\eta^{(k)}z_{t-1}^{(k)})'+\Lambda_0$.

This concludes an MCMC cycle, replace k by $k+1$ and redo Steps (A1) (A2) (B1) (B2) (B3) (B4). The simulated quantities, $\theta^{(k)}$ and $Z^{(k)}$, represent a numerical distribution from the joint posterior.

Appendix III – Adjusting for Test Scores Measurement Error

Adjusting Measurement Error for the prior and CogAT test scores:

$$x_{st}^*=x_{st}+\tau_{st}$$

where x_{st}^* $(s=CogAT,\,prior\,test;t=1,\cdots,T)$ is the observed test score; x_{st} is the true test score; $\tau_{st}\sim N(0,\sigma_{s\tau}^2)$.

The posterior PDF is

$$\pi(\mathrm{b},\sigma^2,\mathrm{x_s}\,|\,\mathrm{D},\mathrm{Z}) \propto \exp\left\{-\frac{1}{2}\left[\frac{tr(\mathrm{Y}-\mathrm{Xb})'(\mathrm{Y}-\mathrm{Xb})}{\sigma^2}+\frac{tr(\mathrm{X}-\mathrm{X}^*)'(\mathrm{X}-\mathrm{X}^*)}{\sigma_\tau^2}\right]\right\}$$

$$\exp\left\{-\frac{1}{2}(\mathrm{b}-\mathrm{b}_0)'\Xi_0^{-1}(\mathrm{b}-\mathrm{b}_0)+\frac{\upsilon^2}{\sigma^2}\right\}(\sigma_{s\tau}\sigma)^{-N-2(m+1)}.$$

The conditional posterior for $\mathrm{x_s}$ is

$$\pi(\mathrm{x_s}\,|\,\mathrm{D},\mathrm{Z},\mathrm{b},\sigma^2,\sigma_{s\tau}^2) \propto \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\sum_{j=1}^{I_t}\left[\frac{(u_{jt}-x_{jst}b_s)^2}{\sigma^2}+\frac{(x_{jst}-x_{jst}^*)^2}{\sigma_{s\tau}^2}\right]\right\}.$$

where $u_{jt}=y_{jt}-\sum_{k=2}^{m}b_k x_{jkt}$.

Computing the true control variable x_{1t} (keep the term associated with x_{1t}):

$$==> \frac{(u_{jt}-x_{jst}b_s)^2}{\sigma^2}+\frac{(x_{jst}-x_{jst}^*)^2}{\sigma_{s\tau}^2}=\frac{x_{jst}^2b_s^2-2u_{jt}x_{jst}b_s}{\sigma^2}+\frac{x_{jst}^2-2x_{jst}x_{jst}^*}{\sigma_{s\tau}^2}$$

$$=\left(\frac{b_s^2}{\sigma^2}+\frac{1}{\sigma_{s\tau}^2}\right)x_{jst}^2-2\left(\frac{u_{jt}b_s}{\sigma^2}+\frac{x_{jst}^*}{\sigma_{s\tau}^2}\right)x_{jst}$$

$$=\left(\frac{b_s^2}{\sigma^2}+\frac{1}{\sigma_{s\tau}^2}\right)\left[x_{jst}-\frac{\left(\frac{u_{jt}b_s}{\sigma^2}+\frac{x_{jst}^*}{\sigma_{s\tau}^2}\right)}{\frac{b_s^2}{\sigma^2}+\frac{1}{\sigma_{s\tau}^2}}\right]^2.$$

To one of the student records for x_{st}, we have the following distributions:

$$x_{st}\sim N\left(\frac{\left(\frac{u_{jt}b_s}{\sigma^2}+\frac{x_{jst}^*}{\sigma_{s\tau}^2}\right)}{\frac{b_s^2}{\sigma^2}+\frac{1}{\sigma_{s\tau}^2}},\frac{1}{\frac{b_s^2}{\sigma^2}+\frac{1}{\sigma_{s\tau}^2}}\right).$$

Appendix IV – SAS Programming for Computation Issues

1. Inv-Wishart Algorithm

The algorithm for computing the posterior $\Omega^{(k)}$ at (B4) goes as follows:

(a) Simulate the vector Z, which is normally distributed. Each element of the vector is independent identical distributed.

$$z_{ii} \sim N(0,1),$$
$$E(ZZ') = I.$$

(b) Compute Ω in the Inv-Wishart distribution

$$\Lambda^{-1} = L_l L_u,$$
$$x_i = L_l Z,$$
$$\Omega = \left(\sum_{i=1}^{T+q_0} x_i x_i'\right)^{-1},$$

where $L_l = L_u'$ and $i = 1, \cdots, 18$.

2. Matrix Computation(Green 3rd Edition p31):

The matrix algebra is

$$[A \pm BCB']^{-1} = A^{-1} \mp A^{-1}B[C^{-1} \pm B'A^{-1}B]^{-1}B'A^{-1}.$$

The target computation is

$$(\hat{\Sigma}_{t|t-1})^{-1} = [(\sigma^2)^{(k-1)} I_t + A_t \hat{P}_{t|t-1} A_t']^{-1}.$$

The simplification will be

$$(\hat{\Sigma}_{t|t-1})^{-1} = [1/(\sigma^2)^{(k-1)}]I_t - [1/(\sigma^2)^{(k-1)}]^2 A_t \{\hat{P}_{t|t-1}^{-1} + [1/(\sigma^2)^{(k-1)}]A_t'A_t\}^{-1} A_t'.$$

Appendix V – Table for Prior and Initial Values Setup

	Setup #1	Setup #2	Setup #3	Setup #4	Setup #5	Setup #6	Setup #7	Setup #8
$\hat{z}_0$	MAP Math Z-score assigned as initial school effects							
$\hat{P}_{1\|0}$	0.01*I(18)	0.01*I(18)	0.01*I(18)	0.01*I(18)	0.001*I(18)	0.001*I(18)	0.001*I(18)	0.005*I(18)
σ_{00}^2	0.03	0.03	0.03	0.03	0.001	0.001	0.001	0.01
m_0	10	10	10	10	1000	500	100	20
υ_0^2	0.3	0.3	0.3	0.3	3	3	3	0.1
μ_{00}	j(9,1,0.2)	j(9,1,0.2)	j(9,1,0.2)	j(9,1,0.2)	j(9,1,0.2)	j(9,1,0.2)	j(9,1,0.2)	j(9,1,0.2)
μ_0	j(9,1,0.2)	j(9,1,0.2)	j(9,1,0.2)	j(9,1,0.2)	j(9,1,0.2)	j(9,1,0.2)	j(9,1,0.2)	j(9,1,0.2)
Ξ_0	0.025*I(9)	0.025*I(9)	0.025*I(9)	0.025*I(9)	0.001*I(9)	0.001*I(9)	0.001*I(9)	0.001*I(9)
η_{00}	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
η_0	0.7	0.7	0.7	0.7	0.5	0.5	0.5	0.5
κ_0^2	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.021
Ω_{00}	0.001*I(18)	0.001*I(18)	0.001*I(18)	0.001*I(18)	0.001*I(18)	0.001*I(18)	0.001*I(18)	0.001*I(18)
Λ_0	0.1*I(18)	0.1*I(18)	0.1*I(18)	0.1*I(18)	0.01*I(18)	0.01*I(18)	0.01*I(18)	0.01*I(18)
T	3	3	3	3	3	3	3	3
q_0	127	127	26	26	37	37	77	50
$\sigma_{1\tau}^2$	0.0225	0.04	0.0625	0.0729	0.05	0.05	0.05	0.02
$\sigma_{2\tau}^2$	0.0225	0.04	0.0625	0.0729	0.05	0.05	0.05	0.02
z_{00}	J(54,1,0.5)	J(54,1,0.5)	J(54,1,0.5)	J(54,1,0.5)	J(54,1,0.5)	J(54,1,0.5)	J(54,1,0.5)	J(54,1,0.5)

Appendix VI – Theory of Prior and Initial Values Setup

This section presents the basic foci and concerns for the initial and prior setting. These initial parameters are $\hat{z}_0, \hat{P}_{1|0}, \sigma^2_{00}, \beta_{00}, \eta_{00}$, and Ω_{00}. These priors are $\beta_0, \Xi_0, m_0, \upsilon_0^2, \eta_0, \kappa_0^2, q_0$, and Λ_0. The following subsection will explain the definition of the initial (or prior) in the model, and discuss the possible interval of the magnitude for each initial (or prior).

1. School Effect, $\hat{z}_0$, in Academic Year 2001

$\hat{z}_0$ is the school effect in the academic year 2001. $\hat{z}_{1|0}$ is the forecast of the school effect in the next academic year 2002, z_1. The forecast of the school effect is

$$\hat{z}_{1|0} = E(z_1) = \eta \hat{z}_0 + \varepsilon.$$

The magnitude of $\hat{z}_0$ would be good and proper if the school effect is located between -0.3 and 0.3. According to the result from OLS regression, the school effect by school by year varies around zero within the above numeric range.

2. Mean Squared Error (MSE) Matrix, $\hat{P}_{1|0}$

$\hat{P}_{1|0}$ is the variance of school effect, z_1, in the academic year 2002 and its forecast, $\hat{z}_{1|0}$ conditional on the observed data and parameters before the year 2002. Larger values for the diagonal elements of $\hat{P}_{1|0}$ register greater uncertainty about the true value of z_1.The variance, $\hat{P}_{t|t-1}$, varies across years. In the academic year t, the MSE matrix is identified by

$$\hat{P}_{t|t-1} = E[(z_t - z_{t|t-1})(z_t - z_{t|t-1})' \mid t-1],$$

where $t-1$ denotes the observed data and parameters before the year t. The larger $\hat{P}_{t|t-1}$ shall increase the distance of the fluctuation for the school effect even though the school effect achieves the steady status after a period of simulation. In other words, the estimates of the school effect is not accurate under the larger $\hat{P}_{t|t-1}$.

3. Scalar of Variance in Measurement Equation, σ^2_{00}, and Prior Distribution of σ^2

The variance of the measurement equation shall be small. The magnitude of σ^2_{00} and σ^2 is the unique effect to the student current performance within one particular school and one year. Larger values for σ^2_{00} and σ^2 register greater unexplained variance about the current estimated student performance given the observed data and estimated parameters. We rewrite the measurement equation,

$$Y = X\beta + \varepsilon,$$

where $y_t = s_t - A_t z_t$. The MSE matrix for y_t and its forecast $y_{t|t-1}$ is

$$\begin{aligned}\hat{\Sigma}_{t|t-1} &= E[(y_t - y_{t|t-1})(y_t - y_{t|t-1})' \mid t], \\ &= (\sigma^2)^{(k-1)} I_t + A_t \hat{P}_{t|t-1} A_t'.\end{aligned}$$

The smaller $\hat{\Sigma}_{t|t-1}$, the better we reach the true estimated values for all parameters. σ_{00}^2 and σ^2 shall be positive and less than 1.

The prior σ^2 has an inverse Gamma distribution with m_0 and υ_0^2. m_0 is the shape parameter and shall be larger than 2 to have the first and second moment (mean and variance). υ_0^2 is a scalar variable. The randomly generated σ^2 in each iteration will be

$$\sigma^2 = \frac{\upsilon^2}{\Gamma(m,1)}.$$

The first moment and second moment of σ^2 are

$$\begin{aligned}E(\sigma^2) &= \frac{\upsilon^2}{m-1}, \\ \mathrm{var}(\sigma^2) &= \frac{\upsilon^4}{(m-1)^2(m-2)},\end{aligned}$$

where $m = m_0 + \dfrac{N}{2}$ and $\upsilon^2 = \upsilon_0^2 + \dfrac{(Y^{(k)} - X\beta^{(k)})'(Y^{(k)} - X\beta^{(k)})}{2}$. The m_0 and υ_0^2 shall be a pair of numbers keeping σ^2 small, and they are negatively correlated.

4. Initial vector β_{00} and Prior Distribution of β

Initial vector β_{00} follows the results from the OLS regression. The prior vector β is normally distributed with the prior mean vector β_0 and the prior variance matrix Ξ_0. Each element of the vector is independent identically distributed. The posterior vector β is simulated from its mean μ and its variance Ξ, associated with the prior mean vector β_0 and the prior variance matrix Ξ_0.

$$\mu = \Xi((\sigma^{-2})^{(k-1)}(X'X)\beta_M^{(k)} + \Xi_0^{-1}\beta_0),$$
$$\Xi = \{(\sigma^{-2})^{(k-1)}(X'X) + \Xi_0^{-1}\}^{-1}.$$

The variance Ξ is small when the prior variance Ξ_0 is small. The prior vector β_0 shall be a vector as similar as the initial β_{00}.

5. Initial η_{00} and Prior Distribution of η

Initial η_{00} is computed from the school effect across years estimated from OLS regression. Moreover, its value shall depend on the assumption to the variability of the school effects crossing years. η is the autoregressive coefficient in the state equation. The analysis of first-difference equation presents the dynamic response of the school effects. If $0 < \eta < 1$, the school effect decays geometrically towards to zero. If $\eta > 1$, the school effect increases exponentially over time. If $\eta = 1$, the school effect keeps constant.

The prior η is normally distributed with the mean η_0 and the variance κ_0^2. The variance κ_0^2 will stay small . η_0 is as similar as η_{00}, or assume that the school effect is constant across year, where $\eta_0 = 1$.

6. Initial Ω_{00} and Prior Distribution of Ω

The Initial Ω_{00} and Ω are the variance-covariance matrix of the residual in the state equation. The magnitude of Ω is the unique effect across the schools. The Initial Ω_{00} and Ω shall be small. The prior Ω has an inverse Wishart distribution with the degree freedom q_0 and the scale matrix Λ_0. The computation of the posterior iterated Ω is

$$\Lambda = \sum_{t=1}^{T}(z_t^{(k)} - \eta^{(k)}z_{t-1}^{(k)})(z_t^{(k)} - \eta^{(k)}z_{t-1}^{(k)})' + \Lambda_0,$$
$$\Lambda^{-1} = L_l L_u,$$
$$x_i = L_l Z,$$
$$\Omega = \left(\sum_{i=1}^{T+q_0} x_i x_i'\right)^{-1},$$

where Z is a vector randomly generated from the normal distribution with the mean 0 and the variance 1. The degree freedom (df) for Ω in the posterior distribution shall be greater than and equal to 18

$$df = T + q_0.$$

The posterior iterated Ω is small when both Λ_0 and df are small.

Tables and Plots

Table 1 – Summary of Ordinary Least Square (OLS) Regression Results with/without Constrained or Error-in-Variance Adjustment Conditions

	OLS Regression	Constrained OLS Regression Estimating Fixed Effects by Controlling Prior, Endowed Ability Test, Student Characteristics						OLS Regression Controlling EIV
INTRCPT	-0.116	-0.004	0.014	0.011	0.035 **	0.054 ***	0.050	-0.156 *
MAP_Math	0.398 ***	0.764 ***		0.414 ***	0.411 ***	0.407 ***	0.398 ***	0.402 ***
CogAT	0.458 ***		0.769 ***	0.466 ***	0.457 ***	0.460 ***	0.458 ***	0.536 ***
F/R Lunch	-0.059 **				-0.066 ***	-0.066 ***	-0.059 **	-0.013
Female	-0.040 **					-0.039 **	-0.040 **	-0.040 **
White	-0.002						-0.002	-0.008
Black	-0.030						-0.030	0.013
Asian	0.243 ***						0.243 ***	0.221 ***
Sco_Miss	-0.137 ***	-0.329 ***	-0.007	-0.132 ***	-0.122 ***	-0.123 ***	-0.137 ***	-0.115 **
	1/53 ***	6/54 ***	9/54 ***	5/54 ***	4/54 ***	4/54 ***	4/54 ***	10/53 ***
Sch_Dummy	2/53 **	6/54 **	2/54 **	2/54 **	3/54 **	3/54 **	2/54 **	13/53 **
	9/53 *	10/54 *	4/54 *	10/54 *	7/54 *	7/54 *	9/54 *	10/53 *
Root MSE	0.521	0.595	0.580	0.524	0.523	0.523	0.521	0.482
R-Sqr	0.719							0.760

*** = p <= 0.05**
**** = p <= 0.01**
***** = p <= 0.001**

Notes: 1. One school dummy in one year is dropped from OLS regression with/without EIV control.
2. Both the prior and CogAT scores are adjusted at 0.9 reliability.

Table 2 – School Effects of OLS Regression Results with/without Constraint and Error-in-Variance Adjustment

Parameters	OLS Estimation		Constrained OLS Estimation		OLS Estimation Contorlling EIV	
school1	0.188		0.023		0.218	**
school2	0.130		-0.036		0.140	
school3	0.142		-0.023		0.135	
school4	0.078		-0.087		0.124	
school5	0.075		-0.091		0.131	
school6	0.214		0.048		0.256	***
school7	0.178		0.013		0.210	*
school8	(dropped)		-0.165		0.033	
school9	0.100		-0.066		0.123	
school10	0.063		-0.103	*	0.102	
school11	0.130		-0.035		0.167	*
school12	0.348	**	0.183	***	0.392	***
school13	0.211		0.045		0.215	**
school14	0.276	*	0.111		0.266	**
school15	0.052		-0.114	*	0.047	
school16	0.216		0.051		0.255	**
school17	0.191		0.026		0.272	*
school18	0.193		0.028		0.243	**
school19	0.129		-0.036		0.132	
school20	0.467	***	0.302	***	0.473	***
school21	0.090		-0.075		0.089	
school22	0.272	*	0.107		0.291	***
school23	0.148		-0.018		0.190	*
school24	0.080		-0.085		0.094	
school25	0.157		-0.008		0.190	*
school26	0.272	*	0.107		0.305	**
school27	0.298	*	0.133	*	0.330	***
school28	0.330	**	0.165	***	0.330	***
school29	0.263	*	0.098	*	0.259	***
school30	0.170		0.005		0.184	**
school31	0.186		0.021		0.222	**
school32	0.208		0.043		0.228	*

To be continued …

Table 2 – School Effects of OLS Regression Results with/without Constraint and Error-in-Variance Adjustment (con't)

Parameters	OLS Estimation		Constrained OLS Estimation		OLS Estimation Contorlling EIV	
school33	0.108		-0.057		0.127	
school34	0.035		-0.130	**	0.066	
school35	0.157		-0.008		0.193	*
school36	0.136		-0.030		0.164	*
school37	0.215		0.050		0.203	*
school38	0.273	*	0.108		0.273	**
school39	0.063		-0.103		0.041	
school40	0.177		0.011		0.208	**
school41	-0.011		-0.176	*	0.017	
school42	0.199		0.034		0.232	**
school43	0.106		-0.059		0.120	
school44	0.289	*	0.124		0.323	***
school45	0.200		0.034		0.217	**
school46	0.254	*	0.088	*	0.274	***
school47	0.001		-0.164	**	0.007	
school48	0.132		-0.034		0.136	
school49	0.038		-0.127	*	0.077	
school50	0.218		0.053		0.268	**
school51	0.143		-0.023		0.178	*
school52	0.296	*	0.131	*	0.346	***
school53	0.076		-0.089		0.117	
school54	-0.034		-0.199	**	(dropped)	
scodum	-0.137	***	-0.137	***	-0.115	**
_cons	-0.116		0.050		-0.156	*

*** = p <= 0.05**
**** = p <= 0.01**
***** = p <= 0.001**

Table 3 – Summary of Two-Level Hierarchical Linear Model Analysis

Two-Level Hierarchical Linear Model Analysis Summary

Fixed Effects	**INTRCPT**	-0.150		-0.144		-0.141		-0.150		-0.236	*
	MAP_Math	0.398	***	0.407	***	0.407	***	0.408	***	0.406	***
	CogAT	0.458	***	0.450	***	0.450	***	0.449	***	0.450	***
	F/R Lunch	-0.059	**	-0.056	**	-0.080		-0.075		-0.079	
	%F/R Lunch					0.062		0.054		0.059	
	Female	-0.040	**	-0.038	*	0.007		0.012		0.048	
	%Female					-0.090		-0.098		-0.167	
	White	-0.002		-0.004		-0.003		-0.005		-0.003	
	Black	-0.030		-0.038		-0.037		-0.039		-0.040	
	Asian	0.243	***	0.244	***	0.246	***	0.242	***	0.233	**
	Sco_Miss	-0.137	***	-0.137	***	-0.136	***	-0.137	***	-0.142	***
		11/53	***	11/53	***	11/53	***	11/53	***	11/53	***
	Sch_Dummy	11/53	**	10/53	**	9/53	**	11/53	**	16/53	**
		10/53	*	11/53	*	9/53	*	9/53	*	11/53	*
	LEVEL-1 Std	0.521		0.518		0.519		0.518		0.517	
Random Effects	**INTRCPT**	0.000	(0.000)								
	MAP_Math			0.071	(0.628)	0.071	(0.626)	0.073	(0.635)	0.075	(0.651)
	CogAT			0.064	(0.601)	0.064	(0.602)	0.062	(0.561)	0.074	(0.641)
	F/R Lunch							0.058	(0.4)	0.066	(0.385)
	Female							0.020	(0.132)	0.022	(0.113)
	White									0.081	(0.652)
	Black									0.029	(0.082)
	Asian									0.117	(0.304)

*** = p <= 0.05**
**** = p <= 0.01**
***** = p <= 0.001**

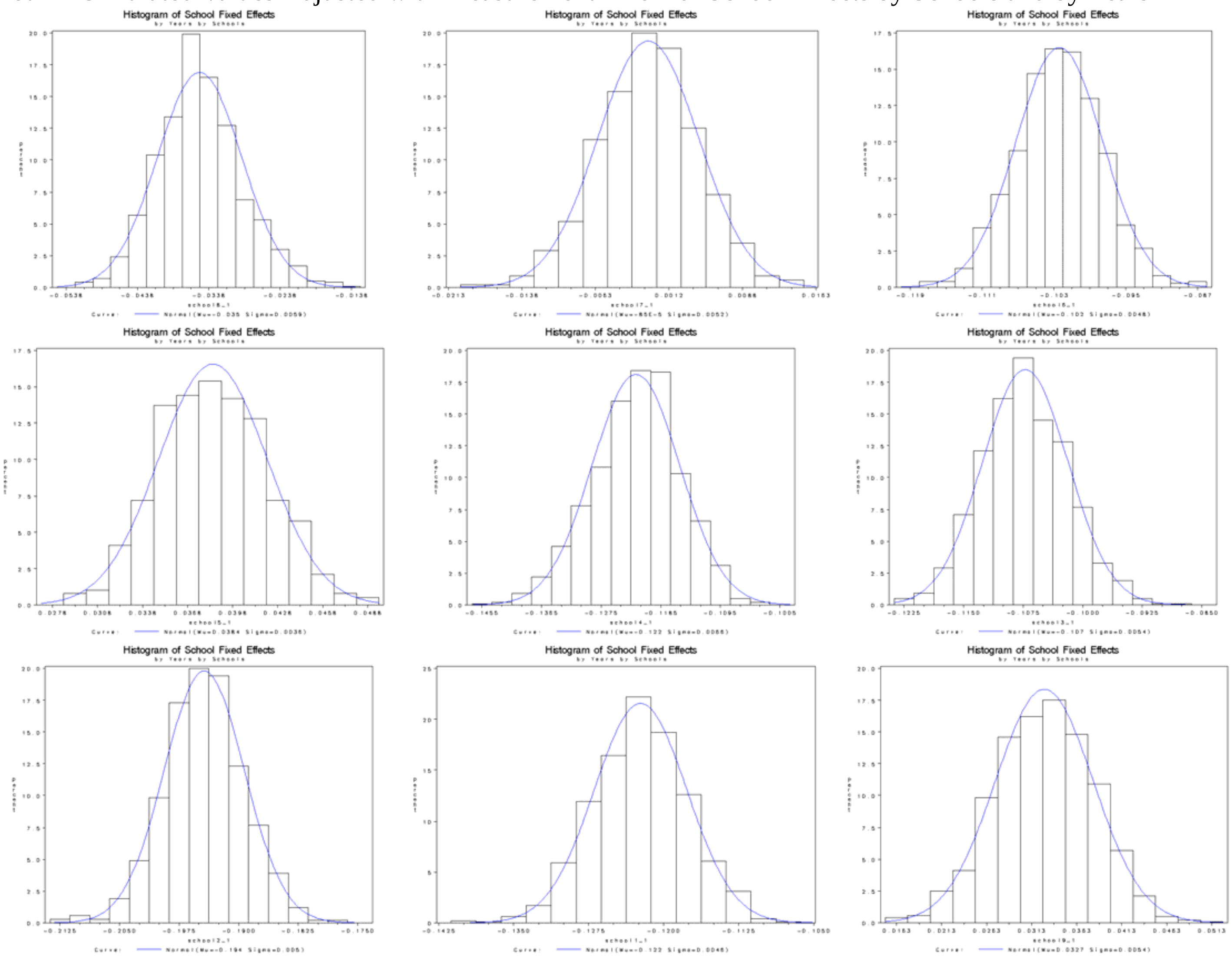
Notes: 1. The last school dummy was dropped from two-level HLM analysis;
2. The reliability of the random variable is in parentheses.

Table 4 – HLM results for Random Effects

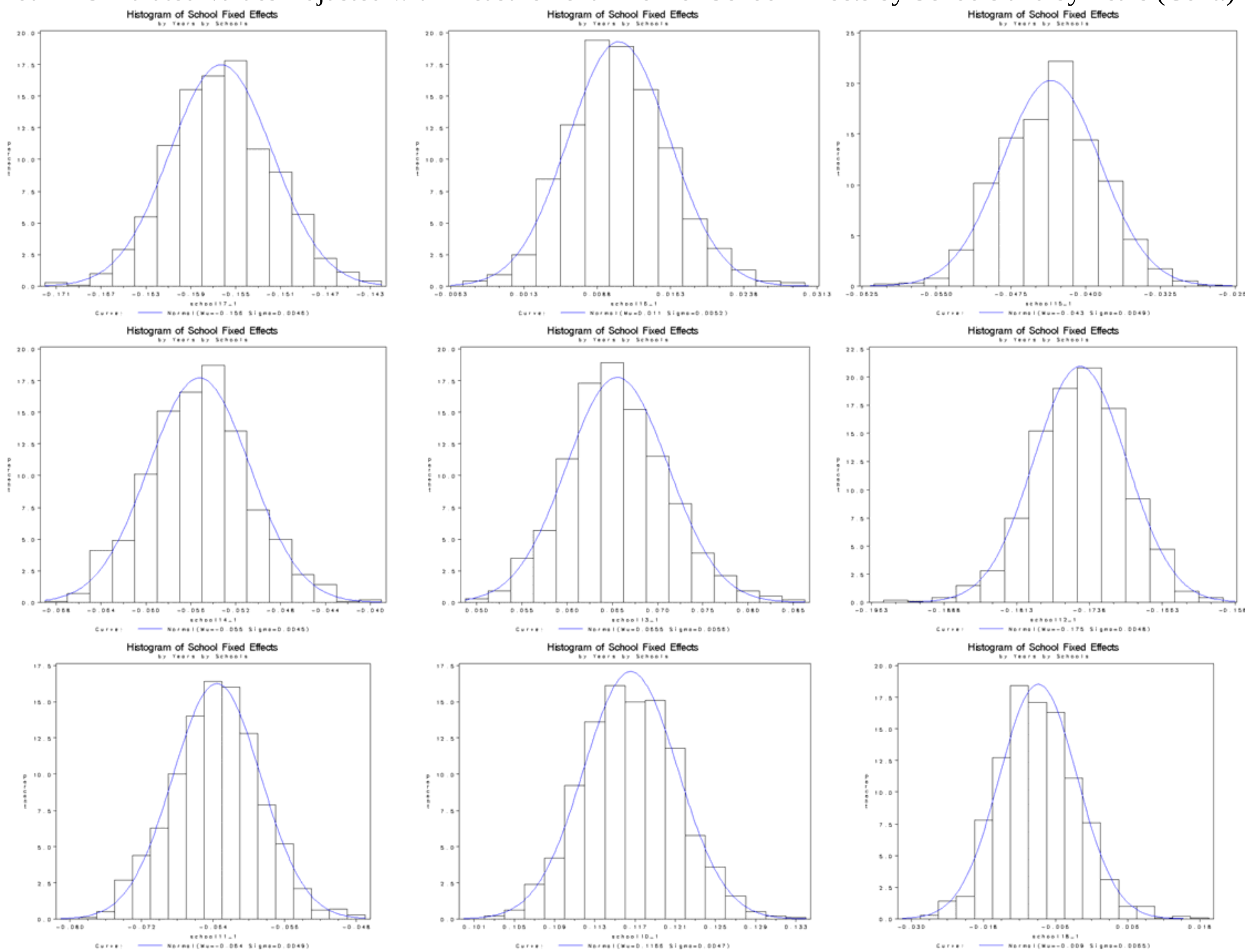
Random Effects	Effect		S.E.	T-ratio	Reliability	Parameter Variance	
CogAT	0.450	*	0.020	22.351	0.602	0.004	*
Prior Math	0.407	*	0.021	19.054	0.626	0.005	*
Level-2 Cov	-0.004						
Level-1 R-sqr	0.269						

* Significant at the 0.001 level.

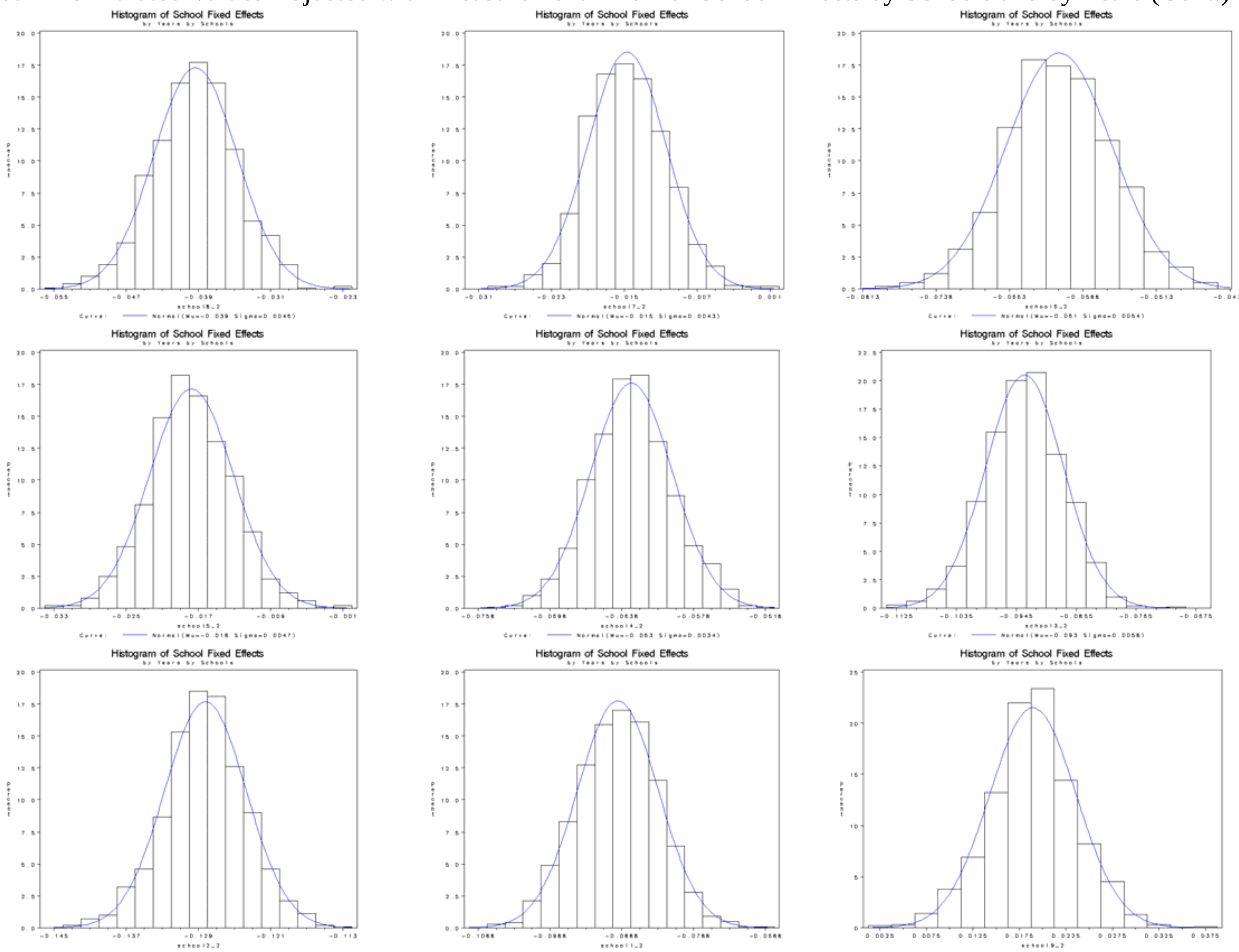
Plot 1 – Simulated Values Adjusted with Measurement Error for School Effects by Schools and by Years



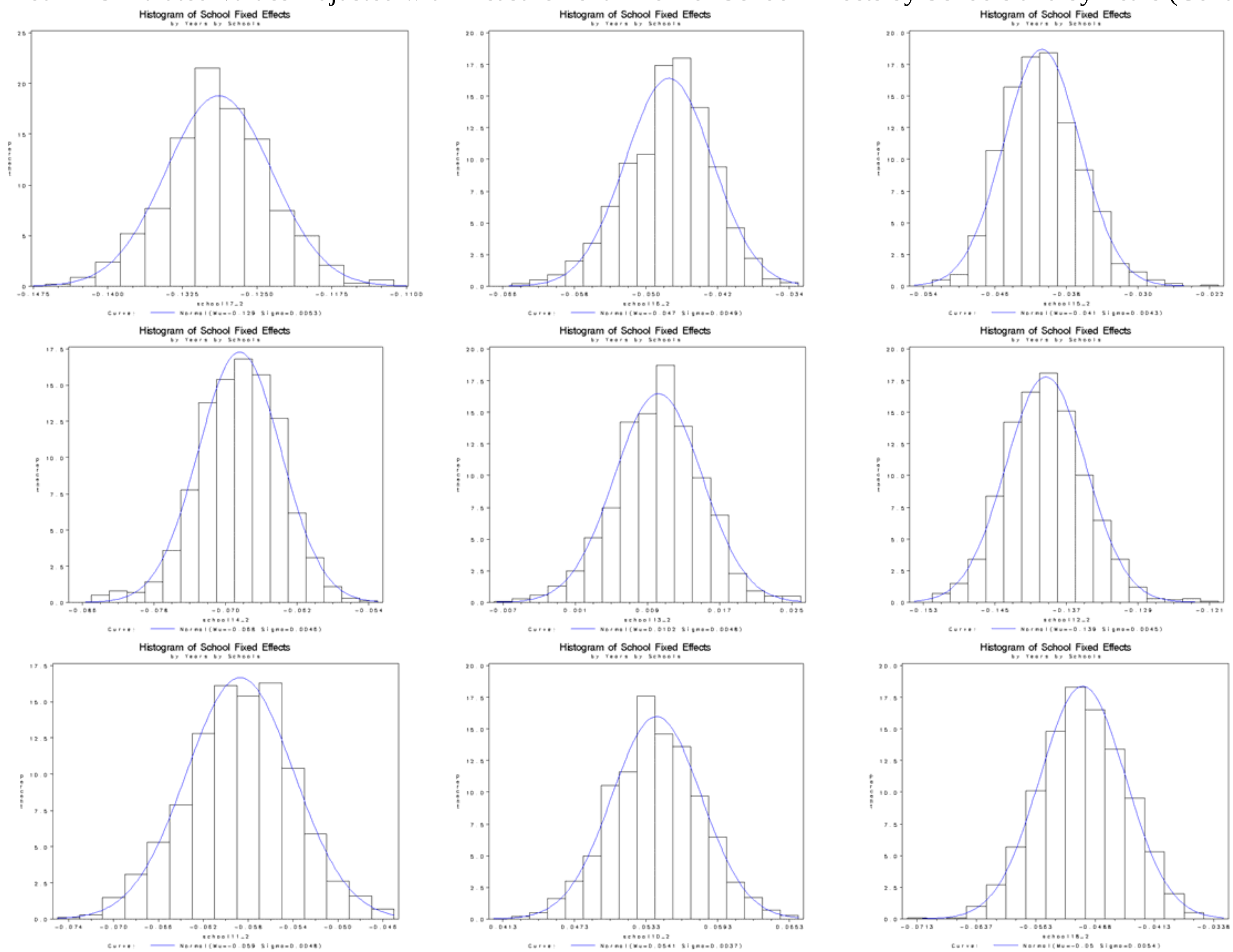
Plot 1 – Simulated Values Adjusted with Measurement Error for School Effects by Schools and by Years (Cont.)



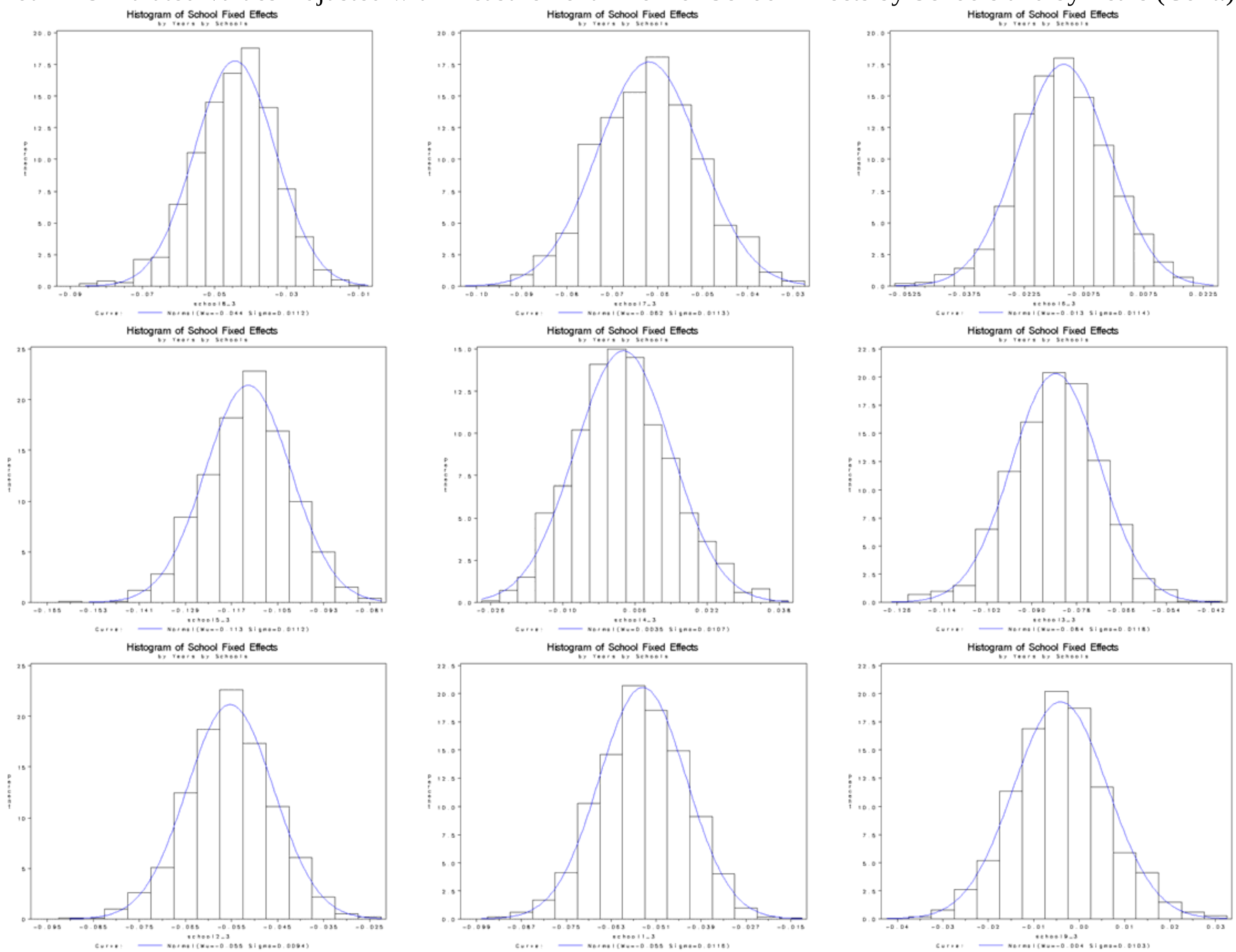
Plot 1 – Simulated Values Adjusted with Measurement Error for School Effects by Schools and by Years (Cont.)



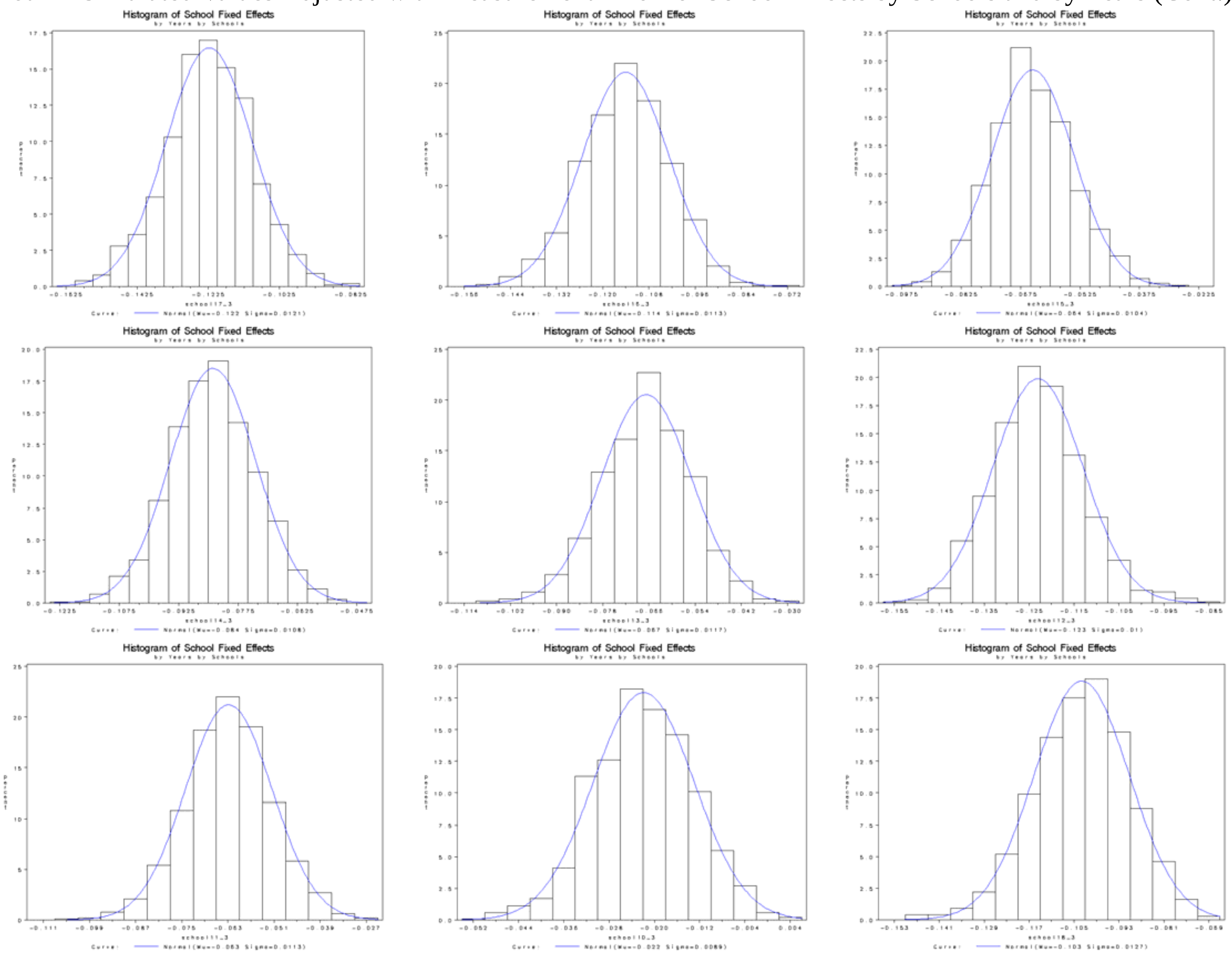
Plot 1 – Simulated Values Adjusted with Measurement Error for School Effects by Schools and by Years (Cont.)



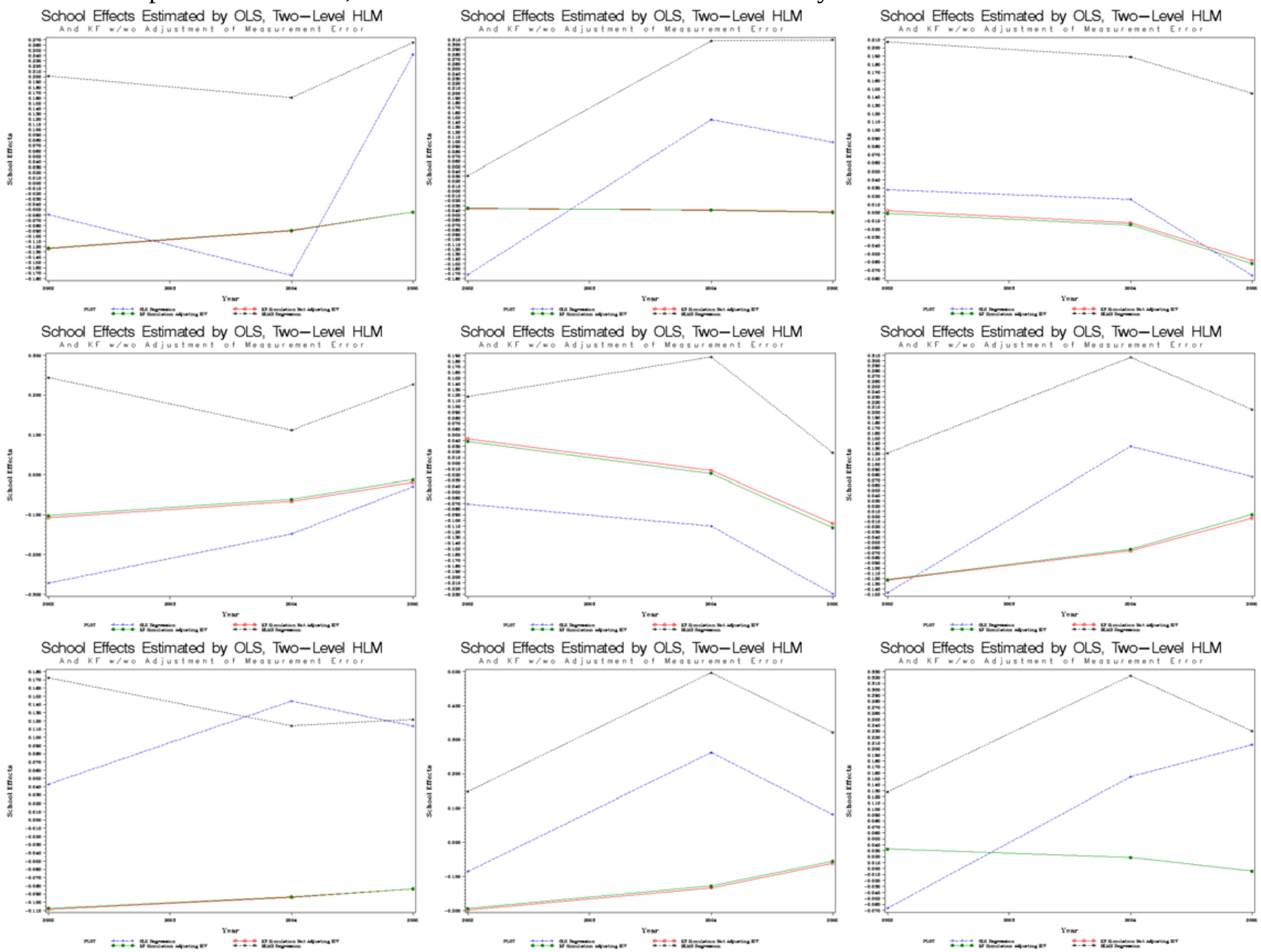
Plot 1 – Simulated Values Adjusted with Measurement Error for School Effects by Schools and by Years (Cont.)



Plot 1 – Simulated Values Adjusted with Measurement Error for School Effects by Schools and by Years (Cont.)



Plot 2 – Comparison with OLS, HLM and KF results for School Effects by Schools



Plot 2 – Comparison with OLS, HLM and KF results for School Effects by Schools (Cont.)