Comparison of Strategies for Value-Added Analyses:
Problems of Regression towards the Mean Artifact and Matthew Effect

Herbert W. Marsh University of Western Sydney
Xiaoxu Li The Chinese University of Hong Kong
Kit-Tai Hau The Chinese University of Hong Kong

Abstract

The analyses of value-added of schools have always been intriguing. We investigated two complications. One, the regression towards the mean artifact (RTMA), is caused by the fact that the measurement error parts of the luckier or the unluckier students scores tend to regress towards the mean zero in another testing). The other, Matthew effect, results from a positive correlation between future growth and gross previous growth. Through mathematical derivation and simulated data with a simplified frame, we compare a traditionally used approach with various remedies and demonstrate how value-added estimation can be systematically biased. These problems, often being ignored even in large-scale studies, will be discussed. We further show that the conditional strategy with individual school admission scores (used in assigning students to different schools) as covariate avoids both RTMAs and Matthew effects within our setting, which reaffirms Marsh and Hau's (2002) simulation conclusion for RTMA.

Keywords: regression towards the mean artifact, Matthew effects, value-added modeling

Literature Reviews

Value-added Modeling (VAM) for School Effects

Value-added has been a popular term (Wainer, 2004; Lissitz, 2005; Lissitz, 2006) since the pioneering project Tennessee Value-Added Accountability System (Sander & Horn, 1998) initiated its usage. VAM can be viewed as specially applying instances of more general quantitative methodologies like longitudinal modeling (Lissitz et al, 2006). McCaffrey et al (2003, pp17) gave a loose definition of VAM – "any education achievement model that uses gain scores or regresses current scores on prior scores". That definition is more extensive than the same banner used by Wright, Sander & Rivers (2006) because it includes the category of hierarchical linear models (HLM), which we utilized following.

The definition suggests the objectives of VAM are certain parts of achievement of students. In a conservative descriptive perspective, it estimates those parts aggregated by schools and divided across grades, in a linear and multivariate normal way. (Although we just discuss here at school level, issues at teacher level or school district level are parallel.) In a bold causal perspective, VAM estimates the difference of achievements of students made by their schools, compared to an implicit average school (Rubin et al, 2004; Rubin 1974). The concept value-added is mostly referred at school level or other group level, but it could also be used at individual student level without any ambiguity.

One key factor in the background of popularity of VAM in US is the adequate annual testing data that is required to states by No Child Left Behind (NCLB) Act (U.S., 2001; Rigney, Doran, & Martineau 2006). Moreover, the accountability movement of researches and appliances of VAM has spread abroad to different education systems, e.g., Hong Kong and Mainland China, where there are no adequate equated or vertically linked testing scores across schools, except at enrollment grade. Thus, the scores of enrollment common exams are inevitably playing a key role in the VAM practice in these education systems. Here we shall compare some value-added modeling strategies with or without enrollment exams data and analyze the mechanisms of two related potential comlications.

RTMA and Regression-toward-the-mean (RTM) within Value-added Context

”Perhaps every applied statistician worth his salt has said to himself at some time, ‘I understand regression-toward-the-mean; others only think they do’”(Roberts, 1980, p. 59). Since F. Galton historically drew out the first regression line, RTM phenomenon has been bewildering not only lay persons but also professional practitioners generation by generation (Galton, 1886; Campbell & Kenny, 1999; Marsh & Hau, 2002). However, the mathematic expression of RTM is so simple as “1-correlation” (Campbell & Kenny, 1999, p. 11), defined in a classic frame of a pair of positively correlated variables, the prediction and the predictor. Nonzero RTM is ubiquitous except in trivial cases with a perfect (1 or -1) correlation. It is absolutely impossible to be avoided, but the various artifacts caused by wrong intuitions involving it, loosely termed RTMAs, do be avoidable in certain settings or with certain knowledge. For example, the original paper of March & Hau (2002) investigated the remedy of RTMA within contexts of HLM longitudinal modeling involving non-equivalent groups and an entrance criterion variable.

The simplest form of RTMA implicates two key concepts. One is z-score. When we compare the predictor and the prediction, say, the grade I scores X and grade II scores Y, asking which is nearer toward the respective mean, we shall at first discount their different scales and calculate their z-scores. For example, z-score of the grade I should be (X-mean(X))/std(X). The other concept is conditional expectation. RTM is not to do with any single sample. What it concerns is the average grade II z-score of the whole subset of students with some special known grade I z-score, say .8. The artifact occurs when someone expects this average grade II z-score should be as constant as .8. With knowledge of regression, we know the true conditional expectation is $E(Z_Y|Z_X=.8)= r*Z_X=.8*r<.8$, wherein r is the correlation between X and Y and always less than 1, except the impossible trivial case for real scores of grades. We compare Z_X with $E(Z_Y|Z_X)$ and note that this simplest RTMA=Z_X- $E(Z_Y|Z_X)=(1-r)* Z_X$=RTM*Z_X. Here we cleared and distinguished the objective RTM from the various subjective potential RTMAs. And it is noted that any special artifact is related to a concrete false expecting.

Surely RTMAs are not limited to this simplest form. Literature has given different cases within various contexts (Campbell & Kenny, 1999; Marsh & Hau, 2002). This paper will follow the research of Marsh & Hau (2002) to simplify their frame and to analyze the essence of the artifact they uncovered. With the value-added terminology, the artifact could be related as -- some value-added models will overestimate those schools with poorer admission scores. Within our VAM setting several entangled questions are presented as following. How to distinguish RTMA from “Bigger Fish” effect -- truly rich-get-poorer and poor-get-richer (Luyten, Wees & Bosker, 2003)? Could RTMA be remedied simply by reducing measurement errors? How to distinguish RTMA from causal complication of Lord paradox? Will RTM last grade by grade or how to define RTM between two or more subsequent predictions? We shall give concrete answer or detailed debate for each in the discussion section.

Matthew Effect within Value-added Context

"For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath." (Gospel according to Matthew, XXV, 29). Merton (1968, 1988) introduced the biblical term Matthew effect into academic literature and Stanovich (1986) adopted it to describe “rich-get-richer and poor-get-poorer patterns of reading achievement” in educational psychology. Later empirical research and literature review of Luyten, Cremers-van Wees & Bosker (2003) protested that “in reality there are also examples that support the Matthew effect but also examples that confirm the 'Bigger Fish' effect”, wherein 'Bigger Fish' effect is just the negative Mathew effect – rich-get-poorer and poor-get-richer. It should be noted that their described “Bigger Fish” effects have excluded neither the alternative RTM explanation nor the threat of RTMA. In discussion section we shall distinguish RTMA cases from true negative Mathew effects cases which are without artifacts but still in parallel with RTM.

Comparatively, we now give a substantial interpretation for Matthew effect within VAM context. If it is a common sense that the previous gross learning speed would positively correlate with the current learning speed, we deduct that the higher entrance ability level the faster current growth, or literally Matthew effect. While if the correlation is negative, we deduct “Bigger Fish” effect or negative Matthew effect. Here it is implicit that the scales of scores at different grades are equated and vertically linked, which is necessary for a consistent definition of growth quantities (Lissitz et al, 2006; Doran & Jiang, 2006).

Educational procedures could not only improve ability level of students but also affect their learning speed through training good studying customs, or brewing the long-term motivations, etc. A lot of current VAM quantitative literatures discuss the lasting of the previous value-added contribution in ability level (McCaffrey et al, 2003, p.36-48; McCaffrey et al, 2004) and the comparison between proficiency criterion and growth assessment (Thum, 2006; Kingsbury and McCall, 2006; Rigney and Martineau, 2006), while few, if not none, has regarded the lasting of the former educational contribution in student learning speed. Here we will analyze the complication concerning whether ignoring of the learning speed dimension of student statuses.

Research Question

We do not involve any validity issue of the testing scores. Our starting presumptions are as simplified as possible. The nature of student ability level is assumed to be single-dimensional with a common long-term scale across all concerned grades. The population of the entrance ability levels and their growth speeds ranges a bi-dimensional normal distribution.

In our imagined control setting, each student keeps his current growth speed invariably within all concerned grades. There are rather true student gains than true school effects in our theoretical frame of analysis. While this study has nothing to do with empirical controversy on whether or how much true educational effect exists in the real world (McCaffrey et al, 2003, p.17-36; Raudenbush, 2004).

We do neither bother the real world linking problems of the testings (Reckase, 2004; Kingsbury and Mccall, 2006, p.360). The reliabilities of those tests are also assumed to be same and known. All of the measurement errors are pure white noises.

What's concerned is the admission procedure (Marsh & Hau, 2002) that differentiates schools into five levels. Those schools at the highest level enroll students with the best entrance scores, which means their scores at end of another grade will regress towards the mean most severely while their ability growths at the same time are the highest. Our interests are the results in value-added estimations concerning the school effect.

Finally we should clear what is the value-added estimation and what is school effect. We shall propose several different models, each of which defines its student level gains. Each school's value-added contribution is simply aggregated from gains of its all students. The school effect is defined as the regression coefficient of the average admission scores within each school as a covariate at school level. The meaning of school effect is the marginal average value-added contribution of a school to each student of it along with every unit of its average admission score.

In our setting, the observed value-added contributions include two parts. The first part is from the measurement errors. The errors at any other non-entrance grade will have zero conditional expectation, impair the observed correlation between the entrance testing scores and any other testing scores. So if this first part of school effect is negative, the RTM of errors also impairs the observed correlation between the school quality and student gains. Anyone mistaking this correlation as its population is making an artifact from RTM. So we call it RTMA. The second part is from the differential growth speeds of students embedded in various schools. It is not artifact at least in a descriptive perspective (Rubin et al, 2004), because the aggregated true individual student gains are differential according to schools. If that part of school effect is not zero, we name it Matthew effect and we know that it also causes part of RTM of the total testing score but no artifact in the population of correlation. The fault of the second part is because we mistake a true Matthew effect as a faked school effect in a causal perspective, which has noting to do with its ubiquitous RTM outlook.

The Theoretical Model

Without loss of generality, we model after Marsh and Hau's (2002) Monte Carlo and compare several analytical approaches in estimating the value-added school effect of each school. In the simulation study, 150 schools each with 150 students are examined.

Student-Level

In this model, we try to estimate the value-added school effect of junior high schools, the population value of which is actually set at zero in a causal perspective. An appropriate analytical would be one that provides this correct zero value-added school effect across all schools. Not too surprisingly, it can be seen that a lot of the commonly used models wrongly over- or under-estimates the value-added of some schools.

Specifically, value-added of these junior high schools are defined as the causal impact (i.e., total effect) they have on their students in Grades (G.) 7, 8 and 9, denoted as t (time) = 1, 2, 3. We also assume that we have access to students' performance in their previous three years at the elementary school, i.e., G. 4, 5, 6, denoted as t = -2, -1, 0. All students take an examination at the end of each year from t = -2 to 3, with respective test scores denoted as T_{-2} to T_3, and T_0 being the G. 6 score used for admission into junior high schools.

The scores (T_{-2} to T_3) actually include two parts, one being the latent academic trait (A_{-2} to A_3) and the other being the independent random measurement errors (E_{-2} to E_3). For simplicity, the reliability of the measurements at each time frame is fixed to be identical to α :

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

Thus,

$$\mathrm{Var}(E_t) \equiv \mathrm{Var}(A_t) * \frac{1-\alpha}{\alpha}, t = -2,-1,0,1,2,3 \tag{2}$$

This framework is simpler than the one used in Marsh and Hau (2002) in that the latent academic traits (A_{-2} to A_3) of each student are now set to be strictly linear (linear growth), with D being the difference between any neighboring years. For each individual student, his A_0 and D will totally determine his A_{-2} to A_3 (i.e. $A_t = A_0 + t*D$). We also assume (D, A_0) to have a centered bivariate normal distribution in which D is standardized across all students (SD = 1). From the above description, the joint normal distribution of D and A_0 with a constant test reliability α defines the centered multivariate normal distribution of the observable 6 year successive test scores (T_{-2} to T_3).

To incorporate the Matthew's effect into the model, we denote the correlation of A_0 and D as r, and the Std(A_0) as σ. A positive value of r implies a faster growth for students with higher initial ability A_0. The Cov-Var matrix of A_0 and D is $\begin{pmatrix} \sigma^2 & r \times \sigma \\ r \times \sigma & 1 \end{pmatrix}$. The ($p$, q) element in the 6 × 6 Cov-Var matrix of (T_{-2} to T_3) is,

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

Let p - 3 = q - 3 = t,

$$\text{Var}(T_t) = (\sigma^2 + 2t \times \sigma \times r + t^2) \times \frac{1}{\alpha} = \frac{1}{\alpha}[t - (-\sigma \times r)]^2 + \frac{1}{\alpha}(\sigma^2 - \sigma^2 \times r^2) \quad (4)$$

It can be demonstrated that from and only from $t_{Min} = -r \times \sigma$, Var($T_t$) increases and the Matthew effect manifests. This also means that a positive Matthew effect manifests only from a certain point in the longitudinal study and it becomes negative if traced backwards.

School-level

In assigning elementary students to junior high schools, Marsh and Hau's (2002) procedure is followed to simulate the commonly seen phenomenon that higher ability students clustered in "competitive" schools. Specifically, students are first ranked by their T_0 scores into 5 ascending ability-groups. Schools of five ability bands/types get random samples of students from each student ability-group in a way that the better schools get relatively larger proportions of higher ability students (see Marsh & Hau, 2002).

For the j^{th} junior high school, the mean of its students' ability at the admission (T_0) is aggregated and averaged for each school and is denoted as Q_j. In hierarchical level terminology (Raudenbush & Bryk, 2002), each student i is nested in the j^{th} high school with the mapping j=J(i). At the student level, one more subscript is needed, e.g., $T_{0,i}$ means the T_0 score of the i^{th} student, with $Q_{J(i)}$ being the school quality effecting on this student.

Analytical Models

Unconditional Models

Our unconditional models share a common form and differ only by their respective definition of the individual student gain, denoted as V.

At the student level, for the i^{th} student in the j^{th} school, j = J(i),

$$V_i = \beta_{J(i)} + e_i, \qquad i = 1 \text{ to } 150 \times 150,\ e_{(.)} \sim N(0, \sigma^2) \quad (5)$$

At the school level, for the j^{th} school,

$$\beta_j = \gamma_0 + \gamma_1 * Q_j + u_j, \quad j = 1 \text{ to } 150, u_{(.)} \sim N(0, \tau) \tag{6}$$

Putting these two levels of equations together:

$$V_i = \gamma_0 + \gamma_1 * Q_{J(i)} + u_{J(i)} + e_i, \tag{7}$$

As schools are constructed with no specific effect on individual students at the population level, a significant non-zero estimation of γ_1 would indicate a wrongly estimated value-added school effect at the school level and hence an inappropriate statistical model to be adopted in a causal perspective.

Conditional Models

There is only one approach of the conditional model, wherein V is defined simply as $V_i = T_{3,i} - T_{0,i}$.

$$\begin{aligned}
&V_i = \beta_{0,J(i)} + \beta_{1,J(i)} * T_{0,i} + r_i \quad ,i = 1 \ldots N \\
&\quad r_i, i = 1 \ldots N \quad \sim N(0, \sigma^2) \\
&\beta_{0,j} = \gamma_{00} + \gamma_{01} Q_j + u_{0,j} \quad ,j = 1 \ldots M \\
&\beta_{1,j} = \gamma_{10} + \gamma_{11} Q_j + u_{1,j} \quad ,j = 1 \ldots M \\
&\begin{pmatrix} u_{0,j} \\ u_{1,j} \end{pmatrix}, j = 1 \ldots M \quad \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix})
\end{aligned} \tag{8}$$

The model is identical to the following, with the non-zero estimates and significance of γ_{01} and γ_{11} being our key interest,

$$\begin{aligned}
&V_i - \gamma_{10} T_{0,i} = \gamma_{00} + \gamma_{01} Q_{J(i)} + \gamma_{11} Q_{J(i)} T_{0,i} + u_{1,J(i)} * T_{0,i} + u_{0,J(i)} + r_i \quad ,i = 1 \ldots N \\
&\quad r_i, i = 1 \ldots N \quad \sim N(0, \sigma^2) \\
&\begin{pmatrix} u_{0,j} \\ u_{1,j} \end{pmatrix}, j = 1 \ldots M \quad \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix})
\end{aligned} \tag{9}$$

Remedies

To prevent RTMA in the unconditional strategy, direct incorporation of the admission score (T_0) in the value-added equation should be avoided. Whenever T_0 is necessary in the V calculation, it should be replaced by T_1 or T_{-1} (see Column 3, comparing Rows 2 and 1, Rows 4 and 3 in Table 1).

To take care of the Matthew effect, instead of comparing the intake and exit (T_3) scores, value-added should be estimated by comparing the growth rates in elementary and junior high schools (see Column 3, comparing Rows 3 and 1, Rows 4 and 2).

The correlations or independence between value-added (V) and observable school level intake (Q) are thoroughly investigated. Although we have no close form expression of this correlation, we have proved that Cov(V,Q) and Cov(V,T_0) share an identical sign (+, 0, or -, shown in full manuscript), results partly shown in Table 1. With these expressions relating the parameters r, σ *and* α, the critical set of parameters can also be determined.

Simulation Study

A simulation study is used to display the differences in approaches and confirm the above mathematical derivations. A total of 150 schools, each with 150 students ($N = 150 \times 150$) are simulated and analyzed with the multilevel regression model using SAS PROC Mixed (Singer, 1998) and SAS Marco (Fan, Felsovalyi, Stephen, Sivo & Keenan, 2002). At the student level, the samples of each variable are generated according to their multiple normal distribution with close form expression.

Result

The results of the unconditional approaches are summarized in Table 1. Cov(V, T_0) indicates the sign (+, -, 0) of the population value of the regression coefficient γ_1 as in Equation (7).

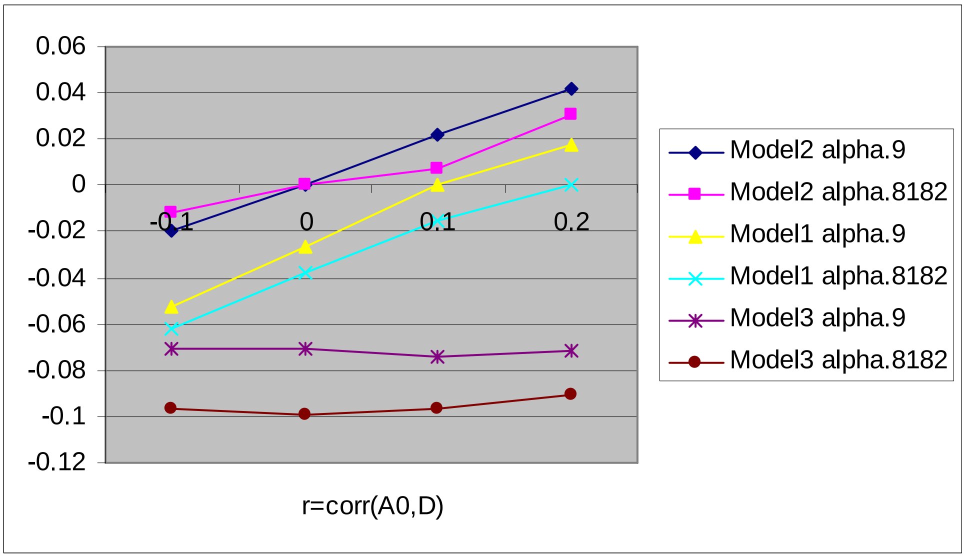
Table 1 Characteristics and Simulation Results of the Unconditional Approaches

Models	**Description**	**How Value-added is operationally measured**	**Cov(V,T_0)**	**Problems (without control for)**
1. Comparison of Entry (T_0) and Exit performance	Take into consideration preexisting differences, most popular model	$V_i = T_{3,i} - T_{0,i}$ $= 3D - E_0 + E_3$	$\sigma^2(3\frac{r}{\sigma} - \frac{1-\alpha}{\alpha})$	RTMA, Matthew Effect
2. Comparison of end of Year 1 (T_1) and Exit Performance	Comparison between entry (pre) and exit (post) but avoid the regression towards mean effect by using the T_1 instead of T_0 as "pre" measure	$V_i = T_{3,i} - T_{1,i}$ $= 2D + (E_3 - E_1)$	$2r\sigma$	Matthew Effect
3. Comparison of primary and high school growth rates	Take into consideration the possible growth trajectories of individual students	$V_i = \frac{T_{3,i} - T_{0,i}}{3} - \frac{T_{0,i} - T_{-2,i}}{2}$ $= -\frac{5}{6}E_0 + (\frac{E_3}{3} - \frac{E_{-2}}{2})$	$-\frac{5}{6}\sigma^2 \times \frac{1-\alpha}{\alpha}$	RTMA
4. Comparison of primary and high school growth rates using T_1	Take into consideration individual growth trajectories and avoid RTMA using T_1 instead of T_0 as "pre" measure; this is the only appropriate (best) model to be adopted	$V_i = \frac{T_{3,i} - T_{1,i}}{2} - (T_{-1,i} - T_{-2,i})$ $= \frac{E_{3,i} - E_{1,i}}{2} - (E_{-1,i} - E_{-2,i})$	0	

With further mathematical derivations to the condition model, in Equation (9), V could be also be decomposed as $V = (V - b \times T_0) + b \times T_0$, where b is chosen such that Cov(($V - b \cdot T_0$), T_0) = 0. It could be proved (details in the full manuscript) that $V - b \cdot T_0$ is independent of $Q_{J(.)}$ and $Q_{J(.)} \cdot T_0$. So the populations of estimates should be $\gamma_{10}=b$; $\gamma_{01}=\gamma_{11}=0$, no matter whether the Matthew effect exists or not.

From the Cov(V, T_0) in Table 1, we select some critical sets of parameters for the simulation study (see Figure 1), with σ fixed at 2.7; r is set at -0.1, 0, 0.1, 0.2 along the x-coordinate; α is set at 0.9000 and 0.8182. The y-coordinate indicates the regression estimation of γ_1 in equation (7) relative to the base Std(V), that is, $\frac{\gamma_1}{Std(V)}$. Results show that all the 8 estimates in Model 4 (not shown) are appropriately lying on the x-axis, while the estimates of other models (see Figure 2) are wrongly estimated to be significantly non-zero, $p < .0005$.

Figure 1 Estimates of $\frac{\gamma_1}{Std(V)}$ using the unconditional approach



Discussion and Conclusion

There are two layers of traps in our analytic frame. The first is at the descriptive perspective layer. The second is at the causal perspective layer. Let's back to the first layer and forget the causal term "school effect", just concentrate on the descriptive correlation between school quality $Q_{J(.)}$ and student gain $V_{(.)}$ in the simplest Model 1.

If reliability α is unit one and $r<0$, we would only face the negative Mathew effect. In this case, the observed $V_{(.)}=3D$ is the population value without any measure error. Then $\mathrm{Corr}(T_3,T_0)<1$ and the $\mathrm{RTM}(T_3,T_0)>0$. While we do not have any descriptive artifact from this RTM. The statement of $\mathrm{Corr}(V_{population},Q_{population})<0$ is true. Now considering someone wrongly expect $E(T_3|T_0)= T_0$, and then conclude that $\mathrm{Cov}(V_{population},Q_{population})=0$, he is just taking a kind of RTMA but ignoring the true "Big Fish" effect in a descriptive perspective. Here we highlight that RTMA is always subjective and relative to a certain wrong intuition mistaking some RTM as zero, or some impaired correlation as its population. While (negative) Mathew effect is an absolute fact in a descriptive perspective. Moreover, in Mathew effect we compare two vertically linked testing scores while in the objective RTM (not any subjective RTMA) we compare the standardized z-scores of two variables.

Remembering that RTMA in our method section is relative to a special wrong intuition of $\mathrm{Corr}(V_{population},Q_{population})= \mathrm{Corr}(V,Q)$. Let's imagine the paradox that reliability α changes from .9 to unit one but the same time the traits instability is set to a fitted case making the multiple-dimensional distribution invariable. Then it becomes true that $\mathrm{Corr}(V_{population},Q_{population})=\mathrm{Corr}(V,Q)$ and the RTMA disappears automatically. Let's presume $D=0$ to highlight the paradox. The reason is in the relativity of the definition of RTMA. Anyone mistaking $E(T_3|T_0)= T_0$ is just making his version of RTMA now. So the answer is only some special kind of RTMA could be remedied by reducing measurement errors.

The setting of the current discussion is just the one to explain the distinction between RTMA and causal complication of Lord paradox (Lord, F., 1967; Wainer, 1991; Rubin et al, 2004; Holland, P. & Rubin, D., 1983). Now, no measure error and no Mathew effect, only the special traits instability makes the descriptive statement $\mathrm{Corr}(V_{\mathrm{population}},Q_{\mathrm{population}})= \mathrm{Corr}(V,Q)<0$ true. There is at least one case that the causal conclusion school effect<0 could be validated. If only it is just school education that is undermining the stability of testing traits. A more explicit statement is, if there would be extreme trait stability without school education, we could conclude that school has a causal effect proportion to its average admission score. Even if the causal effect is confirmed by Rubin's (1974) criterion, it is still distinct from the literal meaning of school effect of every unit of average admission score (this case could be compared to Raudenbush, 2004). Here we highlight that Lord paradox is also relative like RTMA. Their distinction is that Lord paradox relies on usually unknown but objective alternative implication to make causal conclusion/assumption while RTMA relies on some subjective wrong intuition to form descriptive artifacts.

The trait instability also decides whether RTM lasts grade by grade. The mathematical deduction is that $[Z_{T2}-\mathrm{E}(Z_{T2}|T_0)]-[Z_{T1}-\mathrm{E}(Z_{T1}|T_0)]= \mathrm{Corr}(T_2,T_0)-\mathrm{Corr}(T_1,T_0)$. That means, although in our setting with extreme traits stability the RTM lasts only at one other grade, in most real setting, the correlation will decrease according to time span, so the RTM will last grade by grade. This conclusion highlight that RTM is generally related to time span, but not to time direction. It means the remedy of Model 2 for RTMA in a real world setting will give a true descriptive correlation and leave a significant causal conclusion to Lord paradox.

in a descriptive perspective.

As to the remedy of Model 3 for (negative) Matthew effect, the essential issue is the multi-dimensional nature of the effect of educational procedure, no matter how perfect single dimension of the student ability level could reach (see Kingsbury & McCall, 2006, while it has been severely criticized by Reckase, 2004). If most common students ability curves plotted above a grade/time x-axis are parabolas rather than a linear line, vertically linking forces may transform the scare of the y-axis to make most students' curves linear. Such a fact seems that the Mathew effect problem might be reduced to some technical debate of measurement tool quality. For example, it would be a quick solution if someone devices a tool directly for measurements of learning speeds rather than of ability levels. However, it still indicates there is always higher but not currently accountable educational objectives that schools or teachers might hesitate to fulfill with limited resource constraints and a strong but fixed-aim accountability system.

References

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

Doran, H., & Jiang, T. (2006). The Impact of Linking Error in Longitudinal Analysis: An Empirical Demonstration. In Lissitz, R. (Eds.), *Longitudinal and Value Added Models of Student Performance* (pp. 210-229). Maple Grove, MN: JAM Press

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

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland*, 15, 246-263.

Holland, P. & Rubin, D. (1983). On Lord's Paradox. In H. Wainer, & S. Messick, (Eds.), *Principals of modern psychological measurement*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kingsbury, G. & McCall, M. (2006). The Hybrid Success Model: Theory and Practice. In Lissitz, R. (Eds.), *Longitudinal and Value Added Models of Student Performance* (pp. 346-384). Maple Grove, MN: JAM Press

Lissitz, R. (2005). Foreword and Afterword. In Lissitz, R. (Eds.), *Value Added Models in Education: Theory and Applications* (pp. iv-vii; pp. 304-306). Maple Grove, MN: JAM Press

Lissitz, R. (2006). Foreword. In Lissitz, R. (Eds.), *Longitudinal and Value Added Models of Student Performance* (pp. iv-vii). Maple Grove, MN: JAM Press

Lissitz, R., Doran, H., Schafer, W. & Willhoft, J. (2006). Growth Modeling, Value-Added Modeling and Linking: An Introduction. In Lissitz, R. (Eds.), *Longitudinal and Value Added Models of Student Performance* (pp. 1-46). Maple Grove, MN: JAM Press

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, 685, 304-305.

Luyten, H., Wees, C.-V., & 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, 245-282.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: The RAND Corporation.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, *29*, 67-101.

Merton, Robert K. (1968), The Matthew Effect in Science. *Science*, *159,* (3810), January 5: pp. 56-63.

Merton, Robert K. (1988), The Matthew Effect in Science, II: Cumulative Advantage and the Symbolism of Intellectual Property. *Isis, 79*: pp. 606-623.

U. S. (2001). No Child Left Behind Act. Retrieved November 13, 2004, from http://www.whitehouse.gov/news/reports/no-child-left-behind.html

Raudenbush, S. W. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral*

Statistics, *29*, 121-129.

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

Reckase, M. D. (2004). The real world is more complicated than we would like. *Journal of Educational and Behavioral Statistics*, *29*, 117-120.

Rigney, S., Doran & Martineau, J. (2006). NCLB and Growth Models: In Conflict or in Concert? In Lissitz, R. (Eds.), *Longitudinal and Value Added Models of Student Performance* (pp. 47-82). Maple Grove, MN: JAM Press

Roberts, A. O. H. (1980). Regression toward the mean and the regression-effect bias. In G. Echternacht (Ed.), *Measurement aspects of Title I evaluations* (pp. 59-82). San Francisco, CA: Jossey-Bass.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, *66*, 688-701.

Rubin, D. B., Stuart, E. A., & Zanutto, E. L. (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics*, *29*, 103-116.

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, 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, 323-355.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21(4)*, 360-407.

Thum, Y. M. (2006). Designing Gross Productivity Indicators: A Proposal for Connecting Accountability Goals, Data, and Analysis. In Lissitz, R. (Eds.), *Longitudinal and Value Added Models of Student Performance* (pp. 436-480). Maple Grove, MN: JAM Press

Wainer, H. (1991). Adjusting for Differential Base Rates: Lord's Paradox Again. *Psychological Bulletin, 109(1)*, 147-151.

Wainer, H. (2004). Introduction to a Special Issue. *Journal of Educational and Behavioral Statistics*, 29(1), 1-3.

Wright, S., Sanders, W. & Rivers, J. (2006). Measurement of Academic Growth of Individual Students toward Variable and Meaningful Academic Standards. In Lissitz, R. (Eds.), *Longitudinal and Value Added Models of Student Performance* (pp. 385-406). Maple Grove, MN: JAM Press