Confusions about Consistency in Improvement

David Rogosa
Stanford University
June 2003

Acknowledgements
Thanks to Ghassan Ghandour for the usual computing wizardry and to Ed Haertel, Matt Finkelman, and the students in Education 351 for helpful comments.
The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B60002 and Award Number R305B960002-01 as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The findings and opinions expressed in this report do not reflect the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the Office of Educational Research and Improvement, or the U.S. Department of Education. Additional support was provided by the California Department of Education, Policy and Evaluation Division.

Confusions about Consistency in Improvement

David Rogosa

Contents

	pdf page
1. Introduction	5
1.1. LH and KS Assessment of Volatility And Stability	8
1.2. Some introductory examples	9
1.3. Representations of stability (for school accountability systems)	12
2. Common Sense Data Analysis for Consistency of Improvement	17
3. Analytic Results for LH and KS Stability/Volatility Measures	33
3.1 Results for LH and KS using the Constant Rate of Change Model	35
3.2 Results for LH and KS using the Proportional Deceleration Model	39
4. Artificial Data Examples for Consistency of School Improvement	42
4.1 Artificial Data Example 1: LH, KS Volatility	44
4.2 Artificial Data Example 2: LH, KS Stability	51
4.3 Artificial Data Example 3: Up-and-Down, LH, KS stability	57
4.4 Artificial Data Example 4: Proportional Deceleration, LH, KS Volatility	63
5. Discussion	68
References	71
Appendix A: Additional Data Displays, Data Display Archive	73
Appendix B: Algebraic Derivations for Section 3 Results	78

Confusions about Consistency in Improvement

David Rogosa

List of Figures and Tables

pdf page

Figure 1.1. Depictions of Consistency of Individual School Improvement 14
Figure 1.2. 2x2 Array of Stability of Improvement and Stability of Between-school Differences for Collection of Schools. 15

Figure 2.1. Scatterplot of improvement in school API scores '00-01 versus improvement '99-00. 23
Figure 2.2. Set of three scatterplots for four-year improvement in school API scores corresponding to the bottom portion of Table 2.2. 25
Figure 2.3. Scatterplot of improvement in SD subgroup API scores '00-01 versus improvement '99-00. 29
Figure 2.4. Set of three scatterplots for four-year improvement in SD subgroup API scores corresponding to the bottom portion of Table 2.3. 30
Figure 2.5. Scatterplot of improvement in SD subgroup in High SD schools API scores '00-01 versus improvement '99-00. 31
Figure 2.6. Set of three scatterplots for four-year improvement in SD subgroup in High SD schools API scores corresponding to the bottom portion of Table 2.4. 32

Figure 3.1 Population values of r_{LH} and ρ_{KS} as a function of $\rho[D_t]$ for constant rate of change model. 37
Figure 3.2 Population values of r_{LH} and ρ_{KS} as a function of $\rho[D_1]$ for proportional deceleration model with h = 9/10, 2/3, 1/2 . 40

Figure 4.1. Scatterplots corresponding to the KS and LH correlation coefficients for artificial data example 1. 46
Figure 4.2. Scatterplots corresponding to the KS and LH correlation coefficients for artificial data example 2. 53
Figure 4.3. Scatterplots corresponding to the KS and LH correlation coefficients for artificial data example 3 (up-and-down examples). 59
Figure 4.4. Scatterplots corresponding to the KS and LH correlation coefficients for artificial data example 4. 65

Exhibit 1. Consistency In Improvement Caricature 69

pdf page

Table 2.1. Descriptive Statistics for School-Level Longitudinal Data and Year-to-Year Improvement: California API Index for Elementary Schools. 19
Table 2.2. Consecutive Improvement for California Elementary Schools 21
Table 2.3. Consecutive Improvement for SD Subgroups in California Elementary Schools. 27
Table 2.4. Consecutive Improvement for SD Subgroups in High SD Elementary Schools. 28

Table 3.1. Stability Measures for Constant Rate of Change Model 38
Table 3.2. Stability Measures for Proportional Deceleration Model 41

Table 4.1. Descriptive Statistics for School-Level Longitudinal Data and Year-to-Year Improvement: Artificial Data Example 1 45
Table 4.2. Consecutive Improvement for Artificial Data Example 1: Continuous Outcome 48
Table 4.3. Consecutive Improvement for Artificial Data Example 1: Proportion Outcome. 49
Table 4.4. Descriptive Statistics for School-Level Longitudinal Data and Year-to-Year Improvement: Artificial Data Example 2 52
Table 4.5. Consecutive Improvement for Artificial Data Example 2: Continuous Outcome 55
Table 4.6. Consecutive Improvement for Artificial Data Example 2: Proportion Outcome 56
Table 4.7. Descriptive Statistics for School-Level Longitudinal Data and Year-to-Year Improvement: Artificial Data Example 3 58
Table 4.8. Consecutive Improvement for Artificial Data Example 3: Continuous Outcome 60
Table 4.9. Consecutive Improvement for Artificial Data Example 3: Four Year Example, Proportion Outcome 62
Table 4.10. Descriptive Statistics for School-Level Longitudinal Data and Year-to-Year Improvement: Artificial Data Example 4 64
Table 4.11. Consecutive Improvement for Artificial Data Example 4: Continuous Outcome 66

Confusions about Consistency in Improvement

David Rogosa
Stanford University

ABSTRACT

The consistency in improvement of school-level scores from testing programs has attracted considerable interest. Claims of "volatility" in the scores by Linn and Haug (2002) and by Kane and Staiger (2002) represent a serious threat to defensible policy uses of test scores in school accountability systems. In this paper analytic results and artificial data examples are used to demonstrate that the methodology employed by both Linn-Haug and Kane-Staiger has absolutely no value and that their empirical demonstrations of volatility (or lack of stability) have no credibility. In contrast, useful data analysis displays for consistency in improvement are presented using California API data and artificial data examples.

1. Introduction

Linn and Haug (2002) in "Stability of School-Building Accountability Scores and Gains" (hereafter LH) present a correlational methodology for determining the "stability" or "volatility" of year-to-year improvement, and LH apply that methodology to school-level scores (successive fourth graders) from four years (1997-2000) of Colorado assessment data (CSAP). As LH heavily cite and build upon similar methodology and assertions of volatility contained in Kane and Staiger (2002) "Volatility in School Test Scores: Implications for Test-Based Accountability Systems" (hereafter KS), this paper considers the methodology and empirical claims in both LH and KS.

The crux of the failure of both LH and KS is a lack of understanding of what their correlational evidence actually indicates. And this disconnect between their statistical analyses and its appropriate interpretation leads to the mis-attribution of volatility throughout LH and KS. Although in most settings LH and KS will be shown to be wildly pessimistic about consistency in improvement, in other settings LH and KS will be seen to declare great stability in the face of stunning lack of consistency in improvement.

Given the primacy of proper interpretation of statistical evidence, it is important to review a sampling of the assertions made by LH:

> "Year-to-year changes in scores for successive groups of students have a great deal of volatility" (p.29),
> "It was found that the year-to-year changes are quite unstable" (p.29),
> "Moreover, regardless of starting position, schools that gain a lot from year one to year two will generally show a decline in year three, while those that show a decline from year one to year two will generally show a gain in year three." (p.33)
> "The estimates of improvement, however, are quite volatile. This volatility results in some schools being recognized as outstanding and other schools identified as in need of improvement simply as the result of random fluctuations." (p.35)

Similarly, the assertions of volatility in KS would seem to indicate a ubiquitous lack of consistency in improvement. A sampling of KS assertions:

> "We highlight an underappreciated weakness of school accountability systems—the volatility of test score measures—and explore the implications of that volatility for the design of school accountability systems." (p.235)
> "In North Carolina, ... between 50 and 80 percent of the variance in the change in mean fourth-grade scores is nonpersistent. If one were to look for signs of improvement by closely tracking changes in mean scores from one year to the next, 50 to 80 percent of what one observed would be temporary—either due to sampling variation or some other nonpersistent cause." (p.248)

And California, according to KS, is even worse. Almost all improvement is found to be transient or "fleeting":

> "Although the California schools tend to be larger, the data reveal slightly more volatility in the California Academic Performance Index for any given school size. For the smallest fifth of schools, the correlation in the change in adjacent years was –.43, implying that 86 percent of the variance in the changes between any two years is fleeting. For the largest fifth of schools, the correlation was –.36, implying that 72 percent of the variance in the change was nonpersistent." (p.248-9)

This paper has two purposes: first and foremost, to provide tools to detect and describe consistency of school improvement in large-scale assessments, and second, to warn readers away from methods (LH and KS) that falsely promise determinations of stability or volatility. In the remainder of this Introductory Section, the LH and KS empirical measures of stability are presented (section 1.1), small preliminary examples are used to indicate some of the failures of LH and KS methods (section 1.2), and basic representations of stability and volatility are presented (section 1.3). Section 2 develops useful data displays for consistency of school improvement using the California API data. In addition to the basic displays, a conditioning strategy for subgroups, in the spirit of "disaggregation of results," is presented. Section 3 provides the technical results on the properties of the LH and KS measures. Section 4 follows up on those analytic results with a series of large artificial data examples, illustrating for example, that LH and KS will find near 100% volatility in the face of large consistency of improvement (sec 4.1) and, conversely, that LH and KS will determine high stability in the presence of up-and-down trajectories (sec. 4.3).

1.1. LH and KS Assessment of Volatility And Stability

LH procedure.

In their section "Volatility of Change Scores" LH explain: "To investigate this lack of stability of change scores with the CSAP data, we computed change scores based on two-year intervals ... The correlation between change 97 to 99 and change 98 to 00" (p.33). The LH application of this methodology is to scores on "fourth grade reading results for schools in Colorado for four years of administration of the Colorado Student Assessment Program" (p.29) using for the school summary score either the proportion proficient or a "weighted index of all [four] performance levels" (p.31).

To adopt a uniform notation, let the score for school *s* at time t_i be written as $Y_s(t_i)$. Thus for the four successive yearly measurements the school scores are $Y_s(t_1)$, $Y_s(t_2)$, $Y_s(t_3)$, $Y_s(t_4)$. (As the yearly scores can be regarded as taken at equally-spaced integer time-points {1,2,3,4,...}, for convenience, the times of measurement will be often denoted by $t_1 = 1$, $t_2 = 2$, $t_3 = 3$, $t_4 = 4$.) Then the LH stability measure, the sample correlation coefficient denoted by r_{LH}, is written

$$r_{LH} = \text{Correlation}[\ Y_s(t_3) - Y_s(t_1),\ \ Y_s(t_4) - Y_s(t_2)]\ \ .$$

Importantly, r_{LH} is interpreted as the measure of "stability in the two-year change scores" (LH, p.33). This correlation can also be expressed in terms of the obvious difference scores. Denote the observed improvement for school *s* in the time interval (t_1, t_2) by the difference score $D_s[t_1, t_2] = Y_s(t_2) - Y_s(t_1)$, then for the LH stability measure

$$r_{LH} = \text{Correlation}[D_s[t_1, t_3], D_s[t_2, t_4]\ ]\ .$$

KS procedure.

The KS procedure uses three successive years of data from non-overlapping samples, such as fourth grade scores for three years. The KS applications are to "mean fourth-grade scores in North Carolina (combining the scaled scores for math and reading)" for 1997, 1998, 1999 and "also ... the mean Academic Performance Index scores in California for fourth-grade students... in adjacent years" [presumably 1999, 2000, 2001] (KS, p.248). KS compute the sample correlation:

$$r_{KS} = \text{Correlation}[\ Y_s(t_2) - Y_s(t_1),\ \ Y_s(t_3) - Y_s(t_2)]\ .$$
$$= \text{Correlation}[D_s[t_1, t_2], D_s[t_2, t_3]\ ]\ .$$

According to KS, multiply that correlation by –2 to obtain "the proportion of the change in test scores that is attributable to nonpersistent factors" (KS, p.247). (Note that the proportion range of 0 to 1 corresponds to r_{KS} between 0 and –.5; presumably KS would truncate proportion nonpersistent to 0 for $r_{KS} > 0$, and to 0 for $r_{KS} < -.5$.) In order to orient this measure in the same direction as the LH coefficient, define a proportion persistent (i.e., the proportion not nonpersistent) measure p_{KS} as $1 - (-2r_{KS})$ so that the sample quantity

$$p_{KS} = 1 + 2r_{KS}$$

serves as a measure of lack of volatility (stability). (Again, truncate p_{KS} to 1.0 for $r_{KS} > 0$, and to 0 for $r_{KS} < -.5$.)

1.2. Some introductory examples

Before proceeding to the analytic results for the LH and KS procedures in Section 3 and the accompanying artificial data examples in Section 4, some very small examples are introduced. Perhaps these examples will motivate readers who previously accepted the KS and LH assertions as credible to study the more detailed results in the later sections.

Example A: Extreme Volatility? Consider a set of 5 schools with four years of test data (scores are in the California API scale so year-to-year improvement of 50 points is very strong). The four years of API data produce the following results for year-to-year improvement.

	Improvement		
School	Yr1toYr2	Yr2toYr3	Yr3toYr4
A	40	50	50
B	40	40	50
C	50	40	40
D	59	41	49
E	45	45	45

Each school has healthy improvement each year of approximately the same amount. School officials and parents in these schools would cheer. Yet, remarkably, LH would determine 0% stability, 100% volatility for these school scores ($r_{LH} = 0$). KS, whose methods use the first three years of data, would determine p_{KS} = .062, indicating 94% of change nonpersistent, again extreme volatility! (Of course the same results hold for an example with 500 or 5000 schools following this pattern, and statistical properties of these scores are not at issue as the schools could be considered to be very large schools where standard errors of the yearly scores would be 1 or 2 points.) A lingering question from this example is: If the above represents "volatile", then can consistent, stable improvement ever be identified?

Example B: LH volatility, KS persistence. Although LH regard the KS methods as a kindred effort, the two assessments of volatility can violently disagree. The following example of five schools again shows for each school healthy improvement each year of approximately the same amount.

	Improvement		
School	Yr1toYr2	Yr2toYr3	Yr3toYr4
A	40	50	50
B	40	40	50
C	50	40	40
D	50	50	40
E	45	45	45

As r_{LH} = 0, LH would (again) determine 0% stability, 100% volatility for these school scores. However, KS, whose methods use the first three years of data, would obtain r_{KS} = 0 and thus p_{KS} = 1, indicating 0% volatility.

Example C: LH stability, KS volatility. Conversely, a slightly different example of five schools, which once again shows for each school healthy improvement each year of approximately the same amount, reverses the disagreement.

	Improvement		
School	Yr1toYr2	Yr2toYr3	Yr3toYr4
A	40	50	42
B	40	40	35
C	50	40	48
D	50	42	53
E	45	45	45

Here r_{LH} = .977, and LH would determine great stability for these school scores. However, KS, whose methods use the first three years of data, would obtain r_{KS} = –.474 and thus p_{KS}= .052, indicating extreme volatility. Take your pick?

Example D: Up-and-Down, LH stability. LH draw conclusions (c.f. p.33, 35) regarding patterns of scores in which schools initially improve and then decline, thus giving up the gains. In the example below, schools initially improve, then flatten out, then decline, such that the overall improvement over the four years is exactly 0.

	Improvement		
School	Yr1toYr2	Yr2toYr3	Yr3toYr4
A	39	7	–47
B	39	–3	–54
C	49	–3	–41
D	49	–1	–36
E	44	2	–44

Yet for these schools r_{LH} = .977, and thus LH would determine great stability for these year-to-year improvements school scores! In contrast, KS, whose methods use only the first three years of data, would obtain r_{KS} = –.474 and thus p_{KS}= .052, indicating extreme volatility. But before crediting KS methods with good behavior here note that only year 1 through year 3 data is employed (in the example the increase plus leveling off) so that the KS procedure would not include the information about the decline (Yr3toYr4). Whether the KS procedure is prescient is unclear, but it is clear that the LH procedure is oblivious to extreme inconsistency in improvement. An up-and-down example for the KS procedure is given below.

	Improvement	
School	Yr1toYr2	Yr2toYr3
A	40	–40
B	40	–50
C	50	–50
D	50	–40
E	45	–45

Even though the initial improvements in year 1 to year 2 are erased by the declines year 2 to year 3, leaving overall improvement from year 1 to year 3 of exactly 0, KS would obtain 0% volatility; r_{KS} = 0 and thus p_{KS}= 1.

The artificial data examples in Section 4 reiterate the patterns of confusion indicated by these small examples, employing data for 10000 (hypothetical) schools. Furthermore, the analytic results in Section 3 provide some explanations for the bizarre behaviors of the KS and LH methodologies.

1.3. Representations of stability (for school accountability systems)

The final preliminary is to provide some simple depictions of what is meant by the terms: stability, consistency of improvement, lack of stability (volatility). After all, any useful statistical summary must be based on some clear idea of what is being estimated or described. And as stated above, the critical failing in LH (and KS) is the chasm of confusion between their correlational data summary and the asserted interpretation of volatility.

Good common sense in talking about stability has a history in educational psychology that unfortunately has been lost in the current LH and KS work on educational testing. Most notable is the clear thinking in Wohlwill (1973, esp. Ch. 12) in distinguishing between stability of individual processes and stability of individual differences. Building upon Wohlwill's presentation, Rogosa, Floden and Willett (1984) developed statistical procedures (the application being to classroom observation data) for assessment of temporal stability organized around the two main research questions: (i) temporal stability of an individual unit; (ii) temporal stability of individual differences between units (Rogosa et. al. 1984, p.1000).

In the setting of educational assessment and school accountability, the only question of interest is in the temporal stability of the individual school, e.g., consistency of improvement. In accountability systems, the question is:
  Did this school improve?
or more specifically in the current formulation of accountability systems,
  Did this school improve *enough*?
where the word "enough" merits emphasis because it encompasses both quantity and quality of improvement. Typically, the quantity of improvement is stated in terms of a "growth target" or "adequate yearly progress" and the qualitative aspects (think of "quality of earnings" in corporate financial reports) arise from the imposition of growth targets on identified subgroups of students in addition to the entire school.

Figure 1.1 provides depictions of instances of stable and unstable (volatile) school trajectories (smooth trajectories shown assuming no statistical error in the scores). The upper portion of Figure 1.1 shows representations of improvement for a single school, contrasting consistent improvement in which a school continues to improve over consecutive years (frames a and b) with a school exhibiting an up-and-down pattern (frames c and d). LH assert that the up-and-down (and/or down-and-up) pattern may pertain when their indices of stability are small. The constant rate of improvement depiction (frame a) is the clearest representation of consistent improvement for an individual school. The proportional deceleration depiction (frame b) also indicates consistent year to year improvement, but of decreasing magnitude. Deceleration is not decline.

Insert Figure 1.1

State-wide testing programs produce large collections of individual school trajectories. Some exemplary patterns of individual trajectories are shown in Figure 1.2 ("2x2 Array of Stability of Improvement and Stability of Between-school Differences"). Of special note is frame (a) depicting consistent (stable) individual school improvement coupled with stable between-school differences (rank order in school scores maintained over time). Stunningly, this frame (a) represents a setting for which both LH and KS will determine 0% stability, 100% volatility (see results in section 3). On the other hand, frame (b), a configuration with consistent improvement for each school, but for which between-school differences are unstable over time, is a setting for which LH and KS will find strong stability (small volatility). (Data conforming to the configuration in frame b having score reliability around .9 will yield values 1.0 for p_{KS} and about .8 for r_{LH} .) In Wohlwill's depiction (1973, Figure 12-7) Figure 1.2 frame (a) is stable "preservation of individual differences" and frame (b) is unstable or fluctuant "preservation of individual differences". Assessments of the stability of individual differences have a long history in biomedical research under the heading of *tracking* (c.f., Rogosa, and Willett, 1983a; Rogosa et.al. 1984).

Insert Figure 1.2

Perhaps even more strangely, LH and KS procedures will indicate strong stability (small volatility) for settings like frame (d) in which individual improvement is unstable (up-and-down trajectories) and between-school differences are not stable (rank order in school scores not maintained over time). These unfortunate properties of r_{LH} and p_{KS} are demonstrated in detail in artificial data example 3 (section 4.3). On the other hand, a pattern of up-and-down school trajectories for which between-school differences are maintained (frame c) will produce determinations of 100% volatility by LH and KS. Comparing within each column in the 2x2 array, the LH and KS procedures cannot distinguish between frames a and c (indicating zero stability for both), nor can they distinguish between frames b and d

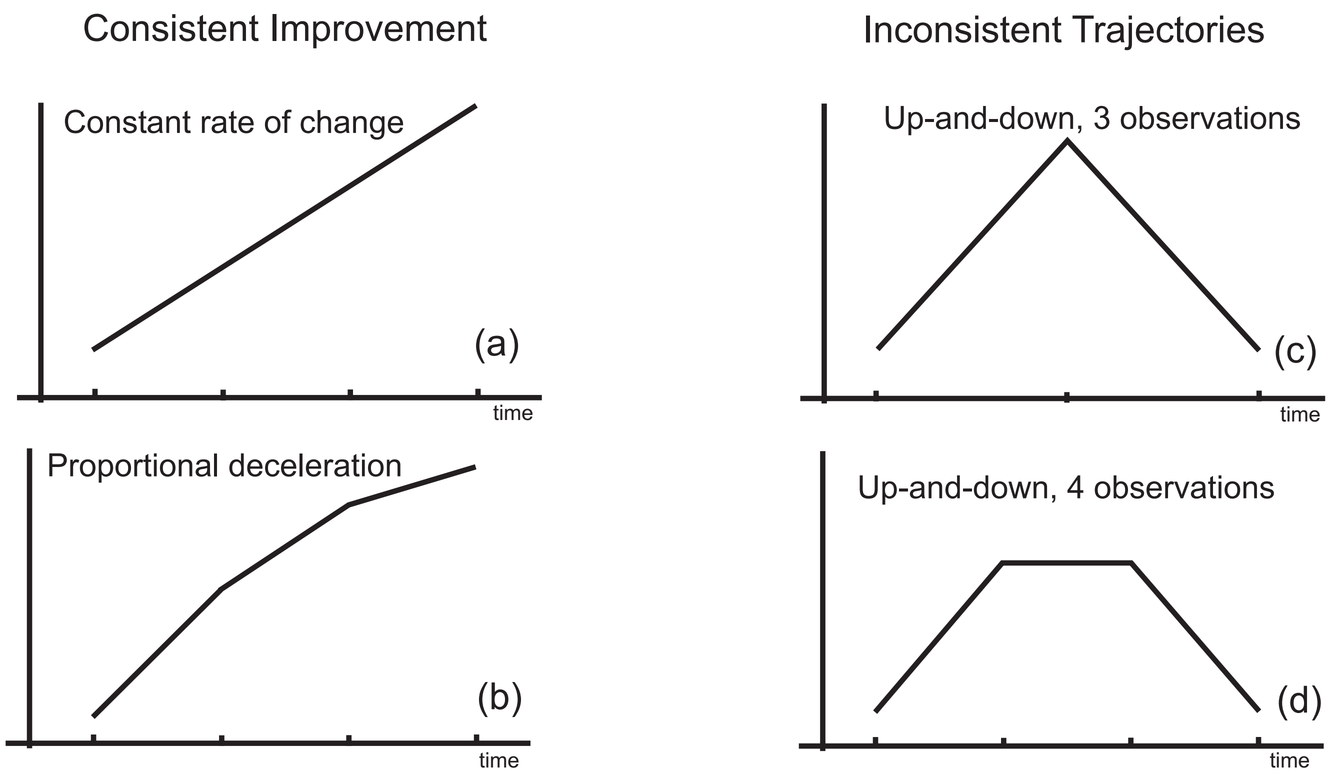


Figure 1.1. Depictions of Consistency of Individual School Improvement

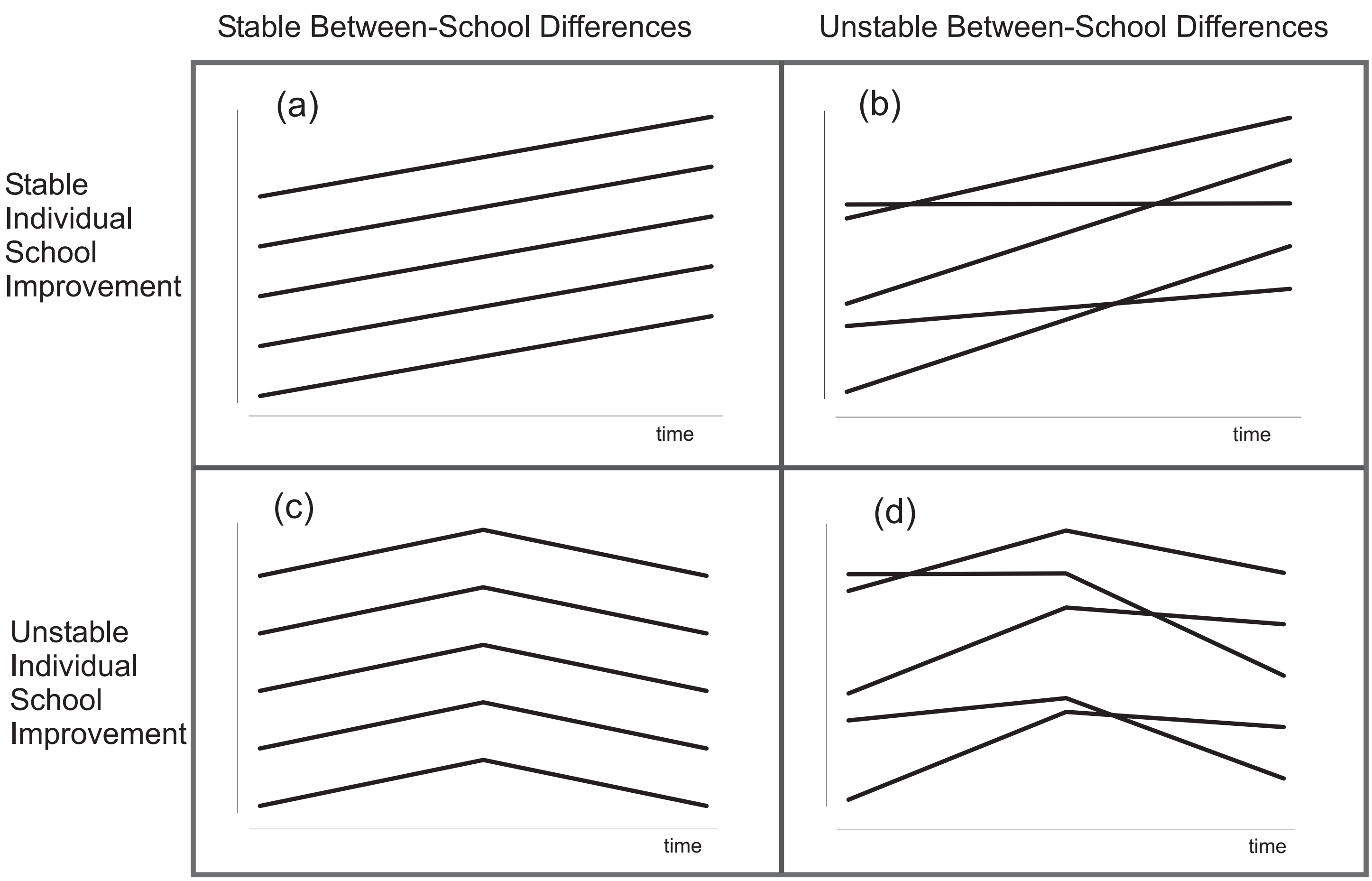


Figure 1.2. 2x2 Array of Stability of Improvement and Stability of Between-school Differences for Collection of Schools.

(indicating high stability for both). Comparing within each row in the 2x2 array, the LH and KS procedures are quite sensitive to between school differences in a direction opposite to common-sense, moving from 100% volatility to little or no volatility as individual differences in improvement increase modestly.

As will be further demonstrated in Sections 3 and 4, the LH and KS procedures simply do not reflect the important feature of the data– consistency of improvement for school trajectories– that is claimed in their substantive interpretations and statements of policy implications. Actually, instead of being measures of stability of school improvement; r_{LH} and p_{KS} appear to indicate an "anti-stability of individual differences" (i.e. measuring the inverse of a property, preservation of individual differences, not central to educational accountability). Consequently, “dead-wrong and irrelevant” are descriptors that simultaneously apply to the LH and KS methodology.

2. Common Sense Data Analysis for Consistency of Improvement

Before returning to the analysis and exposition of the undesirable properties of the LH and KS methodologies in Sections 3 and 4, useful data analysis displays for describing consistency of improvement (or lack thereof) are introduced in this section, with illustrations using California API data. In the context of the stock market, Art Cashin, (Director of Floor Operations, UBS Financial Services) provides the requisite wisdom, explaining on CNBC that the criterion for the credibility of a (short-term) stock market rally is: "Don't give up your gains." That simple statement seems also to be prescriptive for investigation of consistency of improvement in school accountability indices, and it drives the data analysis strategy for describing consecutive (or serial) improvement presented in this section.

In California, one aspect of the ongoing debate on accountability systems was the contention, put forth by the California Teachers Association and others, that year-to-year improvement in API scores was well-described by a "see-saw" metaphor, in that schools with scores that showed strong gains (and achieved awards) in one two-year cycle reversed those gains in the succeeding two-year cycle. The "see-saw" represents a legitimate empirical conjecture about (the lack of) consistency in improvement, which can be investigated by the data analysis methods of this section. The "see-saw" is similar to empirical conclusions (from correlational evidence) voiced in LH, such as the assertion: "schools that gain a lot from year one to year two will generally show a decline in year three, while those that show a decline from year one to year two will generally show a gain in year three" (p.33).

California API data. The California Academic Performance Index (API) is similar in construction to the weighted index for the CSAP used in LH, with weights (200, 500, 700, 875, 1000) assigned to each quintile of performance on a test (Reading, Math, Language etc) from the Stanford 9 battery. The resulting scores are combined across tests, producing a possible range for school scores of 200 to 1000. To provide some calibration of the scale it's useful to note that a school with about half of its students above the national 50^{th} percentile on the tests will have an API score around 650 and that a one percentile point improvement by each student on each test translates into a 8 to 10 point improvement in the school API (Rogosa, 2000). Also, the school-wide yearly growth target for a school with a 600 API is 10 points. Data analyses and interpretation of the statewide API data are provided in Rogosa (2000, 2001a, 2001b). The data used for the analysis of consistency of improvement are test scores for elementary school students over the four year period 1999-2002. A total of 4644 Elementary Schools have API scores all four years. (Here the 2002 API scores use only Stanford 9 test results.)

In addition to the analyses of these school level scores, separate analyses are conducted for the Socioeconomically Disadvantaged (SD) subgroups to illustrate a data analysis strategy for the "disaggregation of results." A student is classified as SD if the student qualifies for the National Student Lunch Program (NSLP) or neither parent is reported as graduating from high school. Of the 4644 Elementary Schools 2520 schools have at least 100 SD students, and the API scores for SD subgroups in these schools comprise the second data set. Conditioning further, a school is classified here as "HighSD" if at least half of its students are SD, and 2045 of the 2520 schools having at least 100 SD students are HighSD. The third data set is comprised of the API scores for the SD subgroups in these 2045 schools.

Table 2.1 contains basic descriptive statistics for the school scores and for year-to-year improvement in scores for these three groups. (Additional correlational information is provided in the Appendix.) All three sets of scores (Elementary Schools, SD subgroups, SD subgroups in HighSD schools) show strong year-to-year improvement in the aggregate, with the greatest aggregate improvement shown by the SD subgroups in HighSD schools (in part a function of the design of API scale which gives greatest increments to improvements of the lower scores.)

Insert Table 2.1

Tables 2.2, 2.3 and 2.4 present the consecutive improvement data displays for the three sets of data: Elementary Schools, SD subgroups, SD subgroups in HighSD elementary schools, respectively. Previous versions of these consistency in improvement analyses for all California schools can be found in Rogosa (2001b, "Year 2001 Growth Update: Interpretive Notes for the Academic Performance Index").

Three-year Displays for Consistency in Improvement. Start with the three-year improvement display for all Elementary Schools (top portion Table 2.2). The first and second columns of the table show that 4309 of the 4644 schools improved from 1999 to 2000 (ImpLevel '99-00 of 0), that 1457 of the 4644 schools improved more than 50 points (ImpLevel '99-00 of 50), and that

Table 2.1

Descriptive Statistics for School-Level Longitudinal Data and Year-to-Year Improvement: California API Index for Elementary Schools

California Elementary Schools (n=4644)

	Yearly School Scores				Year-to-Year Improvement		
percentile	1999	2000	2001	2002	'99-00	'00-01	'01-02
10th	450.56	494.12	525.	551.08	4.62	-11.12	-14.51
25th	523.75	568.38	594.5	618.91	18.75	3.38	-0.75
50th	631.62	674.	693.5	706.17	36.12	19.	14.35
75th	741.88	773.75	789.88	801.58	55.75	37.25	31.85
90th	819.25	846.38	858.88	867.59	75.12	56.5	50.68
Mean	633.17	671.49	692.48	708.73	38.32	20.99	16.26
St.Dev	136.63	130.82	124.37	117.53	29.08	27.58	26.62

SD Subgroups in California Elementary Schools (n=2520)

	Yearly School Scores				Year-to-Year Improvement		
percentile	1999	2000	2001	2002	'99-00	'00-01	'01-02
10th	412.75	458.81	484.62	516.38	5.81	-13.25	-12.68
25th	454.56	500.94	532.5	560.32	24.06	6.88	5.85
50th	508.62	556.25	586.25	611.88	43.38	26.44	25.09
75th	568.25	615.62	641.25	663.99	66.19	48.5	44.73
90th	627.5	670.62	692.	710.61	87.44	67.94	64.09
Mean	514.8	560.45	587.95	613.42	45.65	27.50	25.47
St.Dev	82.88	82.13	79.23	75.95	33.76	32.63	30.79

SD Subgroups in High SD Elementary Schools (n=2045)

	Yearly School Scores				Year-to-Year Improvement		
percentile	1999	2000	2001	2002	'99-00	'00-01	'01-02
10th	405.81	451.12	476.81	509.22	7.25	-11.5	-10.25
25th	444.88	490.56	522.12	549.31	24.69	8.5	7.62
50th	492.06	539.75	570.25	599.36	44.38	28.	27.24
75th	543.5	594.	621.88	648.37	67.19	49.44	46.53
90th	597.25	647.12	669.88	692.69	89.44	69.	65.52
Mean	498.18	544.89	573.62	600.90	46.71	28.74	27.28
St.Dev	75.58	76.48	74.23	72.09	33.19	32.13	30.26

466 of the 4644 schools improved more than 75 points. Columns 3 and 4 provide the information on consistency of the initial improvement in the second testing cycle, 2000-2001. The fraction of the 4309 Elementary Schools that improved in 1999-2000 and also improved in 2000-2001, is .785 (3381/4309). This proportion is a natural indicator for consistency in improvement, addressing the question, Given that the school improved in the first testing cycle ('99-00) , did that school also improve in the second cycle ('00-01)? The magnitude of the second-cycle improvement is summarized in the fourth column; for the 4309 schools improving in '99-00 the median improvement in '00-01 is 18 points, with upper quartile 36.3 (as 25% of these 4309 previously improving schools improved at least 36.3 points in 200-2001).

Insert Table 2.2

The additional rows of the top portion of Table 2.2 repeat this description of consistency in improvement for the subsets of the schools achieving (larger) improvement levels (ImpLevel) in '99-00. Even schools with large improvement 1999-2000 show substantial (but far from perfect) consistency in improvement. Nearly 3/4 (.737) of the 1457 schools improving more than 50 points in 1999-2000 also improve in 2000-2001, with a median improvement in '00-01 for those 1457 schools of 16.2. That median improvement is smaller than the typical improvement of those schools in '99-00 (as all improved more than 50 points). But, remember, deceleration is not decline. Generally, these schools continued to improve, albeit with smaller magnitudes. As seen in Table 2.1, overall school improvement in 2000-2001 was smaller in magnitude than in 1999-2000.

A visual representation of the tabled quantities in the top portion of Table 2.2 is provided by the scatterplot in Figure 2.1, a plot of improvement in school API scores '00-01 versus improvement '99-00. Of course, this is the scatterplot that yields r_{KS} , but instead of evaluating how close the points in the scatterplot lie to a straight-line, the summaries in the top portion of Table 2.2 use the information in the scatterplot to address questions about consistency in improvement. That is, choose an ImpLevel for '99-00, locate that point on the horizontal axis of Figure 2.1 and draw a vertical line corresponding to that value of the horizontal axis. The number of points to the right of that vertical line is the second column of Table 2.2. The proportion in the third column of Table 2.2 is the ratio of number of points contained in the (upper right) quadrant defined by the vertical line at

Table 2.2
Consecutive Improvement for California Elementary Schools (n=4644)

Three year Improvement 1999-2001

ImpLevel 1999-2000	Number exceeding ImpLevel	Proportion of those improving 2000-2001	Amount of Improvement 2000-2001 {lowest decile lower quartile median upper quartile}
0	4309	0.785	-11.5 2.8 18. 36.3
25	3090	0.763	-13.5 1.2 17.4 35.5
50	1457	0.737	-18.5 -0.8 16.2 35.5
75	466	0.727	-19.5 -1.6 16.4 36.5
100	129	0.674	-30.4 -6.7 16. 41.

Fourth-year Improvement, 1999-2002

ImpLevel 1999-2000 and 2000-2001	Number exceeding both ImpLevels	Proportion of those improving 2001-2002	Amount of Improvement 2001-2002 {lowest decile lower quartile median upper quartile}
0	3381	0.733	-14.6 -1. 13.9 31.
25	1199	0.746	-15.3 -0.3 15.9 34.1
50	181	0.718	-18. -1.5 19. 40.2

ImpLevel and the horizontal line intersecting the vertical axis at zero and the number of points to the right of the vertical line. The summary statistics in the fourth column of Table 2.2 describe the vertical axis values of the points in the scatterplot that lie to the right of the vertical line having horizontal axis value ImpLevel. This scatterplot is also a useful adjunct to Table 2.2 in identifying anomalous schools, as erratic instances do exist, but are not typical.

Insert Figure 2.1

Fourth-year Augmentation for Consistency in Improvement Displays. As the LH application uses four years of data, and four years of data are available for the California API data, extensions of the data analysis strategy for consistency in improvement to four years of data are presented (the bottom portions of Tables 2.2-2.4). The display in the bottom portion of Table 2.2 is intended to augment the summary for the first three years of school-level data in the top portion of the table. The conditioning indicated by the ImpLevel is conjunctive for improvement in '99-00 and '00-01. That is, the 1199 schools indicated for ImpLevel 25 in Table 2.2 satisfy the condition of year-to-year improvement exceeding 25 points both in '99-00 and in '00-01. Of those 1199 schools the proportion continuing to improve in '01-02 is .746, and these 1199 schools have median improvement of 15.9 points in '01-02. That most schools exhibiting strong improvement over the first three years continue to improve is an indication of consistency in improvement (and the amount of this improvement is non-trivial). (Alternative four-year data summaries in the spirit of Table 2.2 may be desirable in some settings such as a display resembling the top portion of Table 2.2 which conditions on the cumulative improvement for year 1 to year 3 and then describes consistency of improvement for year 3 to year 4.)

A three dimensional scatterplot displaying the values of the year-to-year improvement triplet for each school would be the four-year data analog to Figure 2.1. Easier to visually digest are the set of two-dimensional scatterplots (derived from the three-dimensional display) shown in Figure 2.2. The three scatterplots correspond to the bottom portion of Table 2.2 and contain 3381, 1199, and 181 schools, respectively. Correspondences between the third and fourth columns of the table and each of the scatterplots are obtained in the same manner as described for Figure 2.1 above. Note that the scatterplot corresponding to r_{LH} (improvement from time 2 to time 4 versus improvement from time 1 to time 3) has no role or relevance in the

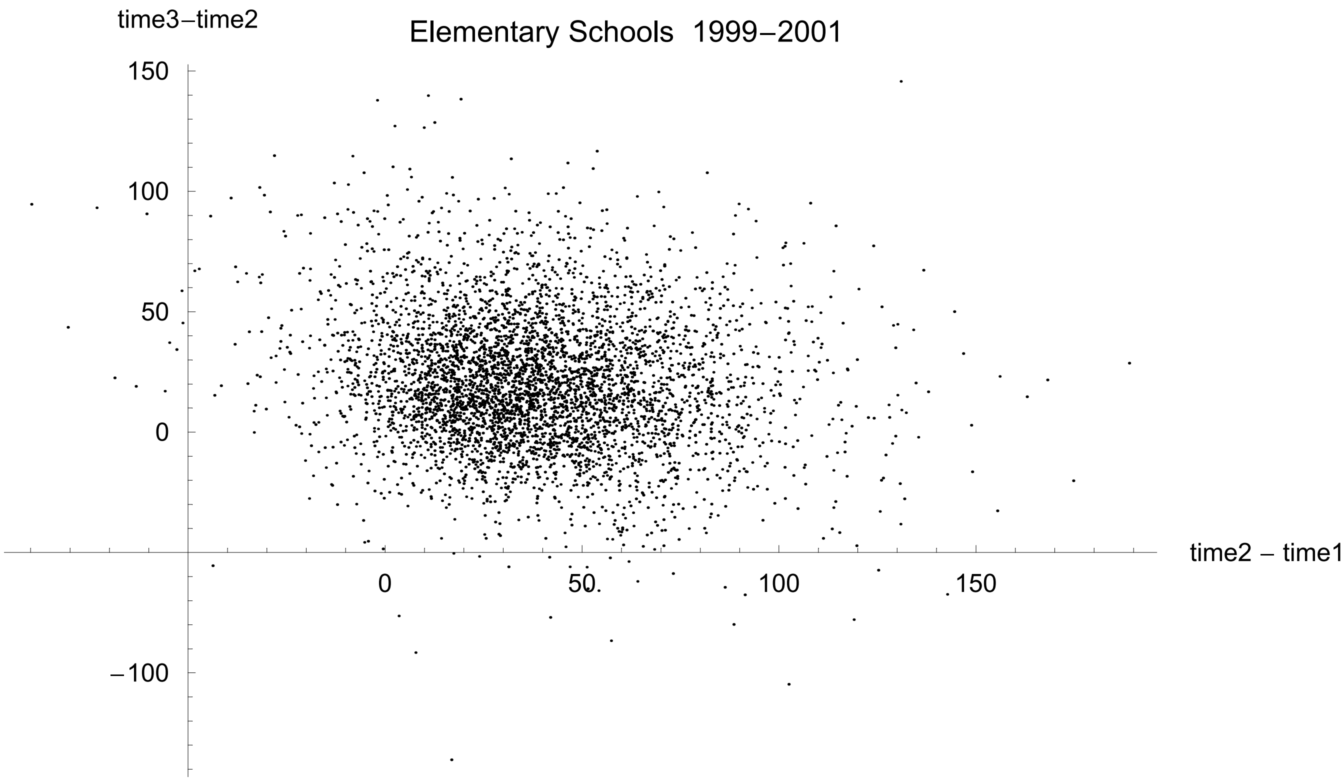


Figure 2.1. Scatterplot of improvement in school API scores '00-01 versus improvement '99-00.

consecutive improvement data analysis over four years (but for completeness the r_{LH} scatterplot is shown in the Appendix).

Insert Figure 2.2

Additional Consistency Information. Additional information not in Table 2.2 provides corroboration for consistency in improvement for the elementary school API scores. To investigate the prevalence of a severe up-and-down pattern, simply count schools that do give up their gains. For the three-year span '99-01 only 24 of the 3090 elementary schools improving more than 25 points in '99-00 reverse their improvement, and only 5 of the 1457 elementary schools improving more than 50 points in '99-00 reverse their improvement. Similarly, over the four-year period '99-02, only 17 of the 3381 elementary schools improving in both '99-00 and '00-01 reverse their cumulative improvement in '01-02, and none of the 1199 elementary schools improving more than 25 points in both '99-00 and '00-01 reverse their cumulative improvement in '01-02. The (rare) reversals can be seen in the scatterplots in Figures 2.1 and 2.2.

Another aspect of consistency of improvement is consistency of decline; i.e., Do schools that initially decline continue to decline? Versions of Table 2.2 could be constructed for consecutive decliners and negative ImpLevels, but the strong positive improvement trend in the California data leaves few schools eligible. For example only 145 of the 4644 elementary schools declined more than 10 points in '99-00 and .876 of these schools improved in '00-01 (median improvement 36 points). Thus there is a down-and-up pattern for this small subset of schools.

Analyses for SD subgroups. Tables 2.3 and 2.4 repeat the consecutive improvement analyses of Table 2.2 for the SD subgroups (Table 2.3) and the SD subgroups in HighSD schools (Table 2.4). The scatterplots corresponding to Figures 2.1 and 2.2 are Figures 2.3 and 2.4 for SD subgroups and Figures 2.5 and 2.6 for SD subgroups in HighSD schools. The major feature to note from Tables 2.3 and 2.4 is that consistency of improvement is stronger for the SD subgroups, over both three years and four years, than for all elementary schools, and consistency in improvement is slightly stronger for SD subgroups in HighSD schools than all SD subgroups. An example from Table 2.4 is that 763 of the 2045 SD subgroups in HighSD schools improve more than 25 points in both '99-00 and '00-01. And over 4/5 (.81) of these SD subgroups also improve in '01-02, with median improvement for these 763

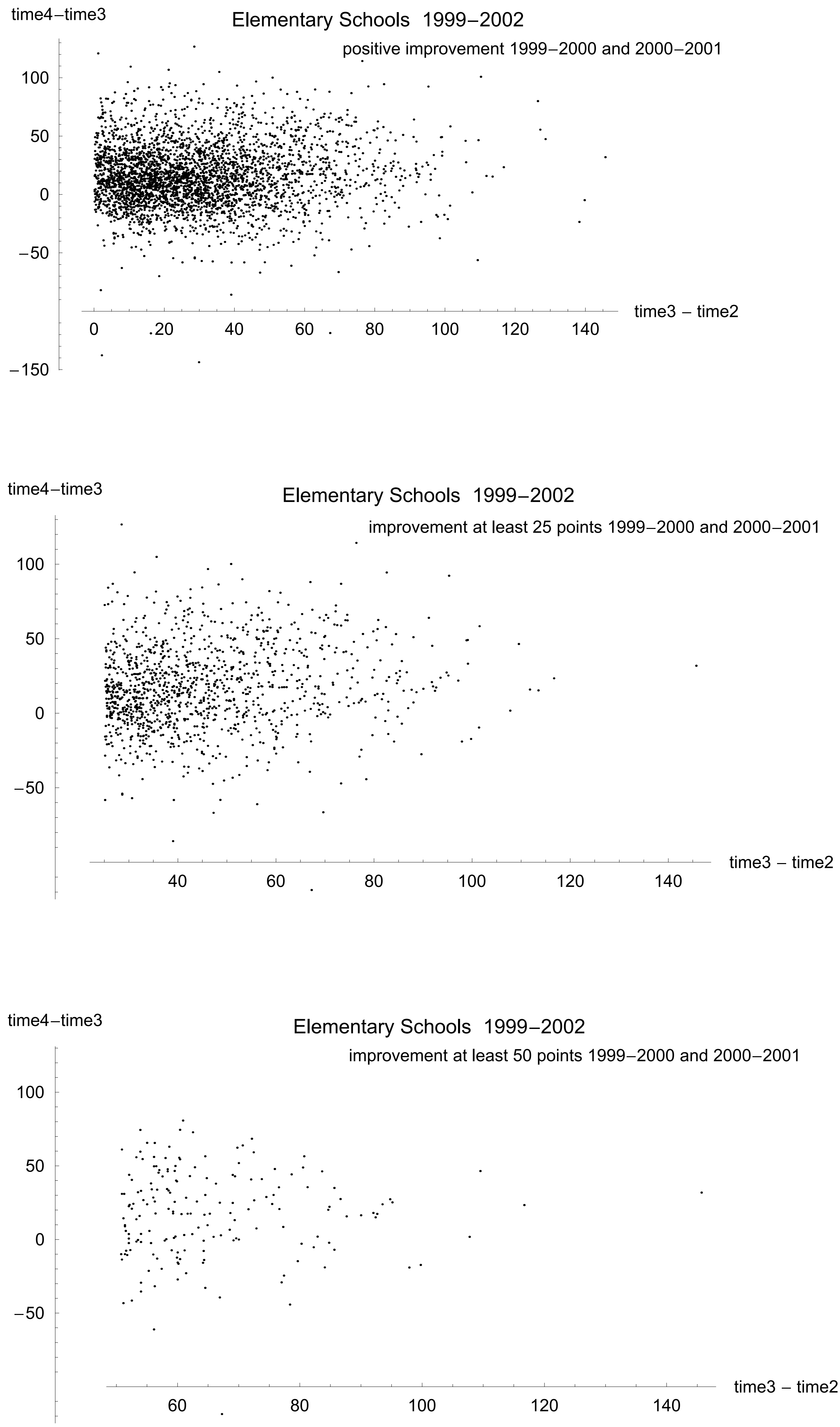


Figure 2.2. Set of three scatterplots for four-year improvement in school API scores.

SD subgroups of 26 points.

Insert Tables 2.3, 2.4
Insert Figures 2.3-2.6

What would LH, KS determine? The applications in both KS and LH are primarily to successive cohorts, such as fourth grade scores in successive years, where overlap of students (same students in repeated years) does not arise. The API is a multi-grade school-level index, and in successive years perhaps half the students in each elementary school appear in both years. The extra correlation induced by this partial overlap should not make a large difference, but to be careful consider the caveat that this presentation of r_{LH} and p_{KS} below for the API data could be considered an "off label" application of the LH and KS procedures.

	Elementary Schools	SD subgroups	SD subgroups in HighSD schools
r_{LH}	.464	.405	.401
p_{KS} (for '99-01)	.721	.615	.59

For each set of data p_{KS} is considerably larger (less pessimistic about consistency in improvement) than r_{LH}, even though the bottom portions of Tables 2.2, 2.3 and 2.4 indicate that the consistency in improvement seen in the first three years of data holds up for year 3 to year 4 improvement. In fact, by serendipity, the p_{KS} value for elementary schools is close to the proportion of schools showing consecutive improvement in Table 2.2. One striking feature of this tabulation is that both r_{LH} and p_{KS} decrease across the columns (groups), whereas the observed consistency of improvement increases from Tables 2.2 to 2.3 to 2.4. The artificial data examples in Section 4 will present dramatic differences between actual consistency in improvement and the indications from r_{LH} and p_{KS} .

Table 2.3
Consecutive Improvement for SD Subgroups in California Elementary Schools (n=2520)

Three year Improvement 1999-2001

ImpLevel 1999-2000	Number exceeding ImpLevel	Proportion of those improving 2000-2001	Amount of Improvement 2000-2001 {lowest decile lower quartile median upper quartile}
0	2344	0.803	-13.8 6.2 25.8 47.6
25	1850	0.793	-15.1 4.9 24. 45.4
50	1051	0.749	-19.3 0 20.2 42.2
75	451	0.707	-24. -4. 17. 41.8
100	153	0.66	-34.1 -7.2 18.6 42.2

Fourth-year Improvement, 1999-2002

ImpLevel 1999-2000 and 2000-2001	Number exceeding both ImpLevels	Proportion of those improving 2001-2002	Amount of Improvement 2001-2002 {lowest decile lower quartile median upper quartile}
0	1883	0.801	-12.9 5.7 24.8 43.9
25	902	0.784	-16.4 3.4 22.9 42.3
50	196	0.765	-17.2 2.8 24.2 45.

Table 2.4
Consecutive Improvement for SD Subgroups in High SD Elementary Schools (n=2045)

Three year Improvement 1999-2001

ImpLevel 1999-2000	Number exceeding ImpLevel	Proportion of those improving 2000-2001	Amount of Improvement 2000-2001 {lowest decile lower quartile median upper quartile}
0	1922	0.816	-11.9 7.6 26.8 48.2
25	1516	0.807	-13. 6. 25.3 46.
50	874	0.767	-18.4 2.1 21.3 43.8
75	375	0.72	-22. -3.7 17.4 42.1
100	129	0.667	-27.5 -6.9 18.6 42.2

Fourth-year Improvement, 1999-2002

ImpLevel 1999-2000 and 2000-2001	Number exceeding both ImpLevels	Proportion of those improving 2001-2002	Amount of Improvement 2001-2002 {lowest decile lower quartile median upper quartile}
0	1569	0.829	-10.1 7.7 27.2 45.3
25	763	0.81	-12.9 5.6 26. 43.9
50	174	0.77	-18.1 3.2 25. 47.1

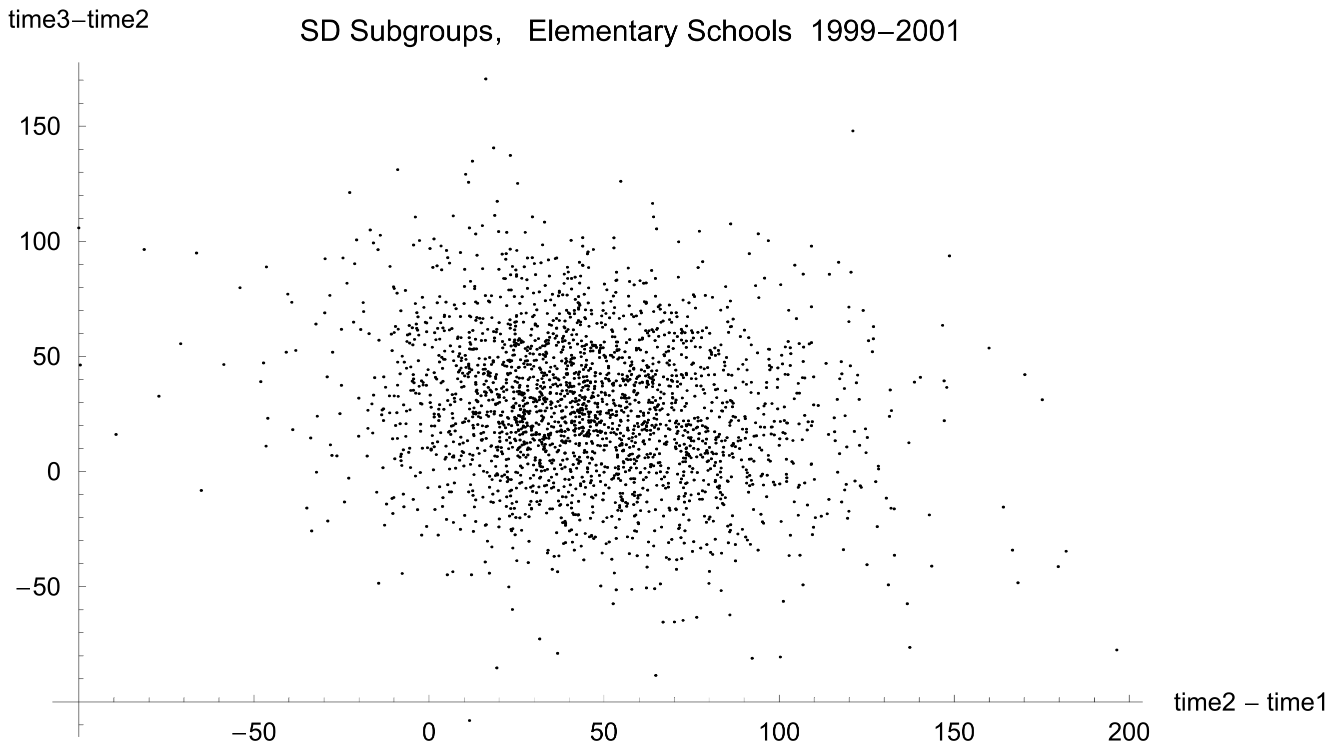


Figure 2.3. Scatterplot of improvement in SD subgroup API scores '00-01 versus improvement '99-00.

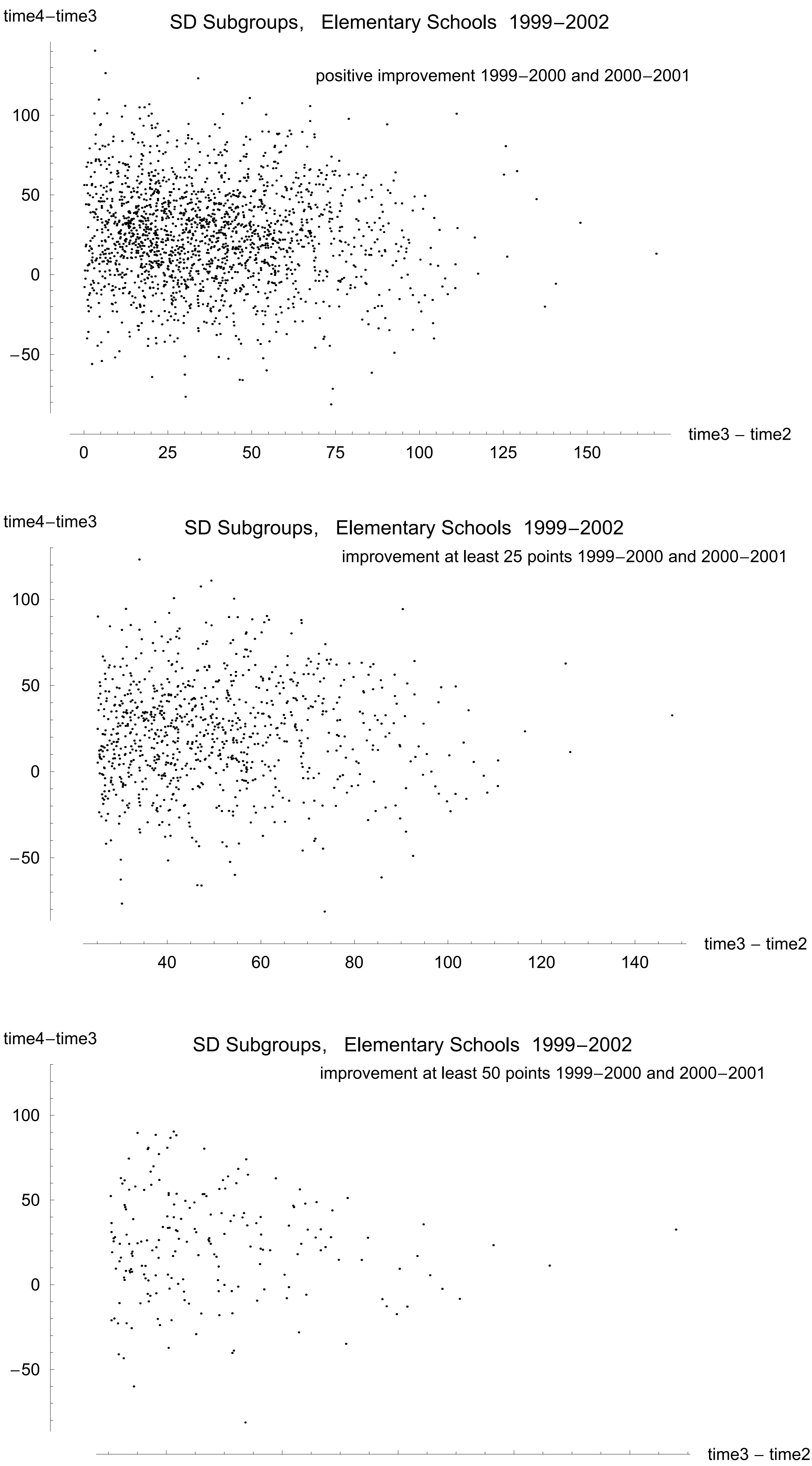


Figure 2.4. Set of three scatterplots for four-year improvement in SD subgroup API scores.

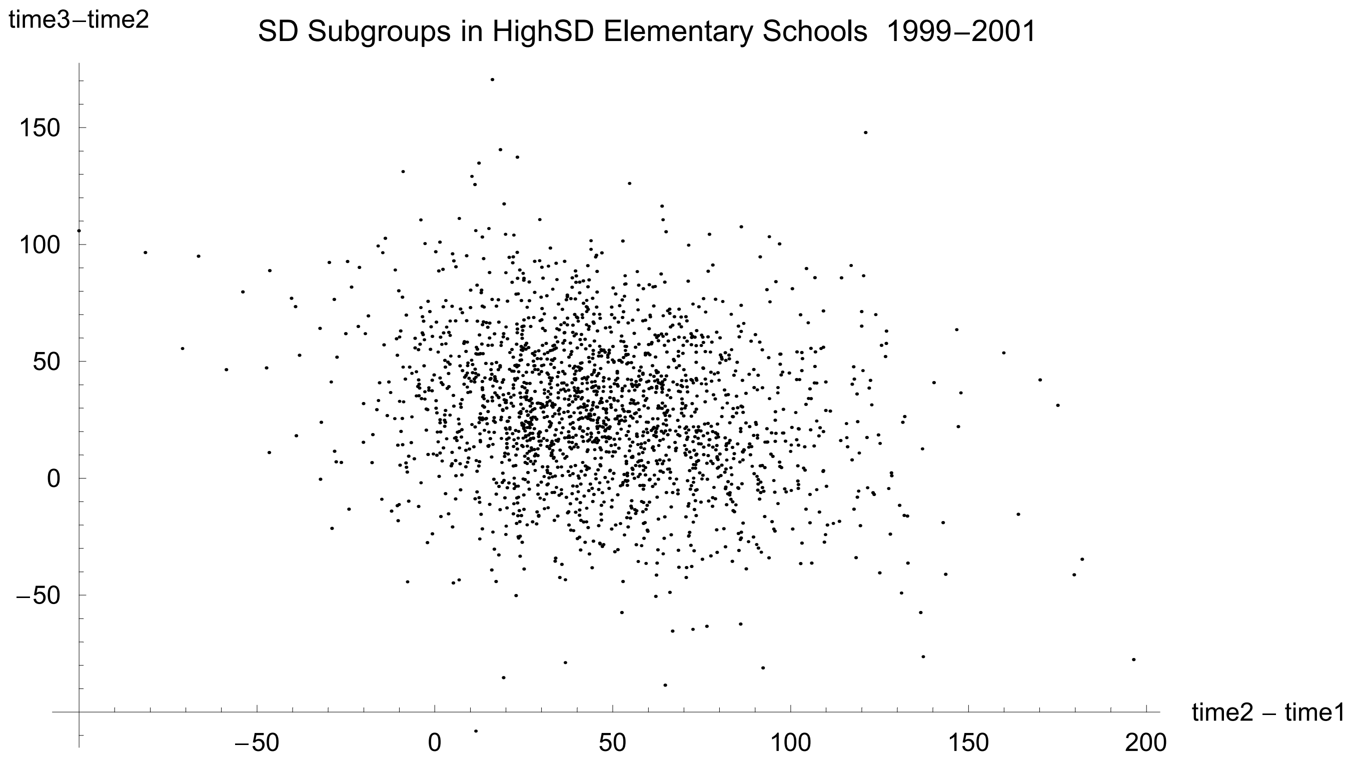


Figure 2.5. Scatterplot of improvement in SD subgroup in High SD schools API scores '00-01 versus improvement '99-00.

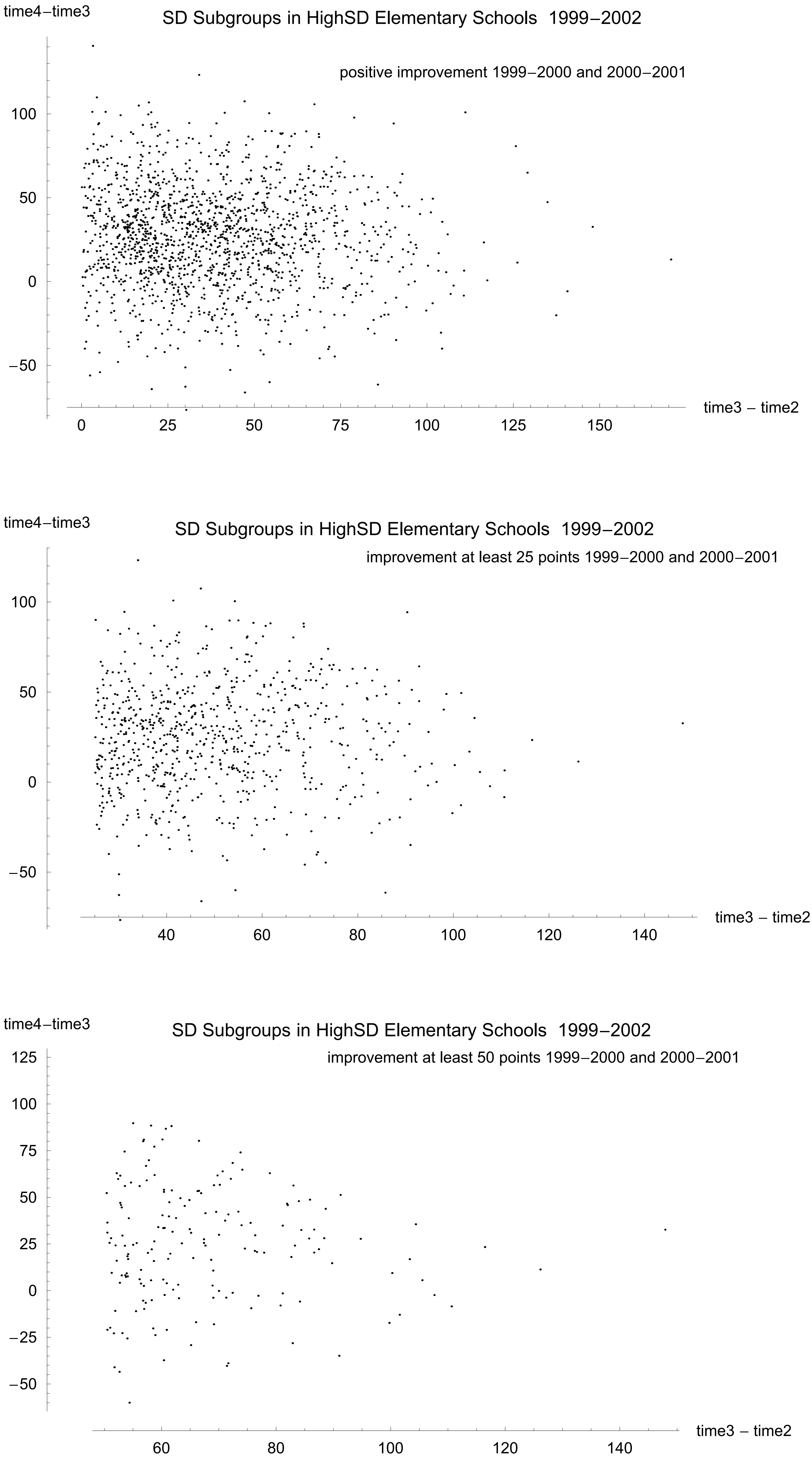


Figure 2.6. Set of three scatterplots for four-year improvement in SD subgroup in High SD schools API scores.

3. Analytic Results for LH and KS Stability/Volatility Measures

This section presents the main technical results for the LH and KS measures, using simple growth curve models to represent the trajectory of an individual school. Both the LH and KS measures are seen to be monotonic functions of the reliability coefficient for the time1, time2 difference score, a quantity which has no discernable relevance to consistency in improvement for individual schools. For more details on results in this section see Appendix B.

Constant Rate of Change Growth Model

Each unit (school) s has true change θ_s and the constant rate of change model is

$$Y_s(t) = \eta_s(0) + \theta_s t + \varepsilon_{ts} \ , \tag{3.1}$$

where the perfectly measured score for unit s at time t is $\eta_s(0) + \theta_s t$. The constant improvement in the perfectly measured school score for any single-year interval [t, t+1] is θ_s (as depicted by the single school trajectory in Figure 1.1, frame a). Over the collection (population) of schools, the θ_s have mean μ_θ and variance σ^2_θ ; a collection of schools with $\sigma^2_\theta = 0$ is depicted in Figure 1.2, frame a, whereas Figure 1.2, frame b depicts sizeable σ^2_θ .

The error of measurement ε_{ts} has mean 0, and the measurement error variance σ^2 is assumed constant across time and units. In the setting of successive cohorts of students no overlap of students is present, and thus the ε_{ts} can be considered independent (*iid*).

For the time interval $[t_1, t_2]$ the observed improvement for school s is the difference score $D_s[t_1, t_2] = Y_s(t_2) - Y_s(t_1)$. Because of its central role in the LH and KS procedures, introduce the reliability coefficient for $D[t_1, t_2]$, which is $\sigma^2_\theta / (\sigma^2_\theta + 2\sigma^2/(t_2 - t_1)^2)$. A convenient notation is to write the reliability coefficient for year-to-year improvement, D[t, t+1], as $\rho[D_t]$. For the constant rate of change growth model $\rho[D_t]$ is the same for all t : $\rho[D_t] = \sigma^2_\theta / (\sigma^2_\theta + 2\sigma^2)$. Additional technical material on the properties of collections of straight-line growth curves can be found in Rogosa (1993, 1995), Rogosa et. al. (1982), Rogosa and Willett (1983b, 1985) and Rogosa and Saner (1995).

Proportional Deceleration in Change.
A modified constant rate of change model, which may better reflect empirical experience in state assessments, has school scores increasing smaller amounts in successive time periods. That is,

$$D[t, t+1] = Y_s(t+1) - Y_s(t) = h^{t-1}\theta_s + \varepsilon_{t+1,s} - \varepsilon_{ts} \qquad (3.2)$$

for $h \leq 1$ and equally-spaced integer time-points {1,2,3,4,...}. This proportional deceleration model simply says that the true improvement between times 1 and 2 is θ_s , between times 2 and 3 is $h\theta_s$, and between times 3 and 4 is $h^2\theta_s$ (e.g., with h = 2/3, "true" improvement of 45, 30, 20 points in the three year-to-year intervals, respectively). Proportional deceleration is depicted by the single school trajectory in Figure 1.1 frame b.

Then the reliability of the year-to-year difference score for times {1,2,3,4}: $\rho[D_1] = \sigma^2_\theta/(\sigma^2_\theta + 2\sigma^2)$, $\rho[D_2] = h^2\rho[D_1]/(1 - \rho[D_1](1 - h^2))$, and $\rho[D_3] = h^4\rho[D_1]/(1 - \rho[D_1](1 - h^4))$.

3.1 Results for LH and KS using the Constant Rate of Change Model

Keep in mind that under the constant rate of change model each school has perfectly consistent improvement (same amount each time period) except for the effect of statistical uncertainty (error term) in the measured score.

KS proportion persistent. For the constant rate of change model with equally spaced observations at (t_1, t_2, t_3) such as yearly measurements, the population value of r_{KS} is a simple function of the reliability of the difference score: $(3\rho[D_t] - 1)/2$. Consequently, for the population value of the proportion persistent measure, p_{KS} :

$$\text{KS persistence} = 3\rho[D_t]. \tag{3.3}$$

LH stability. For the constant rate of change model with equally spaced observations at (t_1, t_2, t_3, t_4) such as yearly measurements, the population value of r_{LH} is also a simple function of the reliability of the difference score:

$$\text{LH stability} = 4\rho[D_t]/(1 + 3\rho[D_t]) . \tag{3.4}$$

As r_{LH} has the form of a parallel-form reliability estimate for the difference score, it is not surprising that the population value of r_{LH} is equal to the reliability coefficient for D[1,3]. This reliability of D[1,3] can be obtained by applying the Spearman-Brown Formula to $\rho(D_t)$ with the augmentation factor being $(3 - 1)^2$, yielding (3.4) .

Thus when the reliability of the difference score is zero, both KS and LH determine 100% volatility. A zero value for reliability of the difference score pertains for the configuration depicted in Figure 1.2 frame a, even though Figure 1.2a shows perfectly consistent improvement (in the perfectly measured score) for each school which is replicated for all schools (i.e. also perfect stability of individual differences). On the other hand, for settings such as depicted in Figure 1.2 frame b, with perfectly consistent improvement for each school but a lack of stability of individual differences (i.e., different rates of improvement for different schools), $\rho[D_t]$ will be substantial (for typical values of uncertainty in the observed school scores) and KS and LH would determine substantial stability. That is, LH and KS determine a lack of volatility when stability of individual differences is not present, and determine 100% volatility when stability of individual differences is present. So it would seem little can be learned, but much misinformation generated, from computing either the KS or LH statistics. The artificial data examples in Section 4 will reinforce and extend this message.

Figure 3.1 and Table 3.1 display population values of the stability measure r_{LH} and the proportion persistent measure, p_{KS} as a function of $\rho[D_t]$ for the constant rate of change model. For $\rho[D_t] > .1$, LH is even more pessimistic than KS about consistency of improvement. The KS measure is truncated to 1 (0% volatility) for $\rho[D_t] > 1/3$; and for $\rho[D_t] = 1/3$, the LH measure is 2/3.

Insert Figure 3.1 and Table 3.1

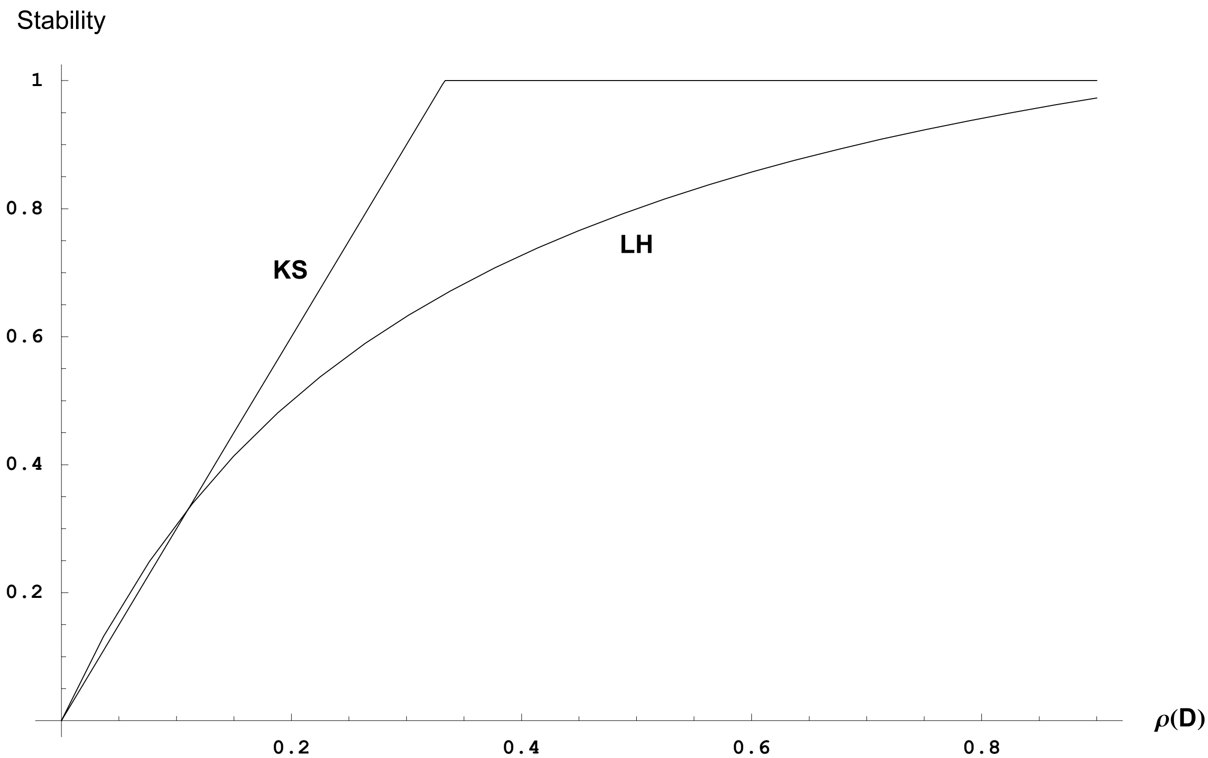


Figure 3.1 Population values of rLH and pKS for constant rate of change model.

Table 3.1
Stability Measures for Constant Rate of Change Model

$\rho[D_t]$	LH	KS
0	0	0
0.03	0.11	0.09
0.05	0.174	0.15
0.1	0.308	0.3
0.15	0.414	0.45
0.2	0.5	0.6
0.25	0.571	0.75
0.3	0.632	0.9
0.35	0.683	1
0.4	0.727	1
0.45	0.766	1
0.5	0.8	1
0.55	0.830	1
0.6	0.857	1
0.65	0.881	1
0.7	0.903	1

3.2 Results for LH and KS using the Proportional Deceleration Model

KS proportion persistent. For the proportional deceleration model with equally spaced observations at (t_1 = 1, t_2 = 2, t_3 = 3) such as yearly measurements, the population value of r_{KS} is a function of h (the proportional deceleration constant) and the reliability of D[1,2] (the difference score for improvement between t_1 and t_2) denoted above by $\rho[D_1]$. The result for the population value of r_{KS} can be written as:

$$(\rho[D_1](1 + 2h) - 1)/2[1 - (1 - h^2)\rho[D_1]]^{1/2} = ((2 + h)\rho[D_2] - h)[\rho[D_1]/4\rho[D_2]]^{1/2} .$$

Then the population value of p_{KS} can be written:

$$\text{KS persistence} = 1 + (\rho[D_1](1 + 2h) - 1)/[1 - (1 - h^2)\rho[D_1]]^{1/2} . \qquad (3.5)$$

Thus KS will show 0% volatility, perfect persistence, for $\rho[D_1] = 1/(1 + 2h)$ such as the {h, $\rho[D_1]$} pairs {1, 1/3}, {3/4, 2/5}, {1/2, 1/2}, {1/3, 3/5}. KS will show 100% volatility for $\rho[D_1] = 0$.

LH stability. For the proportional deceleration model with equally spaced observations at (t_1 = 1, t_2 = 2, t_3 = 3, t_4 = 4) such as yearly measurements, the population value of r_{LH} is a function of h (the proportional deceleration constant) and the reliability of the difference score D[1,2], denoted above by $\rho[D_1]$:

$$\text{LH stability} = h(1 + h)^2\rho[D_1])/[(1 + h(2 + h)\rho[D_1])(1 + (h^2 + h - 1)(h^2 + h + 1)\rho[D_1])]^{1/2} . \qquad (3.6)$$

LH will show 0% stability, 100% volatility for $\rho[D_1] = 0$.

Figure 3.2 and Table 3.2 present population values of the stability measure r_{LH} and the proportion persistent measure, p_{KS} as a function of h and $\rho[D_1]$ for the proportional deceleration model . For almost all {h, $\rho[D_1]$} pairs, LH is notably more pessimistic than KS about consistency of improvement. The KS measure is truncated to 1 (0% volatility) for $(1 + 2h)\rho[D_1] > 1$. For example, with h = 1/2, $\rho[D_1] = 1/2$, KS determines 0% volatility, but the population value of r_{LH} is .499.

Insert Figure 3.2 and Table 3.2

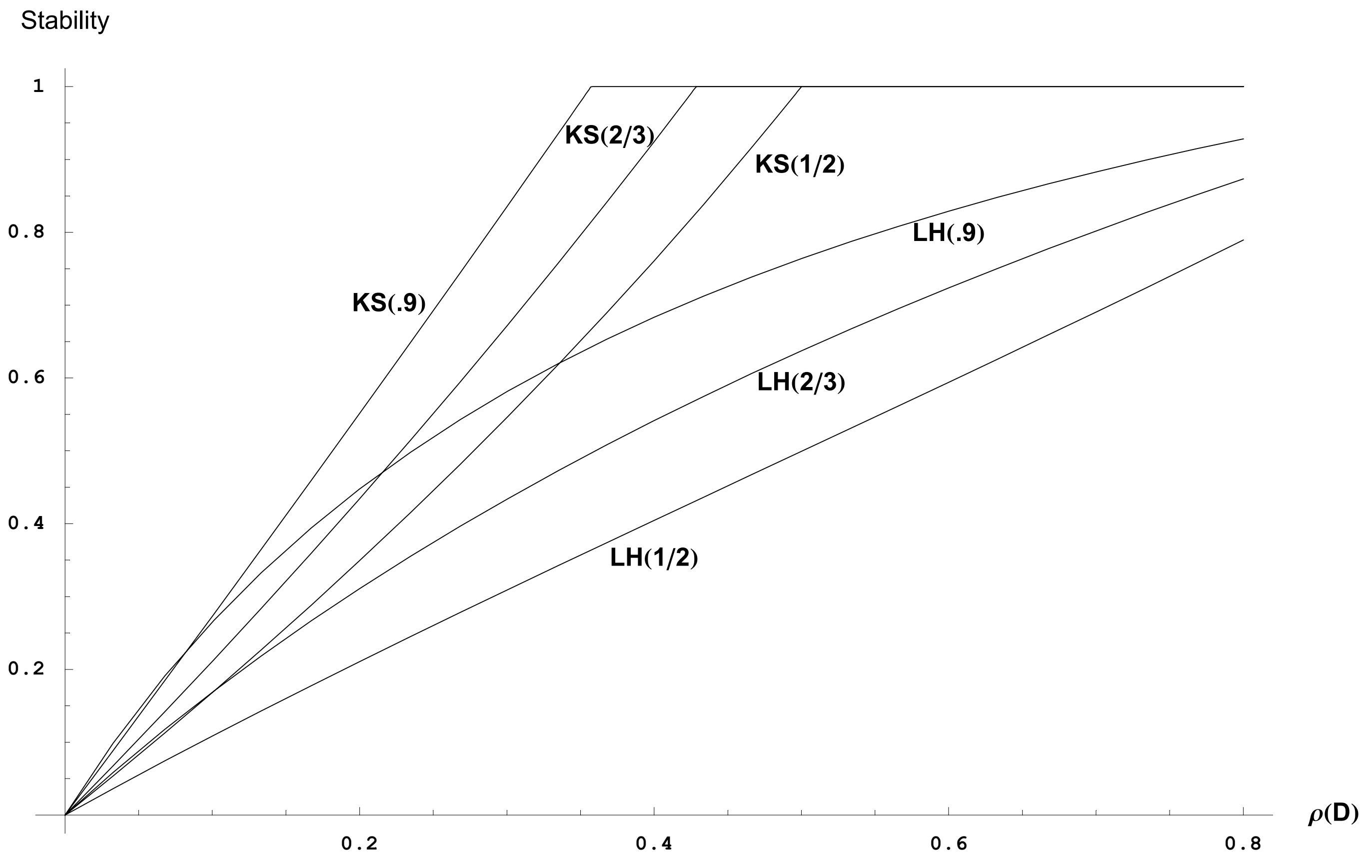


Figure 3.2 Population values of rLH and pKS for proportional deceleration model with h = 9/10, 2/3, 1/2 .

Table 3.2
Stability Measures for Proportional Deceleration Model

	h					
	0.5	**0.6**	**0.7**	**0.8**	**0.9**	**1.**
$\rho[D_1] = 0$						
KS	0.	0.	0.	0.	0.	0.
LH	0.	0.	0.	0.	0.	0.
$\rho[D_1] = .1$						
KS	0.168	0.194	0.220	0.246	0.273	0.3
LH	0.108	0.143	0.182	0.223	0.265	0.308
$\rho[D_1] = .2$						
KS	0.349	0.400	0.451	0.502	0.551	0.6
LH	0.211	0.270	0.331	0.391	0.448	0.5
$\rho[D_1] = .3$						
KS	0.546	0.622	0.696	0.767	0.835	0.9
LH	0.309	0.385	0.457	0.523	0.581	0.632
$\rho[D_1] = .4$						
KS	0.761	0.861	0.955	1	1	1
LH	0.405	0.490	0.565	0.630	0.683	0.727
$\rho[D_1] = .5$						
KS	1.	1	1	1	1	1
LH	0.499	0.587	0.660	0.718	0.764	0.8
$\rho[D_1] = .6$						
KS	1	1	1	1	1	1
LH	0.594	0.679	0.743	0.792	0.829	0.857
$\rho[D_1] = .7$						
KS	1	1	1	1	1	1
LH	0.690	0.765	0.818	0.856	0.883	0.903

4. Artificial Data Examples for Consistency of School Improvement

The analytic results in Section 3, demonstrating that the LH and KS stability measures are actually monotonic functions of the gain score reliability coefficient, make it all too easy to create examples with any properties desired. Low values of the reliability of the time1, time2 difference score can be constructed either with very accurate measurement and little individual differences in change or with poor measurement (low accuracy) and large individual differences in change. Consequently, very good accuracy for estimating change and perfect consistency of true change (i.e., straight-line growth model) are consistent with LH and KS determinations of great volatility (i.e. large proportion of change is transient).

The four subsections each present an artificial data example with known structure. For each example, the conclusions about volatility indicated by the LH and KS procedures are compared with the known structure of the data and especially with the data analysis displays for consistency of improvement introduced in Section 2. Artificial data examples 1 and 2 (sections 4.1 and 4.2) use the constant rate of improvement model to generate examples displaying reasonably strong consistency in improvement. Yet in one case (section 4.1, Tables 4.1-4.3, Figure 4.1) LH and KS determine great volatility, and in the other case (section 4.2, Tables 4.4-4.6, Figure 4.2) LH and KS determine strong stability (persistence). Artificial data example 3 (section 4.3, Tables 4.7-4.9, Figure 4.3) presents data exhibiting the up-and-down patterns (c.f., Figure 1, frames c and d) clearly indicating great inconsistency in improvement, yet for these data LH and KS determine great stability. Artificial data example 4 (section 4.4, Tables 4.10-4.11, Figure 4.4) employs the proportional deceleration mode1 (Figure 1, frame b and Section 3) with h = 2/3, for which LH and KS determine great volatility. These examples reinforce the message that little can be learned, but much misinformation generated, from computing the LH or KS statistics.

School outcome measures: weighted index and proportion proficient. The analytic results in section 3 and artificial data generation in this section employ a continuous quantitative score for the school outcome, denoted by $Y_s(t_i)$. Composite measures such as the California API and the North Carolina test scores used in KS are well-represented by this formulation, as are the weighted index form of the CSAP scores in LH. The generation of artificial longitudinal data based on a collection of growth curve models is described in Rogosa and Saner (1995, esp Appendix A).

The additional LH application to proportion proficient outcomes could be finessed by restating a growth model for proportions above cut-off rather than for the school means, and thus the analytic results of section 3 would directly apply. A more realistic and satisfactory treatment is to use the longitudinal model for improvement in school mean scores $Y_s(t_i)$ along with the statistical uncertainty in the school outcome (error variance σ^2) to obtain a distribution of scores within school. The examples use 75 students within each school (representing the single grade scenario in KS and LH), and thus the within-school distribution has variance $75\sigma^2$. Then the empirical proportion proficient for school *s* at time t_i is taken to be the proportion of the distribution $N(Y_s(t_i)\,,\,75\sigma^2)$ which lies at or above the proficient level. The proficient level in the artificial data examples is set so that median school at time 1 has approximately 40% of students proficient. The artificial examples show that r_{LH} has similar values when computed for the school means or school proportion proficient (a correspondence also seen by LH for the Colorado data). Therefore stability or volatility indications from the LH methods should be given the same regard whether the school outcome score employed be proportion proficient or the weighted index.

4.1 Artificial Data Example 1: LH, KS Volatility

Artificial data example 1 is composed of test outcomes on 10000 schools at each of 4 (yearly) observations. Table 4.1 displays an array of descriptive statistics for this example (similar in content to LH tables 1,2,3,4). Start with the results for the "Continuous Outcome" for a school (top half of Table 4.1), such as the weighted index measure in LH, which for these examples are in the metric of the California API. Median (and mean) improvement is about 60 points each year, with steady overall improvement seen throughout the score distribution. Correlation matrices for the yearly scores and for year-to-year improvement in Table 4.1 have values quite similar to those for CSAP data cited by LH p.33. Also, the observed score correlation between year 1 status and year 1 to year 2 improvement, Corr($Y_s(1)$, $Y_s(2) - Y_s(1)$) is –.348 for example 1 (while LH cite –.35 for CSAP).

Insert Table 4.1

Perhaps the most interesting correspondence lies in the correlations between adjoining changes (e.g. correlation between yr1-yr2 improvement and yr2-yr3 improvement), shown in Table 4.1 to have values of nearly –.5, values also found with the CSAP data. In particular, LH (p.33) cite these correlations as indicating lack of stability: "schools that gain a lot from year one to year two will generally show a decline in year three" and "it should not be surprising that schools that show outstanding gains... do not look so good with respect to their gains the following year." (p.33).

Determinations of consistency in improvement. For the example 1 artificial data, r_{LH} = .097 and p_{KS} = .087. Scatterplots corresponding to the LH and KS correlation coefficients are shown in Figure 4.1. (Theoretical values obtained from equations 3.4 and 3.3 from the parameterization of this example shown below are 1/9 for r_{LH} and 1/11 for p_{KS} .) Thus the indicated conclusion from LH and KS is lack of stability, more than 90% volatility, etc.

Insert Figure 4.1

But as the consecutive improvement analysis in Table 4.2 (top portion) shows, almost 5 of 6 schools (.832) that improved yr1 to yr2 also improved yr2 to yr3. And the median improvement in yr2 to yr3 of those schools improving yr1 to yr2 is 53 points. Moreover, of the schools improving more than 50 points yr1 to yr2, nearly 4/5 (.788) also improved yr2 to yr3 (median

Table 4.1
Descriptive Statistics for School-Level Longitudinal Data and Year-to-Year Improvement: Artificial Data Example 1 (n = 10000)

	Continuous Outcome						
	Yearly Scores				Year-to-Year Improvement		
percentile	yr1	yr2	yr3	yr4	yr1-yr2	yr2-yr3	yr3-yr4
10th	431.73	496.57	556.27	614.72	-14.19	-14.26	-12.42
25th	488.67	548.84	608.7	669.13	21.14	20.98	22.01
50th	549.68	609.72	670.55	731.34	61.05	59.36	60.71
75th	612.07	671.39	729.79	792.74	99.26	97.65	99.12
90th	668.42	726.94	784.63	847.67	134.18	132.88	134.36
Mean	550.32	610.67	670.13	731.09	60.35	59.46	60.97
St.Dev	92.45	90.38	90.05	91.21	57.86	57.20	57.61

Correlation Matrices

Continuous Outcome

Yearly Scores	yr1	yr2	yr3	yr4
yr1	1.	0.8	0.784	0.757
yr2	0.8	1.	0.799	0.783
yr3	0.784	0.799	1.	0.798
yr4	0.757	0.783	0.798	1.

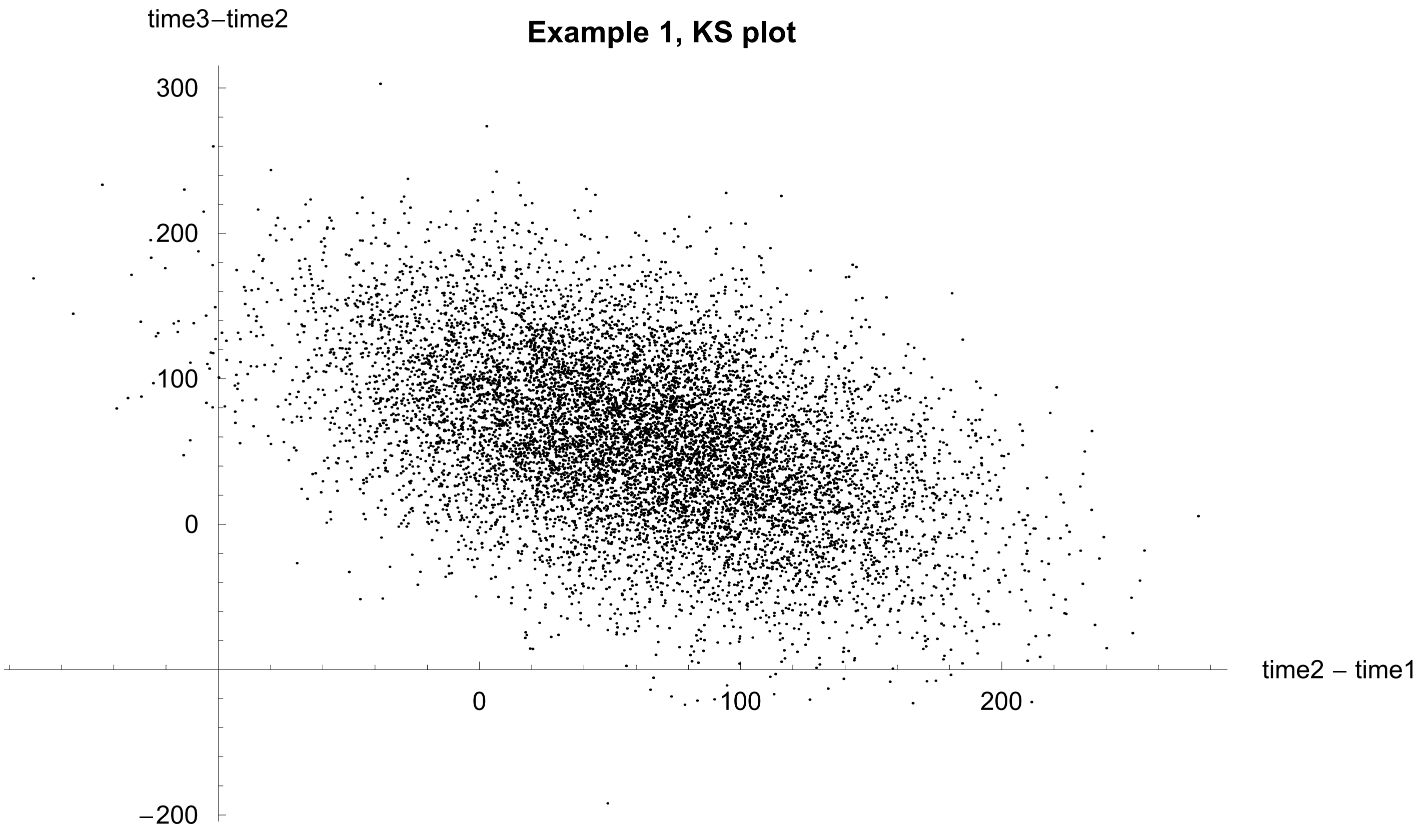
Year-to-Year Improvement	yr1-yr2	yr2-yr3	yr3-yr4
yr1-yr2	1.	-0.456	0.028
yr2-yr3	-0.456	1.	-0.457
yr3-yr4	0.028	-0.457	1.

	Proportion (proficient) Outcome			
	Yearly School Scores			
percentile	yr1	yr2	yr3	yr4
10th	0.28	0.347	0.413	0.48
25th	0.347	0.413	0.48	0.547
50th	0.413	0.493	0.56	0.627
75th	0.493	0.56	0.627	0.693
90th	0.56	0.627	0.693	0.747

Correlation Matrices

Proportion (proficient) Outcome

Yearly School Scores	yr1	yr2	yr3	yr4
yr1	1.	0.718	0.701	0.673
yr2	0.718	1.	0.711	0.694
yr3	0.701	0.711	1.	0.712
yr4	0.673	0.694	0.712	1.



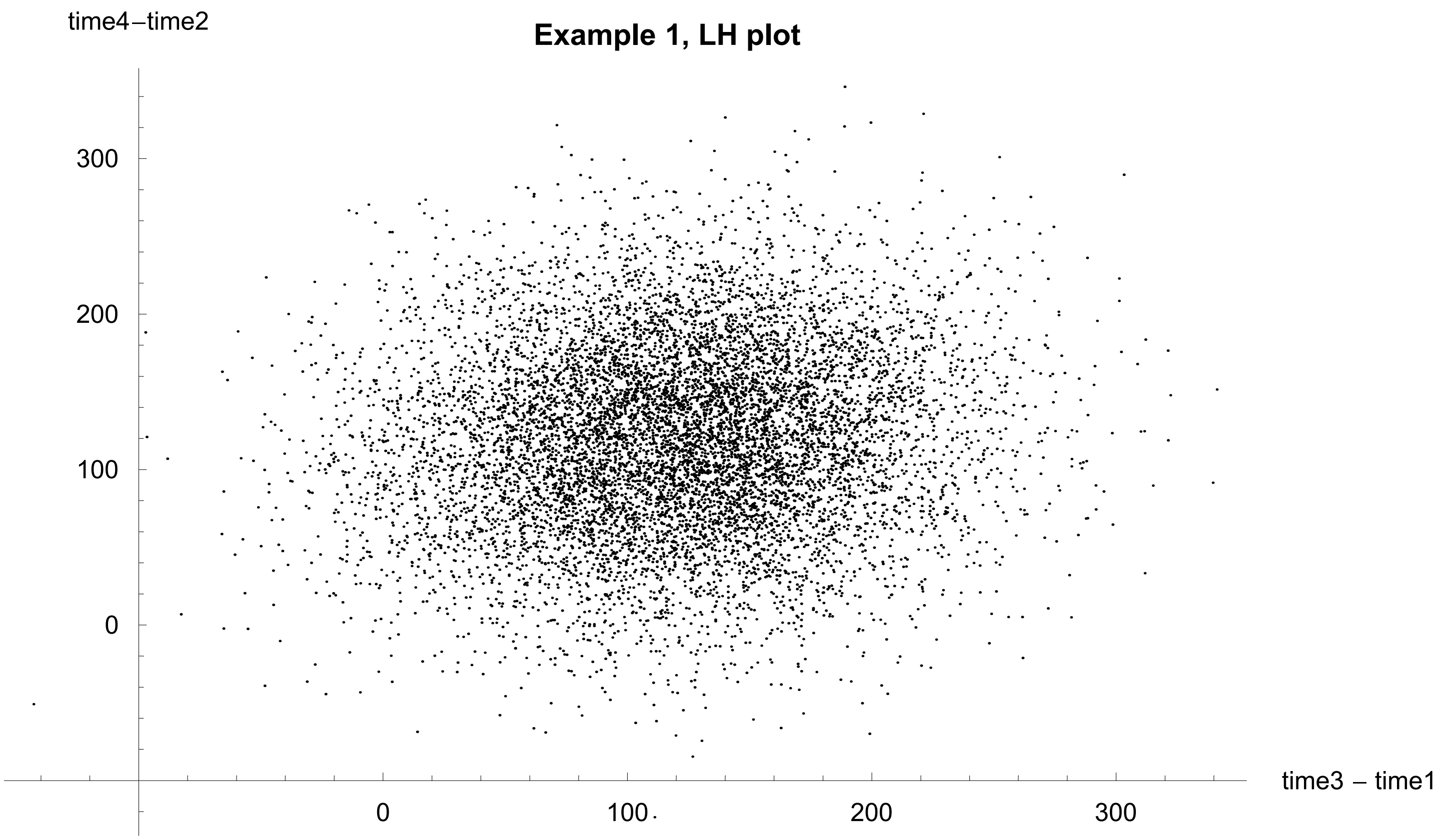


Figure 4.1. Scatterplots corresponding to the KS and LH correlation coefficients for artificial data example 1.

yr2 to yr3 improvement 42 points). From the bottom portion of Table 4.2, we see that 7068 of the 10000 schools improved in both yr1 to yr2 and in yr2 to yr3, and almost 5 of 6 (.831) of these 7068 schools also improved yr3 to yr4 (with the median improvement of these 7068 school 54 points). Similarly, nearly half (4840) of the 10000 schools improved more than 25 points in both yr1 to yr2 and yr2 to yr3, and more than 4 of 5 (.807) of these 4840 schools also improved yr3 to yr4, with these 4840 schools showing median improvement of 48 points. Is this 90% volatility?

Insert Table 4.2

Proportion (proficient) outcome. For completeness in matching LH, a corresponding analysis of example 1 employing a proportion proficient school outcome is presented. Descriptive statistics for the proportion outcome are in the bottom portion of Table 4.1, and consistency in improvement is shown in Table 4.3. The median school at yr1 has 41.3% proficient under this construction, and median year-to-year improvement in proportion proficient is greater than .06. For the proportion proficient outcome r_{LH} = .078 and p_{KS} = .047, again indicating no stability, great volatility in the proportion proficient scores. In contrast, the consecutive improvement display in Table 4.3 shows that more than 3 of 4 (.781) schools that improved yr1 to yr2 also improved yr2 to yr3, and median improvement in yr2 to yr3 of those schools improving yr1 to yr2 is .053. Similarly, nearly half (4712) of the 10000 schools improved proportion proficient more than .025 in both yr1 to yr2 and yr2 to yr3, and nearly 3 of 4 (.748) of these 4712 schools also improved yr3 to yr4, with the median improvement of these 4712 schools .04. Again, Is this great volatility?

Insert Table 4.3

Parameterization. Artificial data example 1 has observations at equally spaced times $t_1 = 1, t_2 = 2, t_3 = 3, t_4 = 4$ generated from the straight-line growth model in (3.1). The individual growth curve parameters are drawn from $\theta_s \sim N(60, 100)$ (i.e., $\mu_\theta = 60, \sigma^2_\theta = 100$) , and $\eta_s(3) \sim N(670, 6400)$, so that the correlation between θ_s and $\eta_s(3)$ is zero. Observations $Y_s(t)$ are obtained by adding to each $\eta_s(t) = \eta_s(3) + \theta_s(t - 3)$ the statistical variability (error) in school score, which is drawn from $\varepsilon_{ts} \sim N(0, 1600)$. With these scores in the metric of the California API a standard error of 40 corresponds to the standard error for a grade-level score (e.g. fourth-grades) in the smaller schools (around 45 students). Note that the ε_{ts} are uncorrelated over

Table 4.2

Consecutive Improvement for Artificial Data Example 1: Continuous Outcome

Three year Improvement: yr1, yr2, yr3 data

ImpLevel yr1 to yr2	Number exceeding ImpLevel	Proportion of those improving in yr2 to yr3	Amount of Improvement yr2 to yr3 {lowest decile lower quartile median upper quartile}
0	8500	0.832	-19. 15.8 52.6 88.8
25	7283	0.812	-23. 11.5 48. 83.8
50	5717	0.788	-27.4 7.3 42.2 78.2
75	4059	0.752	-32.9 0.5 34.6 68.8
100	2455	0.7	-39.8 -8.5 25.2 59.7

Fourth-year Improvement: yr1, yr2, yr3, yr4 data

ImpLevel yr1 to yr2 and yr2 to yr3	Number exceeding both ImpLevels	Proportion of those improving in yr3 to yr4	Amount of Improvement yr3 to yr4 {lowest decile lower quartile median upper quartile}
0	7068	0.831	-17.2 16.3 53.6 89.9
25	4840	0.807	-21.8 10.6 48.1 83.8
50	2535	0.764	-30.6 1.9 38.9 73.5

Table 4.3
Consecutive Improvement for Artificial Data Example 1: Proportion Outcome

Three year Improvement: yr1, yr2, yr3 data

ImpLevel yr1 to yr2	Number exceeding ImpLevel	Proportion of those improving in yr2 to yr3	Amount of Improvement yr2 to yr3 {lowest decile lower quartile median upper quartile}
0	7745	0.781	-0.053 0 0.053 0.107
.025	7183	0.769	-0.053 0 0.053 0.107
.05	5971	0.742	-0.053 -0.013 0.04 0.093
.075	4601	0.709	-0.067 -0.013 0.04 0.08
.10	3333	0.664	-0.069 -0.027 0.027 0.08

Fourth-year Improvement: yr1, yr2, yr3, yr4 data

ImpLevel yr1 to yr2 and yr2 to yr3	Number exceeding both ImpLevels	Proportion of those improving in yr3 to yr4	Amount of Improvement yr3 to yr4 {lowest decile lower quartile median upper quartile}
0	5646	0.766	-0.053 0 0.053 0.093
.025	4712	0.748	-0.053 -0.013 0.04 0.093
.05	2867	0.707	-0.067 -0.013 0.04 0.08

time, corresponding to the setting of successive cohorts, such as fourth grade scores in successive years, where overlap of students (same students in repeated years) does not arise. For these parameter values $\rho[D_t]$ = 1/33 and reliability coefficients for the $Y_s(t)$ are between .8 and .81.

4.2 Artificial Data Example 2: LH, KS Stability

Descriptive statistics for the 10000 school scores comprising example 2 are in Table 4.4. The top half gives results for the continuous outcome for a school, again in the metric of the California API. Median (and mean) improvement is about 40 points each year, with steady overall improvement seen throughout the score distribution. Observed score correlation between year 1 status and subsequent improvement is large and negative (as was also seen by LH p.33); in example 2, Corr($Y_s(1)$, $Y_s(2) - Y_s(1)$) is –.549, Corr($Y_s(1)$, $Y_s(3) - Y_s(2)$) is –.395 and Corr($Y_s(1)$, $Y_s(4) - Y_s(3)$) is –.398.

Insert Table 4.4

Consistency in improvement. For artificial data example 2, r_{LH} = .724 and ρ_{KS} = 1. Scatterplots corresponding to the LH and KS correlation coefficients are shown in Figure 4.2. (Theoretical values obtained from equations 3.4 and 3.3 from the parameterization of this example shown below are 8/11 for r_{LH} and 1.0 for ρ_{KS}). Thus the indicated conclusion from LH and KS is great stability, 0% volatility, etc.

Insert Figure 4.2

Table 4.5 (for example 2) displays just about the same (perhaps a little less) consistency in improvement as was seen in Table 4.2 for example 1. The top portion of Table 4.5 shows more than 3 of 4 schools (.776) that improved yr1 to yr2 also improved yr2 to yr3, and the median improvement in yr2 to yr3 of those schools improving yr1 to yr2 is 41 points. Moreover, of the schools improving more than 50 points yr1 to yr2, nearly 4/5 (.785) also improved yr2 to yr3 (median yr2 to yr3 improvement 43 points). From the bottom portion of Table 4.5, we see that 5990 of the 10000 schools improved in both yr1 to yr2 and yr2 to yr3. More than 4 of 5 (.826) of these 5990 schools also improved yr3 to yr4, with the median improvement of these 5990 schools 49 points. Similarly, 3780 of the 10000 schools improved more than 25 points in both yr1 to yr2 and yr2 to yr3. More than 5 of 6 (.857) of these 3780 schools also improved yr3 to yr4, with the median improvement of these 3780 schools 54 points. How can it be that according to LH and KS, example 1 shows complete volatility while example 2 shows great stability, even though the actual consistency in improvement is quite similar in both examples? The answer, as has been demonstrated in various ways from the examples in the introduction onward, is that neither the LH or KS methods

Table 4.4
Descriptive Statistics for School-Level Longitudinal Data and Year-to-Year Improvement: Artificial Data Example 2 (n = 10000)

	Continuous Outcome						
	Yearly Scores				Year-to-Year Improvement		
percentile	yr1	yr2	yr3	yr4	yr1-yr2	yr2-yr3	yr3-yr4
10th	479.49	541.78	592.61	624.1	-30.12	-30.82	-29.79
25th	545.78	598.71	643.14	678.15	3.53	2.99	2.92
50th	622.58	661.84	699.98	740.06	40.13	38.93	39.02
75th	693.34	723.6	757.2	800.11	76.21	77.36	75.04
90th	757.97	775.81	810.13	855.38	109.72	111.1	108.33
Mean	620.29	660.46	700.31	739.25	40.17	39.85	38.95
St.Dev	109.02	91.26	84.61	90.86	54.82	54.93	54.08

Correlation Matrices

Continuous Outcome

Yearly Scores	yr1	yr2	yr3	yr4
yr1	1.	0.865	0.677	0.393
yr2	0.865	1.	0.807	0.61
yr3	0.677	0.807	1.	0.812
yr4	0.393	0.61	0.812	1.

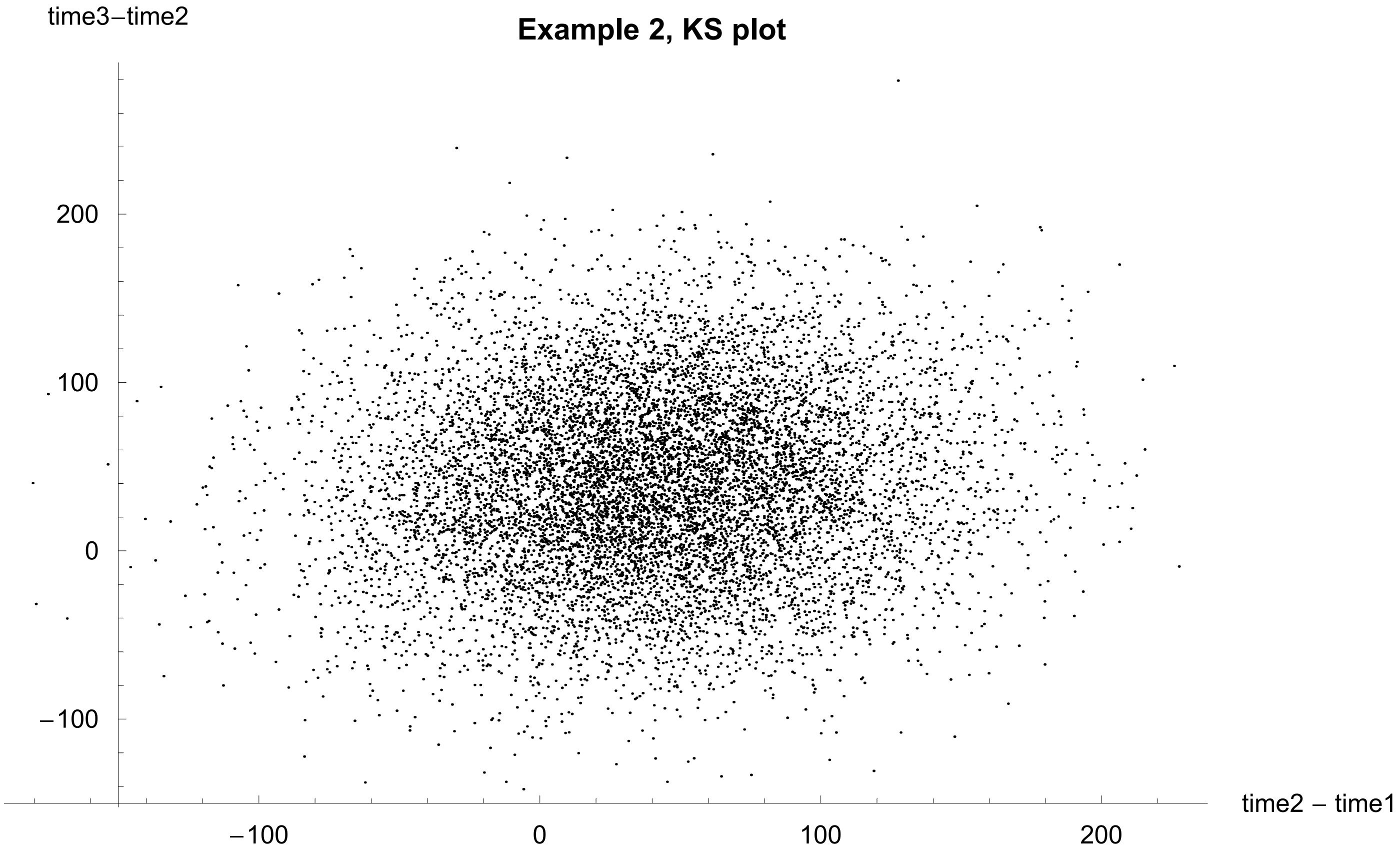
Year-to-Year Improvement	yr1-yr2	yr2-yr3	yr3-yr4
yr1-yr2	1.	0.089	0.394
yr2-yr3	0.089	1.	0.089
yr3-yr4	0.394	0.089	1.

	Proportion (proficient) Outcome			
	Yearly School Scores			
percentile	yr1	yr2	yr3	yr4
10th	0.187	0.253	0.32	0.36
25th	0.267	0.333	0.4	0.453
50th	0.373	0.427	0.48	0.547
75th	0.48	0.52	0.573	0.64
90th	0.587	0.613	0.653	0.72

Correlation Matrices

Proportion (proficient) Outcome

Yearly School Scores	yr1	yr2	yr3	yr4
yr1	1.	0.815	0.625	0.368
yr2	0.815	1.	0.748	0.566
yr3	0.625	0.748	1.	0.756
yr4	0.368	0.566	0.756	1.



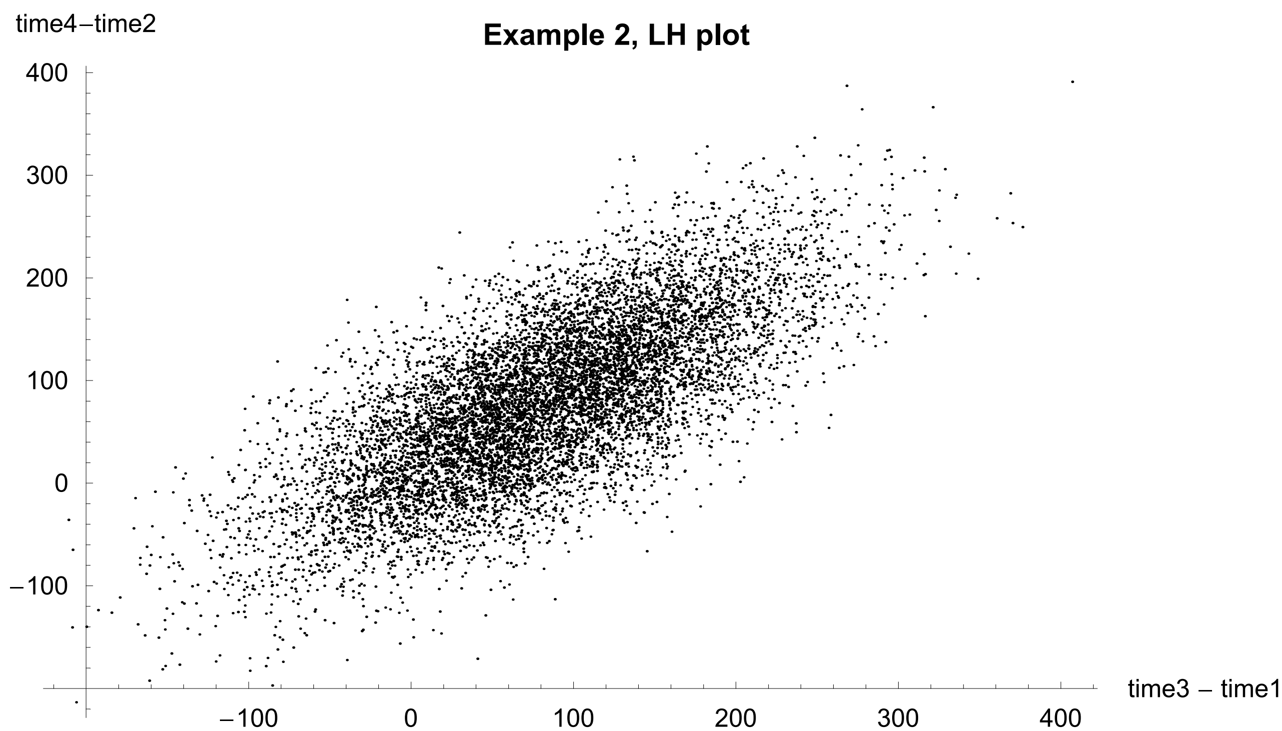


Figure 4.2. Scatterplots corresponding to the KS and LH correlation coefficients for artificial data example 2.

provide useful information about stability or consistency in improvement.

Insert Table 4.5

Proportion (proficient) outcome. For completeness, descriptive statistics for the proportion proficient school outcome in example 2 are presented in the bottom portion of Table 4.4, and consistency in improvement is presented in Table 4.6. The median school at yr1 has 37.3% proficient under this construction, and median year-to-year improvement in proportion proficient is greater than .05. For the proportion proficient outcome r_{LH} = .601 and p_{KS} = .806. These values indicate good stability, small volatility in the proportion proficient scores, but provide a slightly weaker indication than the indices for the continuous outcome. The consecutive improvement display in Table 4.6 has values quite comparable to Table 4.3 for example 1, even though in example 1 great volatility was indicated by KS and LH methods. In Table 4.6 almost 3 of 4 (.741) schools that improved yr1 to yr2 also improved yr2 to yr3 And median improvement in yr2 to yr3 of those schools improving yr1 to yr2 is .053. Similarly, 4197 of the 10000 schools improved proportion proficient more than .025 in both yr1 to yr2 and in yr2 to yr3, and nearly 4 of 5 (.79) of these 4197 schools also improved yr3 to yr4 (median improvement .067).

Insert Table 4.6

Parameterization. Artificial data example 2 has observations at equally spaced times $t_1 = 1$, $t_2 = 2$, $t_3 = 3$, $t_4 = 4$ generated from the straight-line growth model in (3.1). The individual growth curve parameters are drawn from $\theta_s \sim N(40, 1200)$ (i.e., $\mu_\theta = 40$, $\sigma^2_\theta = 1200$) , and $\eta_s(3) \sim N(700, 6400)$ so that the correlation between θ_s and $\eta_s(3)$ is zero. Observations $Y_s(t)$ are obtained by adding to each $\eta_s(t) = \eta_s(3) + \theta_s(t - 3)$ the statistical variability (error) in school score, which is drawn from $\varepsilon_{ts} \sim N(0, 900)$. With these scores in the metric of the California API a standard error of 30 corresponds to the standard error for a grade-level score (e.g. fourth-grades) in the median-sized schools (around 80 students). Note that the ε_{ts} are uncorrelated over time, corresponding to the setting of successive cohorts, such as fourth grade scores in successive years, where overlap of students (same students in repeated years) does not arise. For these parameter values $\rho[D_t] = 2/5$, and reliability coefficients for the $Y_s(t)$ are between .88 and .92.

Table 4.5
Consecutive Improvement for Artificial Data Example 2: Continuous Outcome

Three year Improvement: yr1, yr2, yr3 data

ImpLevel yr1 to yr2	Number exceeding ImpLevel	Proportion of those improving in yr2 to yr3	Amount of Improvement yr2 to yr3 {lowest decile lower quartile median upper quartile}
0	7716	0.776	-29.1 4.8 41.1 79.3
25	6080	0.779	-28.3 5.4 42.1 80.8
50	4280	0.785	-26.8 6.8 43.2 81.8
75	2595	0.797	-25.5 9.8 45. 82.3
100	1364	0.802	-23.1 10.8 44.5 83.1

Fourth-year Improvement: yr1, yr2, yr3, yr4 data

ImpLevel yr1 to yr2 and yr2 to yr3	Number exceeding both ImpLevels	Proportion of those improving in yr3 to yr4	Amount of Improvement yr3 to yr4 {lowest decile lower quartile median upper quartile}
0	5990	0.826	-18.5 13.6 48.5 83.8
25	3780	0.857	-12.2 19.5 54.3 89.2
50	1942	0.893	-2.5 26.2 59.8 93.3

Table 4.6
Consecutive Improvement for Artificial Data Example 2: Proportion Outcome

Three year Improvement: yr1, yr2, yr3 data

ImpLevel yr1 to yr2	Number exceeding ImpLevel	Proportion of those improving in yr2 to yr3	Amount of Improvement yr2 to yr3 {lowest decile lower quartile median upper quartile}
0	7107	0.741	-0.067 -0.013 0.053 0.12
.025	6575	0.738	-0.067 -0.013 0.053 0.12
.05	5410	0.729	-0.08 -0.013 0.053 0.12
.075	4212	0.72	-0.08 -0.013 0.053 0.107
.10	3100	0.71	-0.08 -0.013 0.04 0.107

Fourth-year Improvement: yr1, yr2, yr3, yr4 data

ImpLevel yr1 to yr2 and yr2 to yr3	Number exceeding both ImpLevels	Proportion of those improving in yr3 to yr4	Amount of Improvement yr3 to yr4 {lowest decile lower quartile median upper quartile}
0	4920	0.787	-0.053 0 0.067 0.133
.025	4197	0.79	-0.053 0 0.067 0.133
.05	2779	0.795	-0.048 0.013 0.067 0.133

4.3 Artificial Data Example 3: Up-and-Down, LH, KS stability

Example 3 is an orchestrated example designed to demonstrate that neither the LH or KS methods will detect extreme inconsistency in improvement (real volatility). Individual school trajectory templates for the up-and-down examples are shown in frames c and d of Figures 1.1. Details on the construction of these examples are given in the parameterization paragraph below. Descriptive statistics for three-year (for KS) and four-year (for LH) versions of these up-and-down examples are given in Table 4.7. In the three-year example, during yr2-yr3 the scores reverse the improvement made yr1-yr2. In the four-year example scores improve yr1-yr2, plateau in yr2-yr3, and reverse the improvement yr3-yr4 (with the same correlation matrices as in example 2, Table 4.4). In both versions Corr($Y_s(1)$, $Y_s(2) - Y_s(1)$) is –.549 and Corr($Y_s(1)$, $Y_s(3) - Y_s(2)$) is –.395.

Insert Table 4.7

Consistency in improvement. For the three-year version of up-and-down p_{KS} = 1, and for the four-year version r_{LH} = .724. Scatterplots corresponding to the LH and KS correlation coefficients are shown in Figure 4.3. Thus the indicated conclusion from LH and KS is great stability, 0% volatility, etc.

Insert Figure 4.3

Table 4.8 displays consistency in improvement (or lack thereof) information for the three-year example in the top portion, and separately, for the four-year example in the bottom portion. Even though KS determine 0% volatility, less than 1 of 4 schools (.246) that improved yr1 to yr2 also improved yr2 to yr3. And the median improvement in yr2 to yr3 of those schools improving yr1 to yr2 is –39 points! In the four-year version the bottom portion of Table 4.8 shows that 3918 of the 10000 schools improved in both yr1 to yr2 and yr2 to yr3 but less than 1 of 3 (.281) of these 3918 schools also improved yr3 to yr4, with the median improvement of these 3918 schools –30 points. Yet LH determine strong stability.

Insert Table 4.8

Proportion (proficient) outcome. For completeness, descriptive statistics for the proportion proficient school outcome for the four-year example are presented in the bottom portion of Table 4.7, and consistency in

Table 4.7
Descriptive Statistics for School-Level Longitudinal Data and Year-to-Year Improvement: Artificial Data Example 3 (n = 10000)

Continuous Outcome							
	Yearly Scores, Three-wave example			Yearly Scores Four-wave example			
percentile	yr1	yr2	yr3	yr1	yr2	yr3	yr4
10th	479.49	541.78	512.6	479.49	541.78	552.61	504.1
25th	545.78	598.71	563.14	545.78	598.71	603.14	558.15
50th	622.58	661.84	619.98	622.58	661.84	659.98	620.06
75th	693.34	723.6	677.2	693.34	723.6	717.2	680.11
90th	757.97	775.81	730.13	757.97	775.81	770.13	735.38
Mean	620.29	660.46	620.31	620.29	660.46	660.31	619.25

Correlation Matrices (same as example 2)

Proportion (proficient) Outcome, Four-wave example

Yearly School Scores

percentile	yr1	yr2	yr3	yr4
10th	0.187	0.253	0.267	0.2
25th	0.267	0.333	0.333	0.28
50th	0.373	0.427	0.427	0.36
75th	0.48	0.52	0.52	0.453
90th	0.587	0.613	0.587	0.547

Correlation Matrix

	yr1	yr2	yr3	yr4
yr1	1.	0.718	0.701	0.673
yr2	0.718	1.	0.711	0.694
yr3	0.701	0.711	1.	0.712
yr4	0.673	0.694	0.712	1.

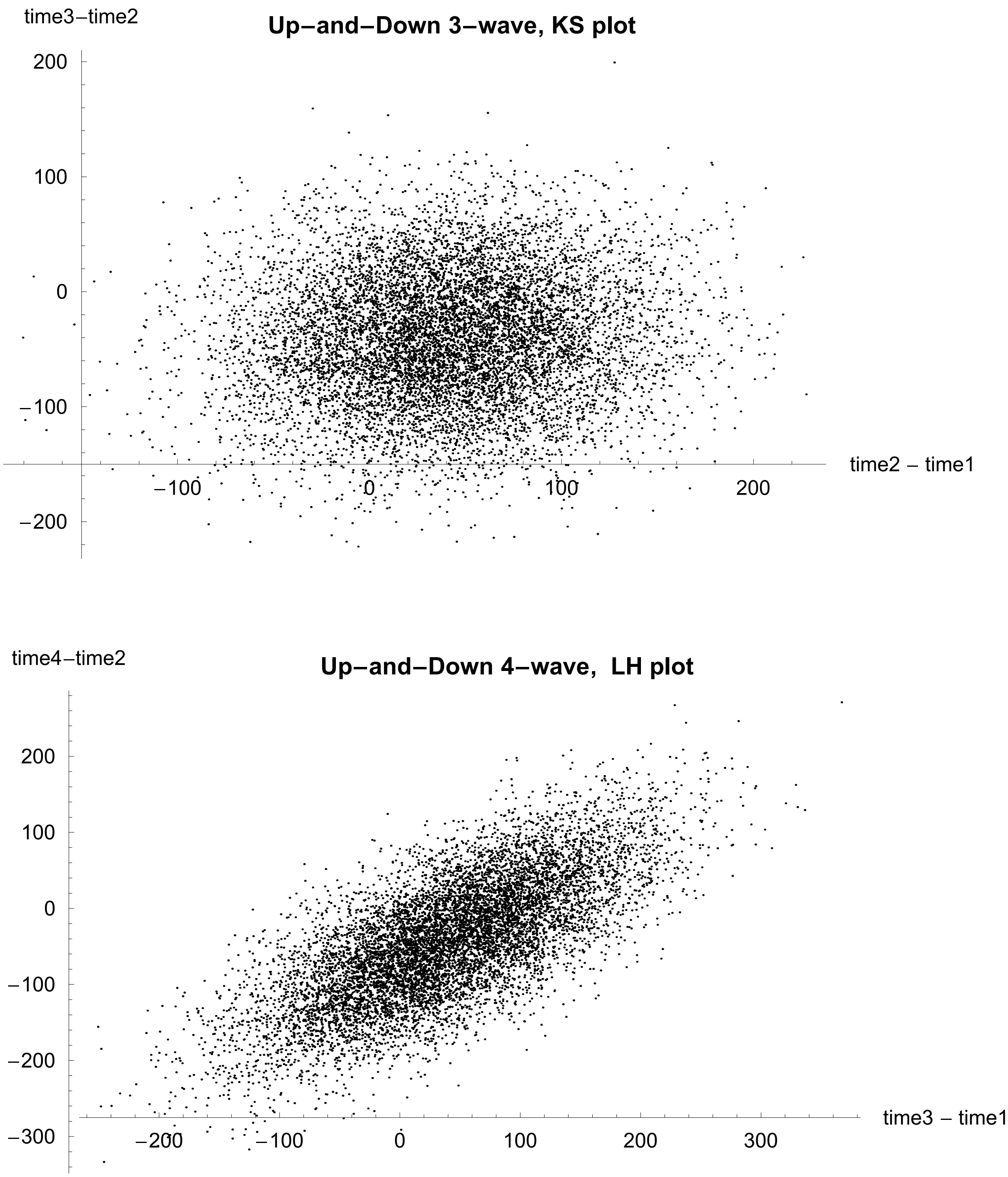


Figure 4.3. Scatterplots corresponding to the KS and LH correlation coefficients for artificial data example 3 (up-and-down examples).

Table 4.8
Consecutive Improvement for Artificial Data Example 3: Continuous Outcome

Three year Example: yr1, yr2, yr3 data

ImpLevel yr1 to yr2	Number exceeding ImpLevel	Proportion of those improving in yr2 to yr3	Amount of Improvement yr2 to yr3 {lowest decile lower quartile median upper quartile}
0	7716	0.246	-109.1 -75.2 -38.9 -0.7
25	6080	0.254	-108.2 -74.6 -37.9 0.8
50	4280	0.262	-106.8 -73.2 -36.8 1.8
75	2595	0.267	-105.5 -70.2 -35. 2.3
100	1364	0.278	-103.1 -69.2 -35.5 3.1

Four-year Example: yr1, yr2, yr3, yr4 data

ImpLevel yr1 to yr2 and yr2 to yr3	Number exceeding both ImpLevels	Proportion of those improving in yr3 to yr4	Amount of Improvement yr3 to yr4 {lowest decile lower quartile median upper quartile}
0	3918	0.281	-96.3 -64.1 -29.8 5.1
25	2063	0.32	-89.4 -59.9 -24.4 11.6
50	871	0.361	-78.6 -52.7 -17.7 16.4

improvement is presented in Table 4.9. The median school at yr1 has 37.3% proficient under this construction, and the up-and-down pattern of these data reverses the improvement by year4 back to the level of year1. Yet for this proportion proficient outcome r_{LH} = .602 (same as for the consistent improvement in proportions seen in example 2). Table 4.9 presents the four-year consistency table for the proportion, showing for example that 3172 of the 10000 schools improved proportion proficient in both yr1 to yr2 and yr2 to yr3, but less than 1 of 3 (.321) of these 3172 schools also improved yr3 to yr4, with median yr3-yr4 improvement of –.053.

Insert Table 4.9

Parameterization. For the three-year up-and-down example (c.f. Fig. 1.1c) with observations at equally spaced times $t_1 = 1$, $t_2 = 2$, $t_3 = 3$, start with the configuration of Example 2, but change each $\eta_s(3)$ to $\eta_s(3) - 2\mu_\theta$. As in example 2 statistical variability (error variance in school score) in $Y_s(t)$ is drawn from $\varepsilon_{ts} \sim N(0, 900)$. Thus correlations and reliability coefficients are unchanged from Example 2. For the four-year up-and-down example (c.f. Fig. 1.1d) with observations at equally spaced times $t_1 = 1$, $t_2 = 2$, $t_3 = 3$, $t_4 = 4$ start with the configuration of Example 2, but change each $\eta_s(3)$ to $\eta_s(3) - \mu_\theta$ and change each $\eta_s(4)$ to $\eta_s(4) - 3\mu_\theta$. As in example 2 statistical variability (error variance in school score) in $Y_s(t)$ is drawn from $\varepsilon_{ts} \sim N(0, 900)$. Here also correlations and reliability coefficients are unchanged from Example 2.

Table 4.9
Consecutive Improvement for Artificial Data Example 3:
Four Year Example, Proportion Outcome

Fourth-year Improvement: yr1, yr2, yr3, yr4 data

ImpLevel yr1 to yr2 and yr2 to yr3	Number exceeding both ImpLevels	Proportion of those improving in yr3 to yr4	Amount of Improvement yr3 to yr4 {lowest decile lower quartile median upper quartile}
0	3172	0.321	-0.16 -0.12 -0.053 0.013
.025	2529	0.326	-0.16 -0.12 -0.053 0.013
.05	1526	0.331	-0.16 -0.12 -0.053 0.013

4.4 Artificial Data Example 4: Proportional Deceleration, LH, KS Volatility

The final example uses the proportional deceleration growth model for school trajectories (consistent improvement of lesser magnitude over time). Descriptive statistics for the 10000 (continuous outcome) school scores comprising example 4 are in Table 4.10. Median (and mean) improvement is about 60 points yr1-yr2, 40 points yr2-yr3, and 27 points yr3-yr4, consistent with the h=2/3 value used in the proportional deceleration growth model (see paramaterization below). Correlations between the yearly scores are slightly higher than in example 1, and correlations between year-to-year improvements are quite similar to example 1. Observed score correlation between year 1 status and subsequent improvement is large and negative: Corr($Y_s(1)$, $Y_s(2) - Y_s(1)$) is –.276, and Corr($Y_s(1)$, $Y_s(4) - Y_s(1)$) is –.305.

Insert Table 4.10

Consistency in improvement. For the example 4 data r_{LH} = .063 and p_{KS} = .111. Scatterplots corresponding to the LH and KS correlation coefficients are shown in Figure 4.4. (Theoretical values obtained from equations 3.6 and 3.5 from the parameterization of this example are .071 for r_{LH} and .083 for p_{KS}). Thus the indicated conclusion from LH and KS is great volatility, lack of stability, etc.

Insert Figure 4.4

Table 4.11 shows reasonable consistency in improvement, in some respects not as strong as examples 1 and 2. The top portion of Table 4.11 shows that more than 4 of 5 schools (.816) improving yr1 to yr2 also improved yr2 to yr3. And the median improvement in yr2 to yr3 of those schools improving yr1 to yr2 is 37 points. Moreover, of the schools improving more than 50 points yr1 to yr2, more than 3 of 4 (.755) also improved yr2 to yr3 (median yr2 to yr3 improvement 27 points). In addition, from the bottom portion of Table 4.11, we see that 7452 of the 10000 schools improved in both yr1 to yr2 and in yr2 to yr3, and more than 2 of 3 (.692) of these 7452 schools also improved yr3 to yr4 (with the median improvement of these 7452 schools 20 points).

Insert Table 4.11

Parameterization. Artificial data example 4 has observations at equally

Table 4.10
Descriptive Statistics for School-Level Longitudinal Data and Year-to-Year Improvement: Artificial Data Example 4 (n = 10000)

	Continuous Outcome						
	Yearly Scores				Year-to-Year Improvement		
percentile	yr1	yr2	yr3	yr4	yr1-yr2	yr2-yr3	yr3-yr4
10th	437.05	499.76	541.3	568.55	3.49	-14.13	-27.97
25th	491.31	552.19	592.16	619.65	31.03	11.03	-1.94
50th	551.36	609.64	649.79	676.12	60.06	39.7	26.74
75th	608.65	668.12	706.02	734.21	88.42	68.99	55.66
90th	661.53	720.38	759.5	785.81	115.32	95.07	81.8
Mean	550.58	610.20	650.17	676.85	59.62	39.97	26.68
St.Dev	87.08	85.85	84.99	85.20	43.07	42.59	42.92

Correlation Matrices

	Continuous Outcome							
	Yearly Scores					Year-to-Year Improvement		
	yr1	yr2	yr3	yr4		yr1-yr2	yr2-yr3	yr3-yr4
yr1	1.	0.876	0.863	0.857	yr1-yr2	1.	-0.444	0.009
yr2	0.876	1.	0.876	0.872	yr2-yr3	-0.444	1.	-0.488
yr3	0.863	0.876	1.	0.873	yr3-yr4	0.009	-0.488	1.
yr4	0.857	0.872	0.873	1.				

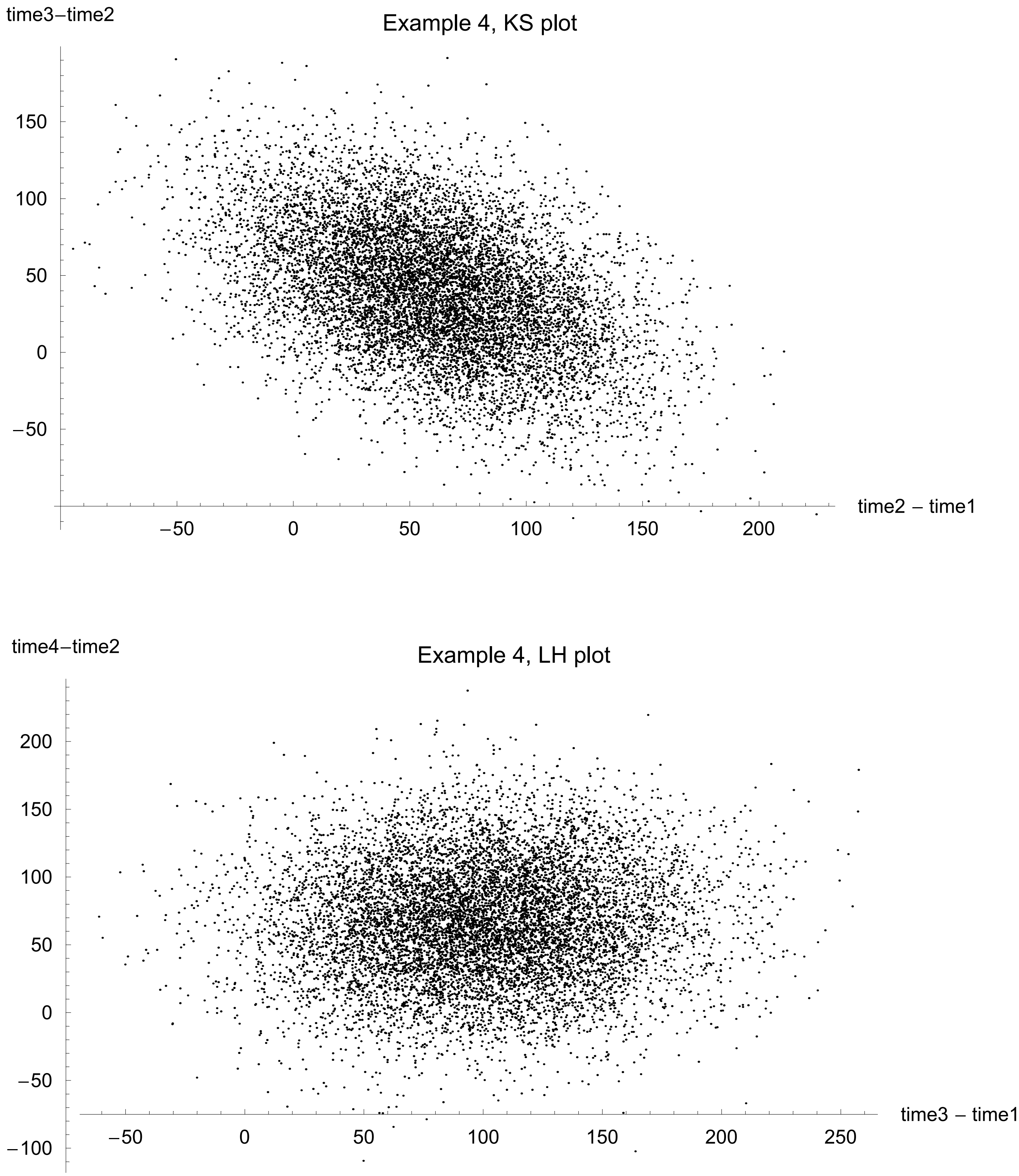


Figure 4.4. Scatterplots corresponding to the KS and LH correlation coefficients for artificial data example 4.

Table 4.11
Consecutive Improvement for Artificial Data Example 4: Continuous Outcome

Three year Improvement: yr1, yr2, yr3 data

ImpLevel yr1 to yr2	Number exceeding ImpLevel	Proportion of those improving in yr2 to yr3	Amount of Improvement yr2 to yr3 {lowest decile lower quartile median upper quartile}
0	9136	0.816	-15.9 8.7 36.5 64.6
25	7916	0.795	-18.6 5.8 33.1 60.6
50	5901	0.755	-23.8 0.5 27.1 54.2
75	3592	0.696	-30.4 -6.5 19.6 46.6
100	1772	0.624	-39.5 -14.3 12.4 37.6

Fourth-year Improvement: yr1, yr2, yr3, yr4 data

ImpLevel yr1 to yr2 and yr2 to yr3	Number exceeding both ImpLevels	Proportion of those improving in yr3 to yr4	Amount of Improvement yr3 to yr4 {lowest decile lower quartile median upper quartile}
0	7452	0.692	-32.6 -7.4 20. 47.
25	4560	0.626	-40.1 -15.1 12.5 38.1
50	1665	0.505	-50.7 -26. 0.6 29.2

spaced times $t_1 = 1$, $t_2 = 2$, $t_3 = 3$, $t_4 = 4$ following the proportional deceleration growth model in (3.2) and Figure 1b. The individual growth curve parameters are drawn from $\theta_s \sim N(60, 75)$ (i.e., $\mu_\theta = 60$, $\sigma^2_\theta = 75$) , and $\eta_s(3) \sim N(670, 6400)$ so that the correlation between θ_s and $\eta_s(3)$ is zero. The $\eta_s(t)$ for the straight-line growth model (3.1) are obtained from $\eta_s(t) = \eta_s(3) + \theta_s(t - 3)$. Then the $\eta_s(t)$ for the proportional deceleration model are modified from the straight-line values by changing each $\eta_s(3)$ to $\eta_s(3) - \theta_s(1 - h)$ and by changing each $\eta_s(4)$ to $\eta_s(4) - \theta_s(2 - h - h^2)$. In example 4, $h = 2/3$. Observations $Y_s(t)$ are obtained adding to each $\eta_s(t)$ the statistical variability (error) in school score which is drawn from $\varepsilon_{ts} \sim N(0, 900)$. With these scores in the metric of the California API a standard error of 30 corresponds to the standard error for a grade-level score (e.g. fourth-grades) in the median-sized schools (around 80 students). For these parameter values $\rho[D_1] = 1/25$, $\rho[D_2] = 1/55$, $\rho[D_3] = 2/245$, and reliability coefficients for the $Y_s(t)$ are around .88.

5. Discussion

The specific goals of this work are:

1. Dissuade researchers from using LH or KS methods in any future data analyses, and convince researchers and policy makers to discard any empirical and substantive conclusions and recommendations voiced by LH and KS or by other users of their methods.
2. Present and illustrate useful data analysis approaches for the important question of consistency of improvement in school-level accountability indices.

Both LH and KS assert conclusions about consistency in improvement without conducting analyses of consistency in improvement. Thus the disconnect between their methods and common sense and good research. In regard to KS, it's important to note that in addition to consistency in improvement, KS also present methodology and results for "volatility" of single year scores, of improvement, and properties of accountability systems. Rogosa (2002) provides a complete catalog of the KS blunders in all of these endeavors.

Consistency in improvement is by no means guaranteed in educational accountability data. The "see-saw" is a legitimate empirical conjecture, albeit one that is repudiated by the California API data and its analysis. For the CSAP data presented in LH, whether or not the school scores show consistency in improvement is unknown.

Closing Caricature. One summary of the methodological results and messages is the 4-year data caricature in Exhibit 1. The caricature emphasizes the key distinction between the progress of individual schools and the (irrelevant) between-school differences in progress. All schools improve consistently, yet KS methods determine high volatility and LH methods determine low stability.

Insert Exhibit 1

Final words from Wohlwill. Thirty years ago, researchers were eloquently warned away from correlational approaches for the assessment of stability. Wohlwill's context was behavioral development but his statistical insights pertain across the spectrum of inquiry about assessments of stability:

> In sum, the field of stability and continuity is one strewn with roadblocks of the investigators' own making, arising from the purely correlational

Exhibit 1: Consistency In Improvement Caricature

Time 1

Collection of Schools

a. True measurements for the collection of schools are distributed Discrete Uniform [498, 502], i.e., mass 1/5 at 498,..., 502. Under perfect measurement the school scores would have mean 500, variance 2.

b. Error process obscuring the school scores are Discrete Uniform [-3, 3], i.e., mass 1/7 at -3, -2, -1, 0, 1, 2, 3. Error variance is 4 points.

c. The observed school scores have magnitude around 500, and error is at most +/- 3 points. (Standard error of the school score is 2.) That may seem a reasonably precise assessment.

Time 2

a. True improvement for each school is drawn from Discrete Uniform [49, 51], mass 1/3 at 49, 50, 51.

b. The observed score for each school at time2 is the sum of the time1 true score, true improvement, and a draw from the error process obscuring the school scores: Discrete Uniform [-3, 3], i.e., mass 1/7 at -3, -2, -1, 0, 1, 2, 3.

c. The observed improvement is around 50, with maximum error +/- 6 points (extremes attained with probability .02). All schools have observed improvement of at least 43 points. Standard error of a school's improvement is 2.83.

d. Reliability of time1-time2 difference score is 1/13 = .077.

Time 3

a. True improvement for each school time2 to time3 is identical to the time1, time2 true improvement.

b. Observed score for each school at time3 is the sum of the time2 true score, true improvement, and a new draw from the error process obscuring the school scores: Discrete Uniform [-3, 3], i.e., mass 1/7 at -3, -2, -1, 0, 1, 2, 3.

c. The observed time2, time3 improvement is around 50. Consistency in improvement is very very strong. All schools improved at least 43 points time1 to time2, and the consecutive improvement is perfect, because all schools also improve at least 43 points time2 to time3.

d. KS would determine the "proportion of variance in changes due to nonpersistent factors" to be .77 (proportion persistent 3/13 = .231). High volatility!

Time 4

a. True improvement for each school time3 to time4 is identical to the time1,time2 true improvement.

b. Observed score for each school at time4 is the sum of the time3 true score, true improvement, and a new draw from the error process obscuring the school scores: Discrete Uniform [-3, 3], i.e., mass 1/7 at -3, -2, -1, 0, 1, 2, 3..

c. The observed time3, time4 improvement is again around 50. Consecutive improvement is perfect over all 3 cycles--all schools improve at least 43 points from time1 to time2 and from time2 to time3, and all of those schools again also improve at least 43 points time3 to time4. Flawless consistency in improvement.

d. LH would determine a stability coefficient of 1/4. Low stability.

sense in which the terms are typically taken, combined with the all too human tendency to lose sight of this limitation in discussions of the subject. (It is interesting to note that the same difficulty is encountered in the field of educational measurement, in the interpretation of test-retest correlations between successive measures of a single test.) Wohlwill, 1973, p.373

The problem of the stability of behavior over age has attracted an enormous amount of attention on the part of developmental researchers, both in the domain of intelligence and abilities and in the personality area. ... The intensity of this concern appears, unfortunately, to have exceeded the imaginativeness of the approaches taken to investigate this question. With only very few exceptions, work on stability, as already pointed out, has consisted in the endless proliferation of correlation coefficients, to indicate the degree of relationship between measures of behavior obtained over some given time interval. ... The result has been that we have learned a little about the "behavior" of variables over age, but nothing concerning the behavior of individuals. Furthermore, even at the level of the focus on dimensions, the information contained in bare correlation coefficients is meager at best, and conveys little sense of a system undergoing change-or, for that matter, remaining invariant in the midst of change. Wohlwill, 1973, p.358-9

References

Kane, T. J., and Staiger, D. O. (2002) "Volatility in School Test Scores: Implications for Test-Based Accountability Systems." Brookings Papers on Education Policy, 2002 (Washington, DC: Brookings Institution), 235-269.

Linn, R. L., & Haug, C. (2002). Stability of School-Building Accountability Scores and Gains. *Educational Evaluation and Policy Analysis, 24*, 29-36.

Rogosa, D. R. (1993). Individual unit models versus structural equations: Growth curve examples. In Statistical modeling and latent variables, K. Haagen, D. Bartholomew, and M. Diestler, Eds. Amsterdam: Elsevier North Holland, 259-281.

Rogosa, D. R. (1995). Myths and methods: "Myths about longitudinal research," plus supplemental questions. In The analysis of change, J. M. Gottman, Ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 3-65.

Rogosa, D.R. (2000). Interpretive Notes for the Academic Performance Index California Department of Education, Policy and Evaluation Division November 2000. California Department of Education website: http://www.cde.ca.gov/psaa/apiresearch.htm

Rogosa, D. R. (2001a). Year 2000 Update: Interpretive Notes for the Academic Performance Index. October 2001. California Department of Education website: http://www.cde.ca.gov/psaa/apiresearch.htm

Rogosa, D. R. (2001b). Year 2001 Growth Update: Interpretive Notes for the Academic Performance Index. December 2001. California Department of Education website: http://www.cde.ca.gov/psaa/apiresearch.htm

Rogosa, D. R. (2002). Irrelevance of Reliability Coefficients to Accountability Systems: Statistical Disconnect in Kane-Staiger "Volatility in School Test Scores" CRESST, November 2002. Available from http://www-stat.stanford.edu/~rag/api/kscresst.pdf

Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. Psychological Bulletin, 92, 726-748.

Rogosa, D. R., Floden, R. E., & Willett, J. B. (1984). Assessing the stability of teacher behavior. Journal of Educational Psychology, 76, 1000-1027.

Rogosa, D. R., & Willett, J. B. (1983a). Comparing two indices of tracking. Biometrics, 39, 795-6.

Rogosa, D. R., & Willett, J. B. (1983b). Demonstrating the reliability of the difference score in the measurement of change. Journal of Educational Measurement, 20, 335-343.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. Psychometrika, 50, 203-228.

Rogosa, D. R., & Saner, H. M. (1995). Longitudinal data analysis examples with random coefficient models. Journal of Educational and Behavioral Statistics, 20, 149-170.

Wohlwill, J. F. (1973). *The study of behavior development*. New York: Academic Press.

Appendix A

Additional Data Displays: Data Display Archive

Correlation Matrices for School-Level Longitudinal Data and Year-to-Year Improvement: California API Index for Elementary Schools

Scatterplots for LH correlation coefficient, California API Index:

- Elementary Schools 1999-2002
- SD Subgroups, Elementary Schools 1999-2002
- SD Subgroups in HighSD Elementary Schools 1999-2002

Correlation Matrices for School-Level Longitudinal Data and Year-to-Year Improvement: California API Index for Elementary Schools

California Elementary Schools (n=4644)

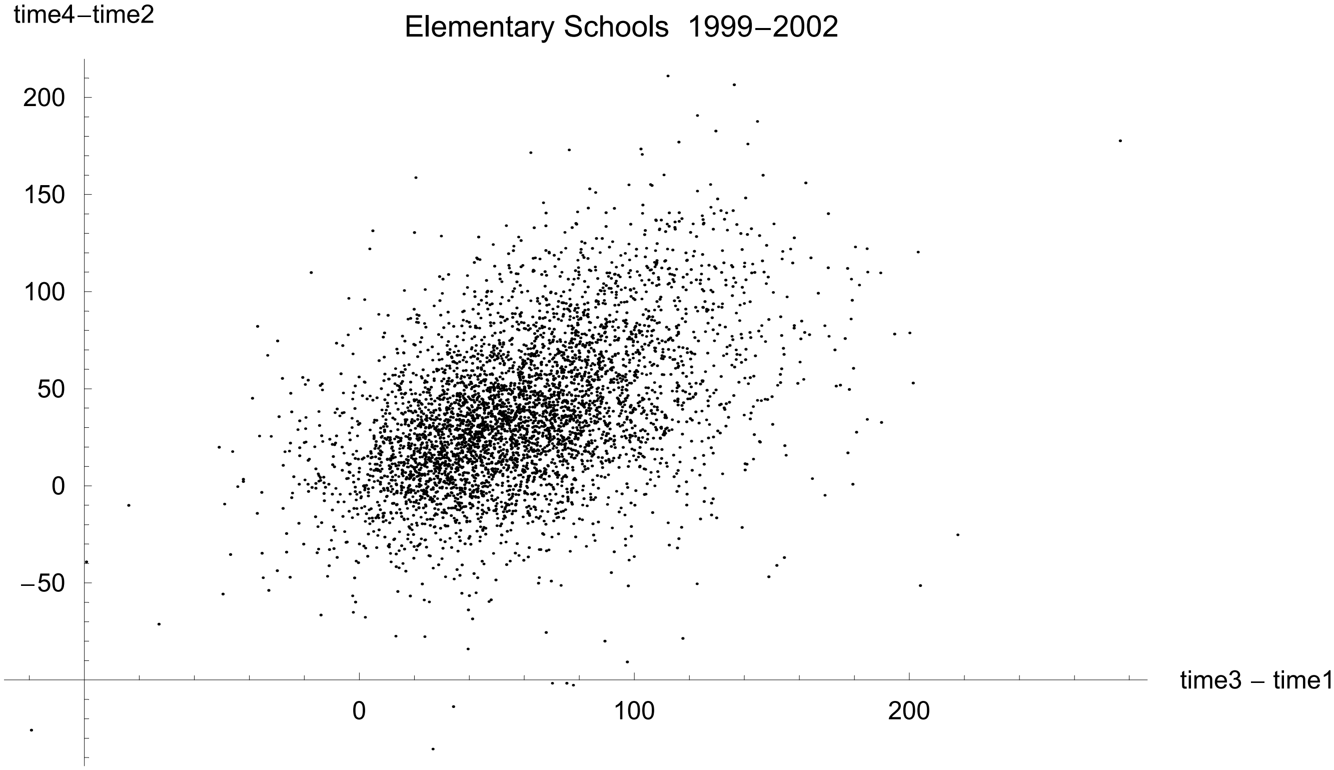
	Yearly School Scores					Year-to-Year Improvement		
	1999	2000	2001	2002		'99-00	'00-01	'01-02
1999	1.	0.977	0.964	0.946	'99-00	1.	-0.139	0.019
2000	0.977	1.	0.978	0.958	'00-01	-0.139	1.	-0.007
2001	0.964	0.978	1.	0.977	'01-02	0.019	-0.007	1.
2002	0.946	0.958	0.977	1.				
Mean	633.17	671.49	692.48	708.73		38.32	20.99	16.26
St.Dev	136.63	130.82	124.37	117.53		29.08	27.58	26.62

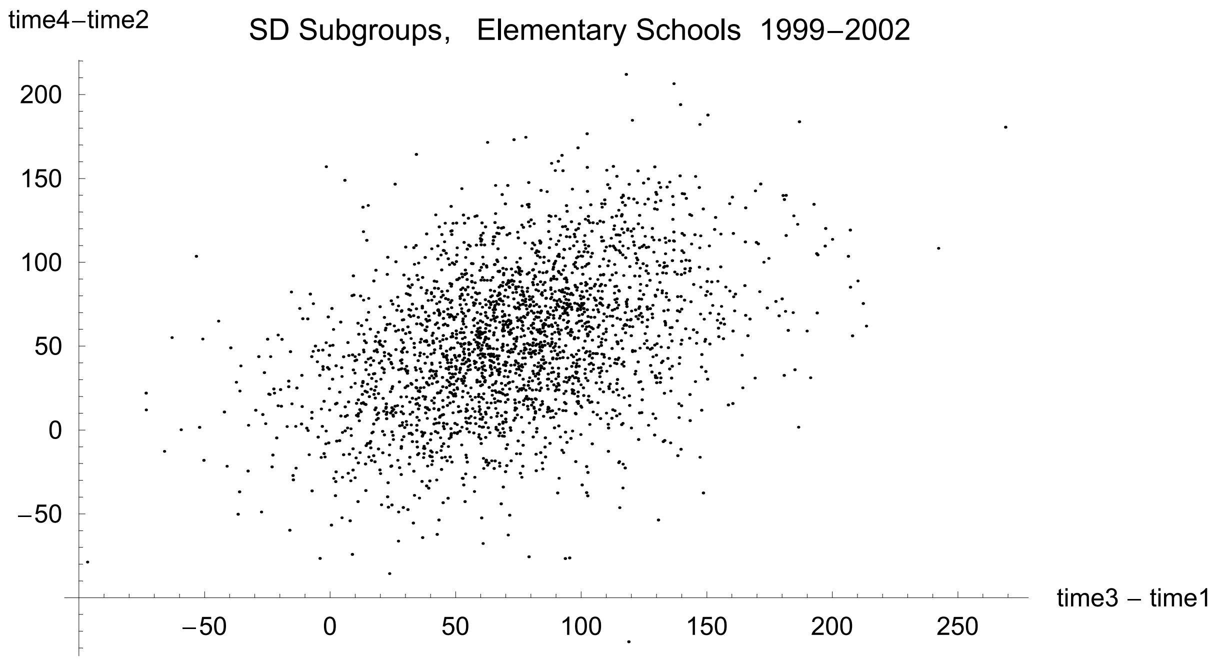
SD Subgroups in California Elementary Schools (n=2520)

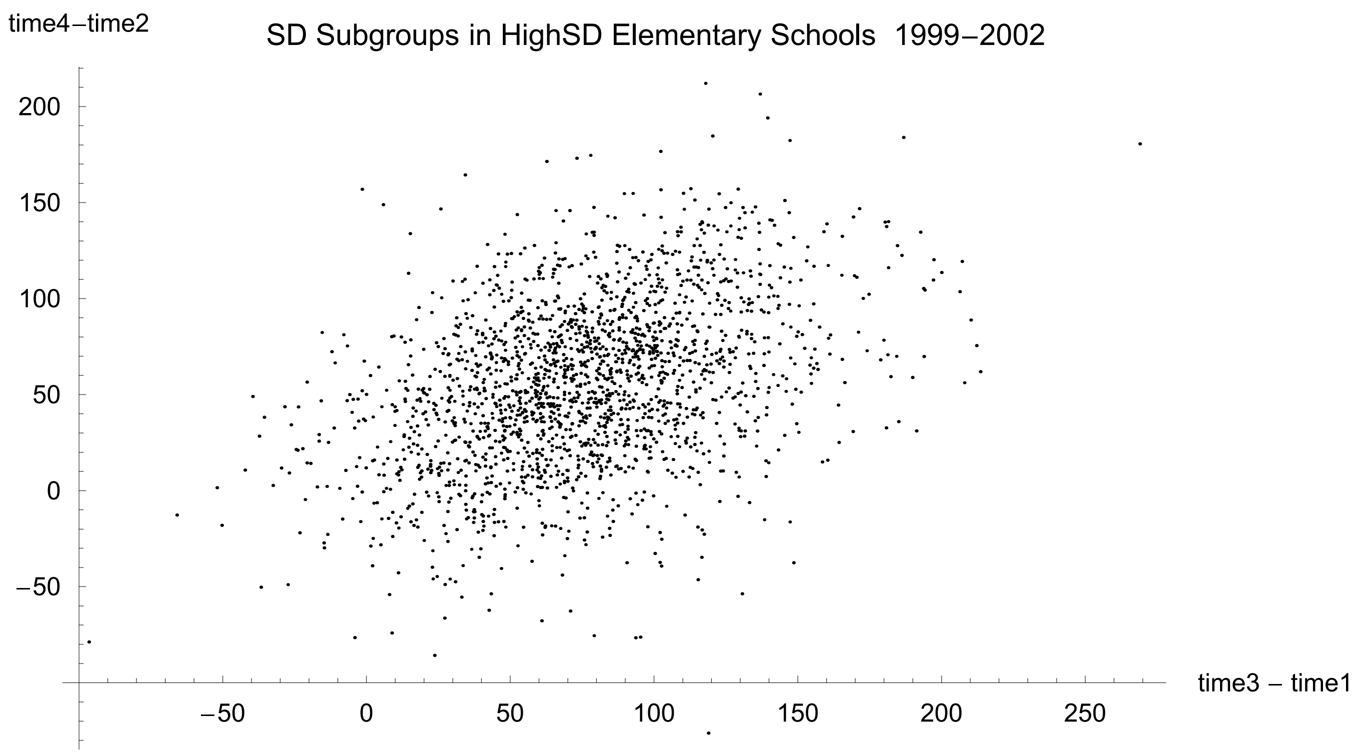
	Yearly School Scores					Year-to-Year Improvement		
	1999	2000	2001	2002		'99-00	'00-01	'01-02
1999	1.	0.916	0.865	0.806	'99-00	1.	-0.192	-0.037
2000	0.916	1.	0.919	0.854	'00-01	-0.192	1.	-0.078
2001	0.865	0.919	1.	0.922	'01-02	-0.037	-0.078	1.
2002	0.806	0.854	0.922	1.				
Mean	514.8	560.45	587.95	613.42		45.65	27.50	25.47
St.Dev	82.88	82.13	79.23	75.95		33.76	32.63	30.79

SD Subgroups in High SD Elementary Schools (n=2045)

	Yearly School Scores					Year-to-Year Improvement		
	1999	2000	2001	2002		'99-00	'00-01	'01-02
1999	1.	0.905	0.849	0.781	'99-00	1.	-0.205	-0.043
2000	0.905	1.	0.91	0.837	'00-01	-0.205	1.	-0.068
2001	0.849	0.91	1.	0.915	'01-02	-0.043	-0.068	1.
2002	0.781	0.837	0.915	1.				
Mean	498.18	544.89	573.62	600.90		46.71	28.74	27.28
St.Dev	75.58	76.48	74.23	72.09		33.19	32.13	30.26


time4−time2
Elementary Schools 1999−2002
200
150
100
50
0
−50
0
100
200
time3 − time1


SD Subgroups, Elementary Schools 1999−2002
time4−time2
time3 − time1
200
150
100
50
0
−50
−50
0
50
100
150
200
250



time4−time2
SD Subgroups in HighSD Elementary Schools 1999−2002
200
150
100
50
0
−50
time3 − time1
−50
0
50
100
150
200
250

Appendix B
Algebraic Derivations for Section 3 Results

This Appendix provides some details on obtaining the results in Equations 3.3, 3.4, 3.5, 3.6. The process is brute-force substitution of the forms of the LH and KS indices into the moments resulting from the growth curves in Equations 3.1 and 3.2. Results are stated in a somewhat more general form for equally spaced observations, which for convenience were taken to be $t_1 = 1$, $t_2 = 2$, $t_3 = 3$, $t_4 = 4$ and can be slightly generalized to an interval Δ between observations at t_1, $t_1 + \Delta$, $t_1 + 2\Delta$, $t_1 + 3\Delta$.

The population values of r_{LH} and r_{KS} can be written in terms of the population moments (variance *Var*, covariance *Cov*):

LH correlation = (B.1)

$$\frac{[Cov(Y(t_3), Y(t_4)) + Cov(Y(t_1), Y(t_2)) - Cov(Y(t_1), Y(t_4)) - Cov(Y(t_3), Y(t_2))]}{\text{Sqrt}\{Var(D[t_1, t_3]) \cdot Var(D[t_2, t_4])\}}$$

and

KS correlation = (B.2)

$$\frac{[Cov(Y(t_3), Y(t_2)) + Cov(Y(t_1), Y(t_2)) - Cov(Y(t_1), Y(t_3)) - Cov(Y(t_2), Y(t_2))]}{\text{Sqrt}\{Var(D[t_1, t_2]) \cdot Var(D[t_2, t_3])\}}$$

Constant Rate of Change Growth Model

Rewrite the growth curve $\eta_s(t) = \eta_s(0) + \theta_s t$ as $\eta_s(t) = \eta_s(\tau) + \theta_s(t - \tau)$ where $\tau = -\sigma_{\eta(0)\theta} / \sigma^2_\theta$, and $\eta_s(\tau)$ and θ_s are uncorrelated (Rogosa & Willett, 1985). Then the population moments for substitution into (B.1) and (B.2) are:

$$Cov(Y(t_j), Y(t_k)) = \sigma^2_{\eta(\tau)} + \sigma^2_\theta(t_j - \tau)(t_k - \tau) + \sigma^2 \qquad \text{if } t_j = t_k ,$$

$$= \sigma^2_{\eta(\tau)} + \sigma^2_\theta(t_j - \tau)(t_k - \tau) \qquad \text{otherwise.}$$

$$Var(D[t_j, t_k]) = \sigma^2_\theta(t_j - t_k)^2 + 2\sigma^2$$

By substitution (B.1) becomes

$$\text{LH correlation} = (2\sigma_\theta^2\Delta^2)/(2\sigma_\theta^2\Delta^2 + \sigma^2) \qquad \text{(B.3)}$$

and (B.2) becomes

$$\text{KS correlation} = (\sigma_\theta^2\Delta^2 - \sigma^2)/(\sigma_\theta^2\Delta^2 + 2\sigma^2) \qquad \text{(B.4)}$$

$$= 1 - 3\sigma^2/(\sigma_\theta^2\Delta^2 + 2\sigma^2)$$

One useful sidenote is that in the limit $\sigma_\theta^2 \rightarrow 0$, the KS correlation goes to –.5 and thus the lower limit of p_{KS} is 0.

Further substitution using reliability of the difference score

$$Reliability(D[t_j, t_k]) = \sigma_\theta^2 / (\sigma_\theta^2 + 2\sigma^2/(t_j - t_k)^2),$$

produces the results in (3.3) and (3.4) for arbitrary Δ .

Proportional Deceleration in Change.
From the growth model defined in (3.2), obtain the moments

$Cov(Y(t_j), Y(t_k))$ =

$$\sigma^2_{\eta(\tau)} + \sigma_\theta^2\{fh[j,h]\Delta + (t_1 - \tau)\}\cdot\{fh[k,h]\Delta + (t_1 - \tau)\} + \sigma^2 \quad \text{if } j = k,$$

$$\sigma^2_{\eta(\tau)} + \sigma_\theta^2\{fh[j,h]\Delta + (t_1 - \tau)\}\cdot\{fh[k,h]\Delta + (t_1 - \tau)\} \quad \text{otherwise,}$$

where the function $fh[t,h] = \Sigma_{i=2,t}\, h^{i-2}$. Also

$$Var(D[t_j, t_k]) = \sigma_\theta^2\{fh[k,h]\Delta - fh[j,h]\Delta\}^2 + 2\sigma^2$$

Substitution of these moments into (B.1) and (B.2) will yield the results for the proportional deceleration model presented in Section 3.2.