LI05405

Regression Toward the Mean Artifacts and Matthew Effects in Multilevel Value-Added Analyses of Individual Schools

Xiaoxu Li, *The Chinese University of Hong Kong, Hong Kong;*
SELF Research Centre, University of Western Sydney, Australia
Herbert W. Marsh, *SELF Research Centre, University of Western Sydney, Australia*
Kit-Tai Hau, *The Chinese University of Hong Kong, Hong Kong*
Irene T. Ho, *The University of Hong Kong, Hong Kong*
Andrew J. Martin, *SELF Research Centre, University of Western Sydney, Australia*

League tables are a problematic approach to infer school effectiveness, but traditional value-added approaches are fraught with statistical complexities. According to the Regression Toward the Mean Artifacts (RTMA), students with initially high or low scores tend to regress toward the mean in subsequent testings, resulting in biased estimates of school growth (Marsh & Hau, 2002). Matthew effects are apparently counter-balancing artifacts in growth to describe the phenomena that achievement gains are systematically larger for students who are initially more able (i.e., the rich becomes richer). Mathematical proof shows that although the Matthew and the RTMA artifacts work in opposite directions and tend to cancel each other, they share a similar mechanism and can be rectified. In this study, mathematical derivations and Monte Carlo simulated data are used to compare four models, namely: (i) without any remedy, (ii) with remedy for the Matthew effect only, (iii) with remedy for the RTMA only, and (iv) with remedies for both the Matthew effect and the RTMA. The conditional strategy with individual assignment test scores (used in assigning students to different schools) as covariate remedies artifacts and the result is consistent with Marsh & Hau's (2002) conclusion for RTMA. The associated problems with the two effects in estimating school value-added information are discussed.

The valid measurement of the quality of education provided by a school is always a topic of academic and public contention. Parents, government education departments, and policy makers have great interest to find out which schools are providing better quality of education than the others. Particularly when the initial ability of the students at admission varies so greatly, it is inappropriate to identify the better schools using the naïve and simple comparison of the final outcome -- students' attainment at graduation. If schools are compared by students' performance at graduation only without taking into consideration their initial ability at intake, then schools with better intake will wrongly and too simplistically be taken as providing higher quality of education. A statistical value-added model that appropriately takes into consideration the difference in intake and output (i.e., the added value) is necessary in determining the quality of the educational process.

The calculation of added value is complicated. Adding value to a student is not enough, a school has to add more value than other schools to be considered good. Understandably for a sufficiently long span of schooling (e.g., 6 years of primary school education), all students learn some knowledge and "improve", thus leading to a positive added value in the strictest sense for all students and for all schools. That means, an average or a net positive added value in achievement for all students in a certain school is not good enough. A more practical, reasonable and common definition for added value is the extra value or improvement students made beyond those in an averaged school. An unavoidable drawback of this definition is that because it is a comparison among all schools within a particular system with no reference to some external standards, it is necessarily a zero-sum game. With this definition using the average as the standard, there will be approximately equal number of schools with positive and negative added value.

The proper analysis of added value relies on the availability of at least three building components, namely, the statistical theory, the analytical software and the necessary database of the targeted students' performance. On the theory component, the multilevel analytical method (e.g., Raudenbush & Bryk, 2002) serves as the standard strategy. Singer (1998) also provides a concise working guide on the mathematical framework of similar analyses, which is accompanied by its implementation on the flexible platform SAS PROC MIXED. Her work has also been extensively referred to in the present complex simulation research.

Multilevel analytical models have been used in actual operating systems, such as the Tennessee Value-Added Assessment System (Sanders & Horn, 1998; Baker, Xu, Detch & Snodgrass, 1995). The added value can be computed from the growth in individual academic performance and is customarily used as the dependent variable in the analyses. It is also possible to attribute, partially or selectively, the outcome dependent variable to some explanatory variables at the school or cohort level. For example, in Marsh and Hau's (2002) model in which the present study is based on, the average admission-test score of all students in each school is used as a school level explanatory variable reflecting the academic competitiveness of the schools.

In this study, we will concentrate on two common phenomena, namely the regression toward the mean artifacts (RTMA) and the Matthew effects, which might jeopardize the correct modeling and interpretation of the value-added analyses. That is, we attempt to compare various statistical models in value-added computation to see their appropriateness in estimating the quality of schools, particularly in the presence of the RTMA and the Matthew effects.

The detailed discussion of the RTMA and Matthew effects is beyond this article. Briefly, the RTMA refers to the tendency of extreme scores to regress toward the mean in the subsequent testing. Marsh and Hau (2002) provided a detailed review and account of RTMA and some potential remedial approaches. In the simulation, the true school effects are identical across schools (i.e., all schools are designed to have an equal impact on their students, there is no school which is better than others). That would imply a zero school effect or added value for all schools. Thus, a mathematical model would be inappropriate if it shows that the added value of schools has a statistically significant non-zero regression coefficient with respective to any explanatory variable (e.g., the mean ability intake of the school). As the true added values of all schools are identical, a statistical model will be erroneous if it shows some schools being better than the others.

One may suspect that the RTMA is a term derived from *regression*. Actually, according to the historical development, it is the other way round. When F. Galton first used the term *regression*, it referred to the RTMA (e.g., tall fathers have shorter sons whose heights "regress" toward the population mean and vice versa; Campbell & Kenny, 1999, pp. 2-3). The term *regression* has a different meaning now and has a much popular popular usage than the RTMA.

In the specific simplistic RTMA situation referred by Galton, the prediction power of the latent explanatory variable and its independent measurement error variance were both assumed to be non-zero. In a slightly more general specification used in Marsh and Hau (2002) and in this study, the explanatory variable could include measurement errors which are correlated with the dependent variable, while the latent explanatory school quality could also have a zero prediction power of the dependent variable.

The above two forms of model specification differ in a number of aspects. Galton's original form can be seen as one in which the explanatory variable is contaminated by measurement errors whose regression coefficient (or simply viewed as correlation) on the dependent variable (i.e., the height of sons in Galton's case, and the added value in the present research) is zero. The error part in the explanatory school quality in Marsh and Hau's (2002) model has a non-zero negative regression relation with the dependent variable. If the regression coefficients (or simply correlations) of the error part and of the latent explanatory part on the dependent variable are both zero, then there is no RTMA. If the above two regression coefficients are identical (equal to any zero or non-zero value), still there will not be any RTMA. Whenever the above two regression coefficients are different, a generalized RTMA would be observed. Thus, a common definition of the RTMA, as used in this study, is that "if the observed explanatory variable has a measurement error and if two latent constructs have different regression relations with the dependent variables, the original regression relation of the latent explanatory part on the dependent variable will be distorted." Actually this variant of RTMA definition is a little more accommodating in the sense that the direction of distortion has not been restricted (i.e., it can be negative as in traditional RTMA or positive in a direction opposite to the traditional RTMA).

We also examine how the Matthew effects may affect the value-added estimation. Matthew effect in this study is operationally defined as the positive correlation between individual students' initial academic

achievement and their growth rates. Although such a definition may have some problems (Bast & Reitsma, 1997), it simplifies our analyses and provides a straightforward understanding. In a lot of cases, previous studies (e.g., Luyten, Cremers-van Wees & Bosker, 2003) suggest that the Matthew effects may not necessarily stay constant or be linear. The actual Matthew effects in operation can be much diversified in form and more complicated than our archetype presented here. Our interest in the present study is not so much on fitting the Matthew effect parameters using empirical data. Rather we use the simplest form of Matthew effect in which there is an increasing trend of the variance of measurement from a certain time point. We try to examine how this increasing variance of measurement affects the value-added modeling.

The Analytical Framework

For sake of simplicity, as in other simulation studies, our model is based on an ideal and theoretical setup rather than on an actual empirical data set though we have included substantial meaning for the design to help interpretation.

Student Level

In this study, we estimate the added value for high schools, i.e., longitudinal and total effects on Grade 7, 8, 9 students. We assume we are able to trace students' performance in their last three years in primary school (Grade 4, 5, 6) and following them after they are allocated to high schools as far as to Grade 9 (the third year in high schools). Added value is calculated for the last 3 years of the 6-year observation period. The school effects are estimated by aggregating the added value of all students in each school. We assume all students, of sample size N, take one examination at the end of each year in the 3 + 3 consecutive years, with respective test scores being T_{-2} to T_3. T_0 is the score at the point of admission into high schools while T_1, T_2, T_3 are the scores at the end of the 1st, 2nd, and 3rd years (Grade 7, 8, 9) in high schools.

For each test, the scores (T_{-2} to T_3) of each student include two parts. One is the latent academic traits (A_2 to A_3); the other is the independently random measurement errors (E_{-2} to E_3). For simplicity, we also fix the reliability of the measurements at each year (T_{-2} to T_3) to be identical, denoted by α as in the following.

$$\alpha \equiv \frac{Var(A_t)}{Var(T_t)}, \ t = -2, -1, 0, 1, 2, 3.$$

Our framework is simpler than that used by Marsh and Hau (2002) in that the trait instability in this study is minimized so that over years the latent academic traits (analogous to true ability, A_{-2} to A_3) of each student are strictly linear, with D being the difference between any neighboring years. For each individual student, his A_0 and D will totally determine his subsequent abilities (A_{-2} to A_3). The distribution of A_0, across all students, is assumed to be normal and centered at zero as a baseline. For simplicity, we also assume D to be standardized and normal across all students (mean = 0, SD=1). So some students grow faster (i.e., positively), some grow slower (i.e., negatively), with the mean growth across students set at zero.

From the above description, students' latent traits at t = -2 to 3 (i.e., A_{-2} to A_3) are generated from the pair of random variables—D and A_0. The joint normal distribution of D and A_0 with a constant test reliability α defines the 6 × 6 COV-VAR matrix of the observable variables, i.e., the 6 successive test scores (T_{-2} to T_3), each of which is used to generate the N samples. Actually, each of the variables (A_{-2} to A_3, T_{-2} to T_3. and D) at the student level is normal and mean centered at zero.

To incorporate the Matthew effects from a certain time point, we denote the correlation between A_0 and D as r, and the SD of A_0 as σ. A positive value of r implies a faster growth for students with higher initial ability A_0, while a zero r represents a growth rate independent of initial ability. The SD of D has been set at unity (1)

and the 2 × 2 COV-VAR matrix of A_0 and D is $\begin{pmatrix} \sigma^2 & r \times \sigma \\ r \times \sigma & 1 \end{pmatrix}$. At each t, t = -2,-1, 0,1,2,3, $T_t = A_0 + t \times D +$ E_t, and the cell on the pth row and the qth column in the 6 × 6 COV-VAR matrix of (T_{-2} to T_3) can be computed as follows,

$$\begin{cases} [\sigma^2 + (p-3+q-3) \times \sigma \times r + (p-3) \times (q-3)] \times \dfrac{1}{\alpha}, p = q \\ \sigma^2 + (p-3+q-3) \times \sigma \times r + (p-3) \times (q-3), p \neq q \end{cases}$$

where p, q = 1,2,3,4,5,6, that is (p-3), (q-3) = -2,-1,0,1,2,3. Let p-3 = q-3 = t , we have the following.

$$\mathrm{VAR}T_t) = (\sigma^2 + 2t \times \sigma \times r + t^2) \times \frac{1}{\alpha} = \frac{1}{\alpha}[t - (-\sigma \times r)]^2 + \frac{1}{\alpha}(\sigma^2 - \sigma^2 \times r^2)$$

It can be demonstrated that from and only from $t_{Min} = -r \times \sigma$, VAR($T_t$) increases (see Figure 1). So, when t ≧ -r·σ, Matthew effects manifest. Due to this relation, perhaps worth noting is that in the present design, positive Matthew effects only start at a certain point on the time line and they can become negative if we trace back beyond this point. Without loss of generality, the above analyses for data points with positive Matthew effects do not restrict the validity of our conclusion in this study to conditions involving positive or negative Matthew effects only.

Figure 1 The relation between var(T_t) and t_{Min}
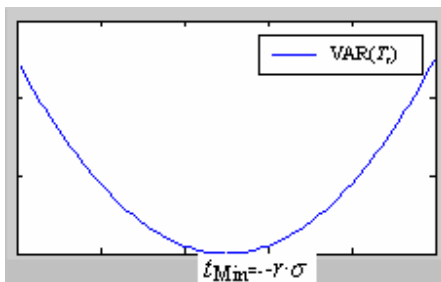
School Level

We assume there is a transition, moving of students from primary to high schools, at t = 0. The promotion procedure from the primary to high schools is identical to that in Marsh and Hau's (2002) study. Specifically, we assume the students are segregated by their entrance test scores T_0 in promoting from primary to high school. In brief, students with higher test scores tend to aggregate into "competitive" schools having more high ability students. In the assignment of students to high schools, the students are first ranked by their T_0 scores and then divided into 5 ability-groups of equal size according to an ascending order of scores. Thus, students in the top 20% belong to the top ability-group; the next 20% belong to ability-group 2 and so on. Five types of schools, Bands 1 to 5, are created with random samples of students according to the following proportions from each ability-group. In general, Band 1 schools take in the most number of students from the highest ability group while Band 5 schools take in the most number of students from the lowest ability group. This reflects a close to reality pattern in which magnet schools tend to attract students with initially higher ability. That is, students with different abilities do tend to segregated among themselves into schools of different competitiveness.

Table 1, Proportions from each ability-group

	Total of 5 Bands	School Band 1	School Band 2	School Band 3	School Band 4	School Band 5
Student group 1 (highest ability)	1	0.5	0.3	0.1	0.1	0
Student group 2	1	0.3	0.2	0.2	0.2	0.1
Student group 3	1	0.1	0.2	0.4	0.2	0.1
Student group 4	1	0.1	0.2	0.2	0.2	0.3
Student group 5 (lowest ability)	1	0	0.1	0.1	0.3	0.5

For the jth high school, the mean T_0 of its students is calculated and denoted as Q_j, which reflects the respective average ability of the intake students (school quality). A positive Q_j denotes a better school (as compared to an average school) with intake of more higher ability students (T_0) while a negative Q_j indicates a relatively weaker school with more lower ability students. Each student i from $i=1$ to N is nested in the jth high school j. The map is noted as $j=$J(i). Thus, $Q_{J(i)}$ means the school quality for the ith student.

To indicate the variables at the student level, one more subscript is needed, e.g., $T_{0,i}$ means the test score in the admission year (which is the performance at the end of the last year at the primary school) of the ith student while $A_{0,i}$ means the latent academic ability in the admission year of the ith student. In the following, $E_{0,i}$, defined as $T_{0,i}$-$A_{0,i}$ (that is the difference between the observed score $T_{0,i}$ and academic ability $A_{0,i}$), indicates the measurement error part in the ith student's test score in the year admitted to the high school.

Model and Result

A number of analytical models have been proposed to assess the added value of schools. In the population, we have set the added value of all schools to be zero. Thus any model that gives a significantly non-zero estimated school effect is inappropriate and should not be recommended.

A. Unconditional Models

Compared to the two-level models used in Marsh and Hau's (2002) unconditional models, our models in this study are relatively simple to reveal the source of artifacts. All four models share a general form. The only differences among them are their respective definition of added value, denoted as V_i for i^{th} student, which is also used as the dependent variable in the models. In model one to four, V_i is defined with the observable test scores as in the followings.

At the student level, for i^{th} student in the jth school, $j=$J(i),
$$V_i = \beta_{J(i)} + e_i, \qquad i=1 \text{ to } N, \ e_{(.)} \sim N(0, \sigma^2)$$

At the school level, for j^{th} school,
$$\beta_j = \gamma_0 + \gamma_1 * Q_j + u_j, \qquad j = 1 \text{ to } M, \ u_{(.)} \sim N(0, \tau)$$

Putting these two levels of equations together, the following model can serve as a general model in all analyses:
$$V_i = \gamma_0 + \gamma_1 * Q_{J(i)} + u_{J(i)} + e_i,$$

As seen in Appendix A, it can be shown that COV(V,Q) and COV(V,T_0) share an identical sign (+, 0, or -), while the former is of crucial importance in our derivation, the latter is used for numeric computation (see Table 2).

Table 2

Models	Description	How added value is operationally measured	COV(V, T_0)	Problems (without controlling for)
1. Comparison of Entry and Exit performance	Take into consideration preexisting differences, most popular model	$V_i = T_{3,i} - T_{0,i}$ $= 3D - E_0 + E_3$	$\sigma^2(3\dfrac{r}{\sigma} - \dfrac{1-\alpha}{\alpha})$	RTMA, Matthew effects
2. Comparison of end of Year 1 and Exit Performance	Comparison between entry and exit but avoiding RTMA by using the end of high school first year performance	$V_i = T_{3,i} - T_{1,i}$ $= 2D + (E_3 - E_1)$	$2r\sigma$	Matthew effects
3. Comparison of primary and high school growth rates	Take into consideration the possible growth trajectories of individual students	$V_i = \dfrac{T_{3,i} - T_{0,i}}{3} - \dfrac{T_{0,i} - T_{-2,i}}{2}$ $= -\dfrac{5}{6}E_0 + (\dfrac{E_3}{3} - \dfrac{E_{-2}}{2})$	$-\dfrac{5}{6}\sigma^2 \times \dfrac{1-\alpha}{\alpha}$	RTMA
4. Comparison of primary and high school growth rates using end of Year 1 performance	Take into consideration individual growth trajectories and avoid RTMA using end of high school first year performance; this is the only appropriate model to be adopted	$V_i = \dfrac{T_{3,i} - T_{1,i}}{2} - (T_{-1,i} - T_{-2,i})$ $= \dfrac{E_{3,i} - E_{1,i}}{2} - (E_{-1,i} - E_{-2,i})$	0	

Figure 2
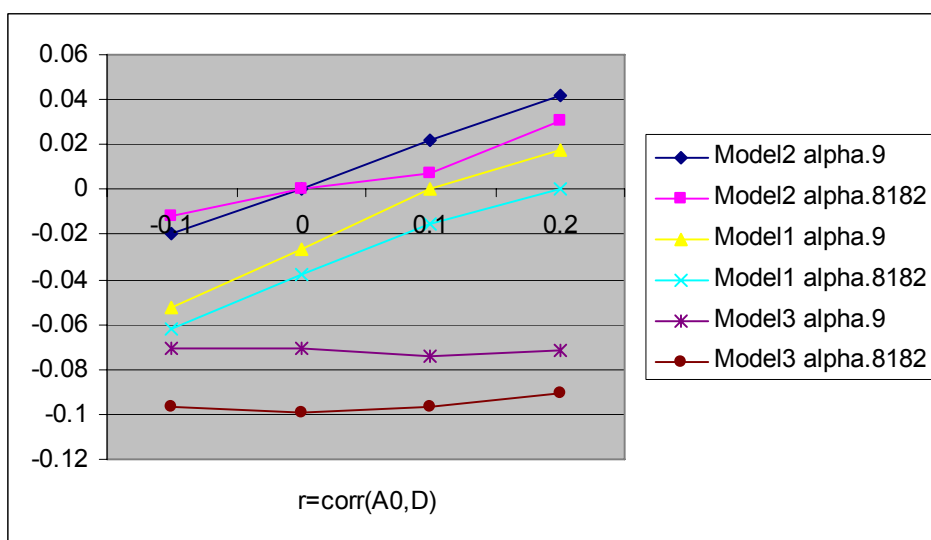
The above figure display the cases σ=2.7 , in which Matthew effects are manifested from t   -2.7r for non-zero values of r.

Though mathematically, the derivation can be used to demonstrate the (in)appropriateness of various models, we are using a simulation study to demonstrate and validate the mathematical derivations. The population model has a zero school effect (added value) on all students and for all schools, so any deviation from this result indicates an inappropriate statistical conclusion. There are 150^2 (N) students nested in 150 (M) schools. The software platform being used is SAS PROC Mixed (Singer, 1998) with SAS Marco (Fan, Felsovalyi, Stephen, Sivo & Keenan, 2002). As can be seen from Table 2 and Figure 2, the simulated effect results fit the mathematical prediction perfectly. Each of the predicted non-zero effects reaches the statistical significance of .0005.

B. Conditional Model

The conditional models are as follow:

$$V_i = \beta_{0,J(i)} + \beta_{1,J(i)} * T_{0,i} + r_i \quad , i = 1...N$$

$$r_i, i = 1...N \quad \sim N(0, Sigma^2)$$

$$\beta_{0,j} = \gamma_{00} + \gamma_{01}Q_j + u_{0,j} \quad , j = 1...M$$

$$\beta_{1,j} = \gamma_{10} + \gamma_{11}Q_j + u_{1,j} \quad , j = 1...M$$

$$\begin{pmatrix} u_{0,j} \\ u_{1,j} \end{pmatrix}, j = 1...M \quad \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix})$$

The models are identical to the following.

$$V_i - \gamma_{10}T_{0,i} = \gamma_{00} + \gamma_{01}Q_{J(i)} + \gamma_{11}Q_{J(i)}T_{0,i} + u_{1,J(i)} * T_{0,i} + u_{0,J(i)} + r_i \quad , i = 1...N$$

$$r_i, i = 1...N \quad \sim N(0, Sigma^2)$$

$$\begin{pmatrix} u_{0,j} \\ u_{1,j} \end{pmatrix}, j = 1...M \quad \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix})$$

$V = (V - b \times T_0) + b \times T_0$, the coefficient b is chosen such that $\text{COV}((V\text{-}b{\cdot}T_0), T_0) = 0$, making $V\text{-}b{\cdot}T_0$ and T_0 zero-correlated, thus, independent. So, $V\text{-}b{\cdot}T_0$ is independent of $Q_{J(.)}$ and $Q_{J(.)}{\cdot}T_0$. So the estimated value is $\gamma_{10}=b$; $\gamma_{01}=\gamma_{11}=0$ irrespective of whether the Matthew effects exist or not. The decomposition of V here shares a similar logic to that in Appendix A for unconditional models.

Discussion

Here we briefly discuss and compare some potential problems of various analytical models. As all school effects are zero in the population by design, the finding of non-zero school effects indicates the respective statistical models are inappropriate. Such artificial effects come from two sources. The first is that we allow correlation of A_0 and D in some of the designs, which means that students gain differentially according to their initial academic ability. For example, in one of the designs, some students with higher initial academic ability grow faster. Thus, it is logical to expect that those schools with these initially higher ability students will have an artificially positive added value. From another perspective, any analytical model that does not appropriately take into consideration the possible differential individual growth will result in wrong conclusion on the added value of schools.

How can the above inappropriate analyses be solved? To avoid the above artificial inflation of added value for schools taking initially higher ability students, we can compare students' growth rates before and after they enter high schools. Specifically, if the students in a certain high school grow faster than their prior growth

rate in primary schools, this high school is said to have extra added value to its students. In that perspective, any school that increases the growth rate is said to have positive added value. The value-added evaluation becomes a problem of comparing the high school $[(T_3 - T_0)/3]$ and primary school $[(T_0 - T_{-2})/2]$ growth rates.

A very common and ignored problem is the RTMA in value-added calculation. In brief, this is the inherent problem in any calculation involving the pre-admission performance T_0. In any educational system in which students with better performance are segregated into high schools, due to the RTMA effect, it can be predicted that these schools will show a negative value-added effects on their students despite their true effect is zero.

More specifically, according to the RTMA, for any measurement with imperfect reliability (i.e., there are measurement errors), the students with higher scores tend to have positive measurement errors (E_0) as compared to other students in lower bands. Looking from another perspective, take as an example two students with the same initial latent academic ability A_0 but having a positive and negative E_0. The one with positive E_0 (thus higher T_0) will end up in a higher banding school, while the one with negative E_0 (thus lower T_0) will end up in a lower banding one even though these two students have identical latent academic ability. The consequence is that in the calculation of the growth rate in high schools, those schools admitting students with positive E_0 will have a slower growth rate as compared to those admitting students with negative E_0. That is, although the comparison between the growth rates during and prior to the high school periods is useful, any analytical models involving T_0 may still be problematic.

The quantity $Q_{J(i)}$, i=1 to N , representing school quality, plays an interesting role in our study. Despite that the variables T, Q, V seem to be distinct measures carrying different meaning to the public, they are closely related. Given that T affects V while T affects Q, Q could be shown to affect V. However, the seemingly distinct features of these three constructs, yet mathematically interrelated, make the derivations of their relations in the Appendix tantalizing.

In sum, the problems of the analytical methods as demonstrated in this study are rather general and commonly encountered in literature. Without a correlated school or cohort explanatory variable and the use of appropriate analytical strategies, the inappropriate value-added estimate could still be attributed to the attainment of the whole school or cohort and directly or indirectly credited to relevant teachers and principals. This would lead to a wrong conclusion that has to be taken care of in the value-added modeling of any school system in general.

About the Authors

Mr. Xiaoxu Li got his bachelor and master degrees at the Peking University, People's Republic of China and is currently PhD student at the Chinese University of Hong Kong. His research interests include time series, multilevel analysis, and structural equation modelling.

Professor Herb Marsh is Professor of Educational Psychology, founding Director of the SELF Research Centre and served as UWS's inaugural Dean of Graduate Research Studies and Pro-Vice-Chancellor. He received UWS's inaugural awards for Research, Postgraduate Supervision, and Doctorate of Science. Herb has published more than 250 peer-reviewed journal articles, 40 chapters, 10 monographs, and 225 conference papers.

Professor Kit-Tai Hau is Professor and Chair of Educational Psychology Department at The Chinese University of Hong Kong. His research interests include structural equation modeling, academic achievement motivation, policy on educational assessment and adolescent suicide. He is one of the international founding members of SELF center and has made frequent visits, including long stays under the sponsorship of the Australian Research Grant Council.

Dr. Irene T. Ho is Assistant Professor at the Department of Psychology and a teacher of the Educational Psychology Program, The University of Hong Kong. Her research interests include learning motivation, teacher efficacy, thinking skills development, and children with special needs.

Dr. Andrew Martin is a Research Fellow at the SELF Research Centre, University of Western Sydney. He is a Registered Psychologist and his research interests are student motivation, parenting, and research methodology. In 2002, his PhD was judged the Most Outstanding Doctoral Dissertation in Educational Psychology by the American Psychological Association and before that was judged the Most Outstanding PhD in Education in Australia by the Australian Association for Research in Education. He has published over 35 refereed articles and chapters and presented 18 invited/keynote addresses in the past three years. Andrew also regularly conducts staff development in schools focused on enhancing student motivation in the classroom, and is author of "How to Motivate Your Child For School and Beyond".

Contact Details

Professor Herbert Marsh
Professor and Director
SELF Research Centre
University of Western Sydney, Bankstown Campus
Australia
Email: h.marsh@uws.edu.au

Reference

Baker, A. P., & Xu, D. (1995), *The measure of education: A review of the Tennessee value-added assessment system.* Nashville, TN: Tennessee Comptroller of the Treasury, Office of Education Accountability.

Bast, J. & Reitsma, P. (1997), Matthew effects in reading: A comparison of latent growth curve models and simplex models with structured means, *Multivariate Behavioral Research,* 32(2), 135-167

Campbell, D. T. & Kenny, D. A. (1999). *A primer on regression artifacts*. New York: Guilford Press.

Fan, X, Felsuovalyi, A., Sivo, S. A., & Keenan, S. C. (2003). *SAS for Monte Carlo studies: A guide for quantitative researchers.* Cary, NC: SAS Publishing

Luyten, H., Cremers-van Wees, L. M. C. M., & Bosker, R. J. (2003) The Matthew effect in Dutch primary education: Differences between schools, cohorts and pupils, *Research Papers in Education*, 18(2), 167-195.

Marsh, H. W. & Hau, K.T.(2002), Multilevel modeling of longitudinal growth and change: Substantive effects or regression toward the mean artifacts? *Multivariate Behavioral Research*, 37(2), 245-282

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (2nd ed.)*. Thousand Oaks, CA: Sage.

Sanders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee value-added assessment system (TVAAS) database: Implications for educational evaluation and research, *Journal of Personal Evaluation in Education, 12*(3), 247-256.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 24(4). 323-355.

Appendix A

To prove: $COV(V,Q)$ and $COV(V,T_0)$ share an identical sign (+, 0, or -).

Step 1: To prove $CORR(T_{0,(\cdot)}, Q_{J(\cdot)}) > 0$.

The N samples of T_0, notated ($T_{0,i}$; i=1 to N) with some independent random event ω decide the J(i) mapping and subsequently the corresponding samples of school quality ($Q_{J(i)}$; i=1 to N). The larger $T_{0,i}$ corresponds to the larger conditional expectation of $Q_{J(i)}$. Given the perfect symmetry, it can be shown that $CORR(T_{0,(\cdot)}, Q_{J(\cdot)}) > 0$.

Step 2: To make orthogonal decompositions within which the zero correlation implies independence.

$$Q_{J(\cdot)} = a \times T_{0,(\cdot)} + (Q_{J(\cdot)} - a \times T_{0,(\cdot)})$$

wherein $a = \dfrac{COV(T_{0,(\cdot)}, Q_{J(\cdot)})}{VAR(T_{0,(\cdot)})}$ is positive and fit to $CORR(Q_{J(\cdot)} - a \times T_{0,(\cdot)}, T_{0(\cdot)}) = 0$.

For any normal variable V defined from T_0, D and measure errors (E_{-2} to E_3),

$$V = b \times T_0 + (V - b \times T_0), \text{ wherein } b = \dfrac{COV(T_0, V)}{VAR(T_0)}$$ and makes $CORR(V - b \times T_0, T_0) = 0$. Given the

normality, $(V - b \times T_0)$ is independent from T_0. Given its independence from ω, it can be derived that $(V - b \times T_0)$ is independent from $(Q_{J(\cdot)} - a \times T_{0,(\cdot)})$, which is determined by ($T_{0,i}$; i=1 to N) and ω.

Step 3: $COV(V,Q)=a \cdot b \cdot VAR(T_0)$, whose sign is determined by b, thus by $CORR(V,T_0)$.

In the model, the coefficient γ_1 is an indication of size of the school effect on added value. Thus an appropriate analytical method should give us a zero γ_1. From the above equation, since the other terms are either random errors or constant terms, the zero γ_1 would indicate that an appropriately defined V should be uncorrelated with $Q_{J(\cdot)}$ for any combinations of the parameter setting. In Table 2, in the comparison of school effects between models, the estimated value and SE of γ_1 should be adjusted by the scale of $\dfrac{1}{SD(V)}$. In this study, we use simulated data to demonstrate and validate the above mathematical derivation.