seniority. *Public Opinion Quarterly, 38,* 69–80.
Werts, C. E., Breland, H. M., Grandy, J., & Rock, D. A. (1980). Using longitudinal data to estimate reliability in the presence of correlated measurement errors. *Educational and Psychological Measurement, 40,* 19–29.
Werts, C. E., Jöreskog, K. G., & Linn R. L. (1971). Comment on "The estimation of measurement error in panel data." *American Sociological Review, 36,* 110–113.
Werts, C. E., Jöreskog, K. G., & Linn, R. L. (1973a). Identification and estimation in path analysis with unmeasured variables. *American Journal of Sociology, 78,* 1469–1484.
Werts, C. E., Jöreskog, K. G., & Linn, R. L. (1973b). A multitrait-multimethod model for studying growth. *Educational and Psychological Measurement, 33,* 665–678.
Werts, C. E., & Linn, R. L. (1970). Path analysis: Psychological examples. *Psychological Bulletin, 74,* 194–212.
Werts, C. E., Linn, R. L. & Jöreskog, K. G. (1974). Quantifying unmeasured variables. In H. M. Blalock (Ed.), *Measurement in the social sciences.* Chicago: Aldine.
Werts, C. E., Pike, L. W., Rock, D. A., & Grandy, J. (1981). Applications of quasi-Markov simplex models across populations. *Educational and Psychological Measurement, 41,* 295–307.
Werts, C. E., Rock, D. A., Linn, R. L., & Jöreskog, K. G. (1977). Validating psychometric assumptions within and between several populations. *Educational and Psychological Measurement, 37,* 863–872.
Wheaton, B., Muthén, B., Alwin, D. F., & Summers, G. F. (1977). Assessing reliability and stability in panel models. In D. R. Heise (Ed.), *Sociological methodology 1977* (pp. 85–136). San Francisco: Jossey-Bass.
Wiley, D. E., & Wiley, J. A. (1970). The estimation of measurement error in panel data. *American Sociological Review, 35,* 112–117.
Wilson, K. L. (1981). On population comparisons using factor indexes or latent variables. *Social Science Research, 10,* 301–313.
Winship, C., & Mare, R. D. (1983). Structural equations and path analysis for discrete data. *American Journal of Sociology, 89,* 54–110.
Woelfel, J., & Haller, A. O. (1971). Significant others, the self-reflexive act and the attitude formation process. *American Sociological Review, 36,* 74–87.
Wold, H. A., & Jureen, L. (1953). *Demand analysis.* New York: John Wiley.
Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics, 5,* 161–215.
Wright, S. (1954). The interpretation of multivariate systems. In O. Kempthorne, T. A. Bancroft, J. W. Gowen, & J. L. Lush (Eds.), *Statistics and mathematics in biology* (pp. 11–33). Ames, IA: Iowa State College Press.
Wright, S. (1960). Path coefficients and path regressions: Alternative or complementary concepts? *Biometrics, 16,* 189–202.
Zeller, R. A., & Warnecke, R. B. (1973). The utility of intervening constructs in experiments. *Sociological Methods and Research, 2,* 85–110.

5

Myths about Longitudinal Research*

David Rogosa

This chapter is concerned with methods for the analysis of longitudinal data. Longitudinal research in the behavioral and social sciences has been dominated, for the past 50 years or more, by a collection of damaging myths and misunderstandings. The development and application of useful methods for the analysis of longitudinal data have been impeded by these myths. In debunking these myths the chapter seeks to convey "right thinking" about longitudinal research; in particular, productive statistical analyses require the identification of sensible research questions, appropriate statistical models, and unambiguous quantities to be estimated. The heroes of this chapter are statistical models for collections of individual growth (learning) curves. The myths to be discussed are:

1. Two observations a longitudinal study make.
2. The difference score is intrinsically unreliable and unfair.
3. You can determine from the correlation matrix for the longitudinal data whether or not you are measuring the same thing over time.
4. The correlation between change and initial status is
   (a) negative
   (b) zero
   (c) positive
   (d) all of the above

*This chapter is a revised version of a coloquium of the same title presented at National Institutes of Health, Stanford University, University of California–Berkeley, Center for Advanced Studies in the Behavioral Sciences, and Vanderbilt University. Preparation of this chapter has been supported by a Seed Grant from the Spencer Foundation. I would like to thank Ghassan Ghandour, John B. Willett, and Gary Williamson for computational assistance in preparing the examples.

5. You can't avoid regression toward the mean.
6. Residual change cures what ails the difference score.
7. Analyses of covariance matrices inform about change.
8. Stability coefficients estimate
   (a) the consistency over time of an individual
   (b) the consistency over time of an average individual
   (c) the consistency over time of individual differences
   (d) none of the above
   (e) some of the above
9. Casual analyses support causal inferences about reciprocal effects.

The most prevalent type of longitudinal data in the behavioral and social sciences is longitudinal panel data. Longitudinal panel data consist of observations on many individual cases (persons) on relatively few (two or more) occasions (waves) of observation. An observation on a variable X at time t_i for individual p is written as X_{ip} where $i = 1, \ldots, T$, and $p = 1, \ldots, n$. (For statistical methods based on individual growth curves, observations need not be made at the same times for all individuals. But as this is necessary for the standard methods that predominate in the behavioral and social sciences, in my examples all individuals have the same values of t_i, which means everyone is measured at the same times.)

The X_{ip} are presumed to be composed of a true score $\xi_p(t_i)$ and an error of measurement ε_{ip} according to the classical test theory model: $X_{ip} = \xi_p(t_i) + \varepsilon_{ip}$. Many of the examples are in terms of the $\xi_p(t_i)$ and thus assume good measurement. The justification is that perfect measurement serves as a baseline for the examination of analysis methods. A statistical procedure that works poorly even with perfect measurement is clearly not attractive. Estimation of individual growth curves is not jeopardized by the presence of measurement error within reasonable bounds, but measurement errors cause more severe problems for methods based on the covariance matrix of the X_i (e.g., regression-based procedures).

The individual growth curves are functions of true score over time, $\xi_p(t)$. Research questions about growth, development, learning, and the like center on the systematic change in an attribute over time, and thus the individual growth curves are the natural foundation for modeling the longitudinal data. The growth curve models are kept relatively simple because the basic ideas and approaches remain valid for more complex growth models. The simplest and most widely used example will be straight-line growth, which specifies a constant rate of change denoted by θ. A second growth curve example is exponential growth to an asymptote.

The straight-line growth curve for individual p is written:

$$\xi_p(t) = \xi_p(0) + \theta_p t. \tag{5.1}$$

A collection of straight-line growth curves is shown in Figure 5-1; the individual growth curves have different values of rate of change θ_p and level $\xi_p(0)$. The value of the growth curve at a discrete time t_i yields the $\xi_p(t_i)$, and the X_{ip} are formed by the addition of measurement error. (In particular, for the many examples based on the collection of growth curves in Figure 5-1, the numerical values are obtained for a population of growth curves illustrated by the 15 growth curves in Figure 5-1, not for a sample or population of size 15.)

In some variables, such as attitudinal measures, the volatility over time may be far more important in the data than a systematic trend. The myth about stability over occasions will address this, using measures of consistency over time based on growth curve models.

The discussion of each myth is based on simple numerical examples, using either the $\xi_p(t_i)$ or X_{ip}. Although these examples are constructed to illustrate a particular message, each message is supported by technical results from my papers on statistical methods for the analysis of longitudinal data. This chapter is intended to serve as a less formal, and more accessible, exposition of the key ideas in those publications. In fact, the exposition deliberately avoids the presentation of mathematical results; citations throughout the text and the "reference notes" section at the end of each myth locate the relevant technical presen-

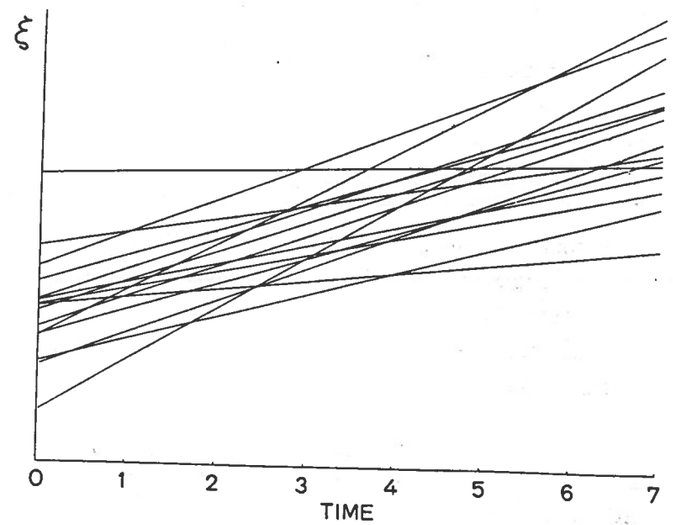

FIGURE 5-1. An illustrative collection of 15 straight-line growth curves in ξ (cf. Equation 5.1).

tations. A partial listing of the papers that serve as primary sources for this chapter are:

Rogosa, D. R. (1980). A critique of cross-lagged correlation. *Psychological Bulletin, 88,* 245-258.

Rogosa, D. R. (1985). Analysis of reciprocal effects. In T. Husen & N. Postlethwaite (Eds.), *International Encyclopedia of Education* (pp. 4221–4225). London: Pergamon Press.

Rogosa, D. R. (1987). Causal models do not support scientific conclusions. *Journal of Educational Statistics, 12,* 185–195.

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. (1983). 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. (1985a). Satisfying a simplex structure is simpler than it should be. *Journal of Educational Statistics, 10,* 99–107.

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

MYTH 1: TWO OBSERVATIONS A LONGITUDINAL STUDY MAKE

Strictly speaking, two repeated observations do constitute a longitudinal study. A more exact statement of the myth would be that two observations are presumed to be adequate for studying change. This misunderstanding is inspired by the dominance of pre-test, post-test longitudinal designs in the methodological and empirical work of the behavioral and social sciences. Two observations do provide some information about change over time, but this design has many critical limitations. In Rogosa et al. (1982, p. 744), I expressed this by the motto: "Two waves of data are better than one, but maybe not much better." Longitudinal designs with only two observations may address some research questions marginally well—but many others rather poorly.

Two Observations Permit Estimation of the Amount of Change but Not of the Individual Growth Curve

Consider two observations on true score ξ for a single individual plotted against time; that is, time is on the horizontal axis and true score is on the vertical axis. With just two observations over time, what can be learned about an individual?

Although it is statistically shaky, a growth curve can be fit to two points in time. A straight line passing through the two points is the most complex functional form that can be fit. Even then, the data contain no information on the adequacy of the straight-line functional form for growth or on the amount of scatter in the data. Furthermore, two points in time provide no basis for distinguishing among alternative growth curves; for example, a variety of exponential or logistic growth curves could pass perfectly through the two points. Even if the form of the growth curve were known (e.g., exponential), two observations are not sufficient to provide any estimates of the parameters of the growth curve. Although the investigation of the functional form of growth will often require far more than two points in time, two observations do allow estimation of the amount of change between t_1 and t_2. These remarks are obvious, and this discussion would be of little import if it were not for the preponderance of two-wave panel designs in methodological discussions and empirical studies of change and development.

The formulation of Coleman (1968) founded an alternative tradition for the study of change, mainly among sociologists. In this formulation the parameters of the growth function do not differ over individuals. This tradition assumes that "the process is identical for all persons" (p. 437) and allows the estimation of complex growth curves (e.g., exponential, logistic) where "the data may be two waves of a panel with two observations on many individuals or many observations on the same individual" (p. 432). Additional examples of this tradition are Nielson and Rosenfeld (1981), Salemi and Tauchen (1982), and Tuma and Hannan (1984, chap.11). In order to estimate complex growth curves from only two observations on each individual, observations from many individuals must be "combined" into a single growth curve. Individual differences in growth preclude the validity of this approach unless some exogenous individual characteristics can be used to completely account for the individual differences. That is, violations of the assumption that the parameters of the growth function are the same for all individuals can be extremely consequential.

The Amount of Change Will Often Be Deceptive

The amount of change over a specified time interval is a natural quantity to estimate from longitudinal data. Define $\Delta_p(t, t + c) = \xi_p(t + c) - \xi_p(t)$ as the amount of true change for individual p over the time interval starting at time t and extending c units. For straight-line growth, $\Delta_p(t, t + c) = \theta_p c$. The amount of change between times t and time $t + c$ depends on t for growth curves having a nonconstant rate of change (i.e., growth curves other than straight-line) and will often be a complex function of t and c. Thus, in a two-wave study, choices of time 1–time 2 measurements are likely to be extremely consequential. In particular, the amount of change may be especially deceptive in comparing growth

among individuals because observations over alternative time intervals may yield contradictory information. Below is an example showing that the amount of change is no guide to individual differences in growth.

Consider a collection of six individual growth curves, for individuals labeled A, B, C, D, E, F. Each growth curve has the form of exponential growth toward a ceiling or asymptote governed by the equation

$$\xi_p(t_i) = \lambda_p - (\lambda_p - \xi_p(0))e^{-\gamma_p t_i}. \qquad (5.2)$$

Table 5-1 gives the parameter values for these six growth curves. The individuals differ on the asymptote λ, on the starting level $\xi(0)$, and on the curvature parameter γ. These growth curves also produce individual differences in the amount of change. Table 5-2 presents the amount of change $\Delta\ (t_I, t_I + 1)$ for individuals A, B, C, D, E, F for initial observation at time t_I and final observation at $t_I + 1$, with $t_I = 0, 4, 10$. For $t_I = 0$ the individual ranking on Δ is A, B, C, D, E, F (A improves the most in the interval [0, 1], B the next most, C the next, and so on), with the largest Δ nearly double the smallest. If instead $t_I = 10$, the ranking for the amount of change is reversed, with the largest Δ nearly three times the smallest. So two different studies might obtain exactly the opposite results for individual differences in change depending on the choice of initial time of measurement. Furthermore, for $t_I = 4$, the Δ values are nearly equal (smaller individual differences in change) with yet a different ranking of individuals.

The reversals of individual standing on the amount of change may be most consequential for studies of the correlation of change with an exogenous background variable. Such a correlation might be found to be big, positive for a study using $t_I = 0$; big, negative for $t_I = 10$; and about zero for $t_I = 4$, even if all three studies had perfect measurement.

The example above illustrates the danger of characterizing the growth of individuals by the amount of change over a specific time interval. Even with perfect measurement, the pre-post longitudinal design provides meager information. Two-wave designs permit at best the study of individual differences in Δ or, equivalently, in some sort of average rate of change. Consequently, designs with only two observations are usually inadequate for the study of individual growth and individual differences in growth.

TABLE 5-1. Parameter Values for the Six Exponential Growth Curves

Individual	$\xi(0)$	λ	γ
A	50	80	.25
B	40	70	.22
C	30	60	.19
D	20	50	.16
E	10	40	.13
F	0	30	.10

TABLE 5-2. Amount of True Change $\Delta(t_I, t_I + 1)$ for Exponential Growth Example

Individual	$t_I = 0$	$t_I = 4$	$t_I = 10$
A	6.64	2.44	.54
B	6.32	2.62	.70
C	5.88	2.75	.88
D	5.37	2.81	1.07
E	4.63	2.75	1.26
F	3.81	2.55	1.40

Reference notes

The limitations of two-wave designs for the measurement of change are examined in Rogosa et al. (1982) and Rogosa and Willett (1985b). Mathematical results corresponding to the example in Table 5-1 are given in Section 1.4 of Rogosa and Willett (1985b). The advantages of multiwave data for the estimation of individual change are enumerated in Rogosa et al. (1982, pp. 741–743).

MYTH 2: THE DIFFERENCE SCORE IS INTRINSICALLY UNRELIABLE AND UNFAIR

An impressive amount of psychometric literature over the last 50 years has sought to demonstrate deficiencies in the difference score. With only two observations, the difference score, $D = X_2 - X_1$, is a natural estimate of the amount of true change, $\Delta(t_1, t_2)$, regardless of the form of the growth curve. For a straight-line growth curve model the difference score estimates the (constant) rate of change times the time interval. In general, the difference score divided by the time elapsed estimates an average rate of change over the time interval.

Unreliability of the Difference Score

The traditional tabulation of the reliability of the difference score is shown by Table 5.3, which also appears in Linn and Slinde (1977) and in various forms in

TABLE 5-3. Traditional Tabulation of the Reliability of the Difference Score

	$\rho(X)$		
$\rho_{X_1X_2}$	.7	.8	.9
.5	.40	.60	.80
.6	.25	.50	.75
.7	.00	.33	.67
.8	—	.00	.50
.9	—	—	.00

many other publications. The pre-test-, post-test correlation of observed scores and the reliability of observed scores look reasonable, and for most combinations the difference score has little reliability. This type of numerical demonstration supports the assertions by Lord that "differences between scores tend to be much more unreliable than the scores themselves" (1956, p. 429) and that "the difference between two fallible measures is frequently much more fallible than either" (1963, p. 32).

The untold story is the limited and constrained nature of this table. The table employs the constraints of equal reliabilities $\rho(X_1) = \rho(X_2) = \rho(X)$ and equal variances $\sigma^2_{X_1} = \sigma^2_{X_2} = \sigma^2_X$ for the fallible observed scores, X_1 and X_2. Also, these constraints imply equal true-score variances at times 1 and 2 and also a negative value of $\rho_{\xi_1\Delta}$, the correlation between true change and true initial status.

The most prominent feature of Table 5-3 is that the time 1–time 2 true-score correlation $\rho_{\xi_1\xi_2}$ is very large in almost all regions; this can be seen from the standard disattenuation formula $\rho_{\xi_1\xi_2} = \rho_{X_1X_2}/\rho(X)$. In particular, $\rho_{\xi_1\xi_2}$ is 1.0 along the diagonal of zero reliability for the difference score. What are the implications for individual growth of the table's restriction to this small portion of the parameter space? A collection of growth curves that exhibit high time 1–time 2 correlation and equal variances at times 1 and 2 will have all the growth curves nearly parallel. Thus all individuals are growing at nearly the same rate which translates into almost no individual differences in true change. (Figure 1 of Rogosa et al., 1982, shows such a collection of straight-line growth curves with time 1–time 2 correlation about .95.) If there are no individual differences in true change, the difference score cannot be expected to detect them. So after building into the traditional tabulations the constraints that there be almost no individual differences in growth, the low reliability of the difference score should be no surprise.

If, instead, a moderate correlation, $\rho_{\xi_1\xi_2} = .4$, is used in conjunction with the other constraints in Table 5-3, the difference score appears much stronger. The quantity $\rho(D) / \rho(X)$ has values. 83, .88, and .94 for the $\rho(X)$ values .7, .8, and .9, respectively. Thus, even with the other constraints the difference score is nearly as reliable as the measure X. A moderate time 1–time 2 correlation corresponds to numerous crossings of the growth curves and considerable individual differences in change. (Rogosa et al. 1982, Figure 2, shows a time 1–time 2 correlation of about .5 for a collection of straight-line growth curves.)

Table 5-4 presents a slightly different tabulation of the reliability of the difference score in terms of the time 1–time 2 true-score correlation and the reliability of X. The reliability of X_2 is set to .9, and the reliability of X_1 is varied. (Setting $\rho(X_2) > \rho(X_1)$ maintains approximately equal error variances at times 1 and 2.) The correlation between true change and initial status is set to zero, which is a useful benchmark case, also known as the Overlap Hypothesis. For these parameter values, the difference score does extremely well; for a moderate true-score correlation the difference score is *more* reliable than the average reliability of the measures. Even for a high correlation, the difference score does rather well compared to reliability of X, and in absolute terms, $\rho(D)$ is also substantial. In sum, when there are individual differences in change, the difference score has decent reliability.

The message that debunks this myth is that the difference score is reliable when individual differences in true change exist. After all, the reliability of the difference score is the variance of true change divided by the sum of the variance of true change and the variance of the difference of the errors. For parameter configurations that require all individuals to grow at about the same rate, the low reliability of the difference score properly reveals that you can't detect individual differences that ain't there.

Unfairness of the Difference Score

The belief that the difference score is somehow not a "fair" measure of change is reflected in the statements that difference scores "give an advantage to persons with certain values of the pretest score" (Linn & Slinde, 1977, p. 125) and "the correlation between change and initial status made it inappropriate to use change

TABLE 5-4. Values of $\rho(D)/\bar{\rho}(X)$ when $\rho_{\xi_1\Delta} = 0$ and $\rho(X_2) = .90$

	$\rho(X_1)$		
$\rho_{\xi_1\xi_2}$	.6	.7	.8
.4	1.06	1.03	1.00
.6	.86	.88	.90
.8	.53	.60	.67

[difference] scores to evaluate individuals with different initial scores" (O'Connor, 1972, p. 78). The difference score is an unbiased estimate of true change. How can an unbiased estimate be inequitable? That is a question to which I have no answer. The confusion is bound up with misunderstandings about the correlation between change and initial status and with misguided motivations for the use of residual change measures. These will be untangled in subsequent myths.

Reference notes

A presentation of the reliability of the difference score in terms of individual differences in growth is given in Rogosa et al. (1982, pp. 731–734). The nontechnical exposition of Rogosa and Willett (1983b) provides numerical examples demonstrating the reliability of the difference score when individual differences in growth exist. Statistical properties of D_p for estimating Δ_p are described in Rogosa et al. (1982); in particular, the construction and properties of "improved difference score" (Kelley-type, Lord-McNemar, and empirical Bayes) estimates, which use information from all n individuals in the estimation of Δ_p are examined in detail in Rogosa et al. (1982, pp. 735–738, 742–743, and the Appendix).

MYTH 3: YOU CAN DETERMINE FROM THE CORRELATION MATRIX FOR THE LONGITUDINAL DATA WHETHER OR NOT YOU ARE MEASURING THE SAME THING OVER TIME

A typical statement of the third myth is that with low correlations over time "it is questionable whether one is measuring the same thing on both occasions, and consequently the notion of change becomes questionable" (Bond, 1979). A very serious question in studies of development (whether it be in early child development or later in the aging process) is whether measures "change out from under you" in the sense of measuring something different on different occasions of observation. The important issue is whether asking about quantitative change in the measures over time is meaningful. The assumption that the psychological variable or dimension being studied retains the same meaning over the occasions of observation is a logical prerequisite for the measurement of quantitative change. This view is reflected by Lord (1958), who discussed an instructional setting in which "the test no longer measures the same thing when given after instruction as it did before instruction. If this is asserted, then the pretest and posttest are measuring different dimensions and no amount of statistical manipulation will produce a measure of gain or of growth" (p. 440). Similarly, Bereiter (1963) wrote: "Once it is allowed that the pretest and posttest measure

different things it becomes embarrassing to talk about change. There seems no longer any way to answer the question, change on what?" (p. 11). (See also Cronbach & Furby, 1970, p. 76; Linn & Slinde, 1977, p. 24; Lord, 1963, p. 21).

In many situations these concerns may preclude the study of quantitative change. Nonetheless, valid and answerable questions about change should be pursued. Thus, the myth addresses a very important consideration; the misunderstanding is in thinking that this issue can be resolved by the between-wave correlation matrix. The truth is that much more and very different information may be required to resolve this issue.

Consider the picture of a collection of straight-line growth curves in Figure 5-1. Table 5-5 presents the corresponding correlation matrix, with entries of the correlation between ξ_i and $\xi_{i'}$ for $t_i, t_{i'} = 0, 1, \ldots, 8$. Now, between times 5 and 7 the correlation between true scores is very big, .94; even with some measurement error there would be a healthy correlation. Should we "conclude" that the same thing is being measured over this time interval? If, instead, the interval is from time 1 to 5, the correlation is .385. Should this correlation be taken to indicate that different things are being measured at times 1 and 5? Furthermore, for the interval with end points at times 1 and 7 (the concatenation of the two time intervals above) the correlation is .056. Are unrelated quantities being measured at times 1 and 7? According to the myth, the above three questions receive affirmative answers. Furthermore, the correlation between times 0 and 8 is −.24; should this correlation be taken to indicate that *opposite* attributes are being measured at times 0 and 8?

The correlations in Table 5-5 correspond to the collection of straight-line growth curves in Figure 5-1. As each individual has a constant rate of change on the attribute ξ, it is hard to imagine a configuration of individual growth that shows less discontinuity. Clearly, a way of thinking that indicates that different things are measured by ξ_i and $\xi_{i'}$ has deep flaws. In the same vein, large correlations cannot "prove" that the same thing is being measured at both ends of

TABLE 5-5. True-Score Correlation Matrix for Straight-line Growth Example

	0	1	2	3	4	5	6	7	8
0	1								
1	.981	1							
2	.894	.965	1						
3	.707	.832	.949	1					
4	.448	.614	.800	.949	1				
5	.197	.385	.614	.832	.965	1			
6	.001	.197	.447	.707	.894	.981	1		
7	−.140	.056	.317	.600	.822	.943	.990	1	
8	−.241	−.047	.218	.515	.759	.904	.970	.995	1

the observation interval, only that the ordering of individuals in the initial measure is similar to the ordering of individuals in the final measure. Whether or not the same thing is being measured over time simply cannot be answered from the correlation matrix on a couple of occasions of measurement, and it is dangerous to do so. Even plotting the individual growth curves cannot completely resolve this question, although large discontinuities in individual growth would be cause for concern.

A sidenote message to this myth is that large individual differences in growth lower the between-wave correlations. Myth 3 serves to discourage the study of change for variables that have sizable individual differences in growth on the grounds that these variables do not retain the same meaning over time . Thus, variables that are chosen for study have high time 1–time 2 correlations, which often results in low σ^2_Δ (i.e., not much individual differences in change). In reference to Myth 2, if there are little individual differences in change, what will the difference score show? Low reliability.

Reference notes

The results of Rogosa and Willett (1985b) can be used to obtain the between-wave covariance and correlation functions for different forms of individual growth; the results for straight-line growth were used in constructing the example in Table 5-5. Rogosa et al. (1982, pp. 731–733) discuss the consequences for the reliability of the difference score of limiting studies of change to variables with high between-wave correlations (stability).

MYTH 4: THE CORRELATION BETWEEN CHANGE AND INITIAL STATUS IS

(a) negative
(b) zero
(c) positive
(d) all of the above

Myth 4 is a multiple choice myth whose distractors have long-standing substantive interpretations. A negative correlation between change and initial status is best known as the Law of Initial Values (Lacey & Lacey, 1962; Wilder, 1957). The negative correlation is also bound up with Regression Toward the Mean, as will be seen in Myth 5. A zero correlation between change in initial status is known as the Overlap Hypothesis, which dates back to Anderson (1939) and was prominent in Bloom (1964). One interpretation of the Overlap Hypothesis is that growth occurs via independent increments (similar to the formulation of simplex models in Humphreys, 1960). A positive correlation between change

and initial status corresponds to "fanspread" where variances increase over time. The positive correlation can be described as "them that has, gets."

The correct answer is (d), " all of the above, " because the correlation between change and initial status depends crucially on the choice of t_I, the time at which initial status is measured. For straight-line growth, the correlation between change and initial status is monotonically increasing, having a lower asymptote of -1.0 for $t_I = -\infty$, passing through 0 for a single t_I and increasing to an upper asymptote of 1.0 for $t_I = \infty$. For almost any collection of growth curves, a very different correlation between true change and true initial status will be obtained, depending on whether the time of initial status is chosen to be later, earlier, or in between—a likely reason that studies of academic growth obtain disparate estimates of the correlation between true change and true initial status.

One sidenote to the myth is that with fallible scores, the correlation between observed change and observed initial status is a poor estimate of the correlation between true change and true initial status. The estimate is negatively biased in addition to the attenuation (see, e.g., Rogosa et. al., 1982, Eq. 11). Thus, because of the poor properties of this estimate, negative correlations between observed change and observed initial status are often obtained when the true-score correlation is zero or positive. The myth is stated and discussed in terms of true scores because these are of primary substantive interest; although of less interest, a similar dependence on time of initial status also holds for the observed score correlation.

Table 5-6 gives values of the correlation between the amount of true change $\Delta(t_I, t_I + c)$ and true initial status $\xi(t_I)$ for $t_I = 0, \ldots, 7$, using the collection of straight-line growth curves for true scores shown in Figure 5-1. The correlation does not depend on c. For each choice of t_I a different value for the correlation

TABLE 5-6. Correlation between Change and Initial Status for Straight-line Growth Example in Figure 5-1

		$\rho_{X_i(X_{i+c}-X_i)}$	
t_I	$\rho_{\xi(t_I)\Delta}$	$c=1$	$c=3$
0	−.71	−.50	−.69
1	−.55	−.48	−.59
2	−.32	−.44	−.47
3	0	−.36	−.29
4	.32	−.25	.00
5	.55	−.12	.17
6	.71	−.02	.30
7	.80	.02	.42

between change and initial status will be obtained. In this example, if initial status is chosen to be time 1, the correlation is big and negative. If initial status is time 3, the correlation is zero. And if initial status is time 5, the correlation is positive. Time 3 is the only time of initial status that would satisfy Anderson's Overlap Hypothesis. The Law of Initial Values would be satisfied for any $t_I < 3$.

Table 5-6 also gives values of the correlation between observed initial status X_i and observed change $X_{i+c} - X_i$ for $c = 1, 3$. The X_i are based on the $\xi(t_i)$ for this example, with the addition of measurement error (having equal error variance over the t_i), producing reliabilities of the X_i between .74 and .87. The difference between the $c = 1$ and $c = 3$ values is attributable to the larger reliability of the difference score for $c = 3$; except for $t_I = 2$, the $c = 3$ observed score correlation is closer to the true-score correlation. The difference between the observed-score and true-score correlations is somewhat complex. For $t_I > 2$ the observed-score correlation is always less than the true-score correlation, especially for non-negative values of the true-score correlation ($t_I \geqslant 3$). For large negative values of the true-score correlation, the attenuation and negative bias in the observed-score correlation may offset each other.

Table 5-7 repeats the example for a different type of growth curve: exponential growth to an asymptote λ instead of straight-line growth. This collection of growth curves is illustrated in Figure 5-2. The exponential growth curves have the form of equation (5.2) with $\gamma_p = \gamma$. This collection of growth curves was constructed to have a between-wave correlation structure similar to that for the straight-line growth example (with a translation of the time scale by 3 units). The correlation between change and initial status is monotone increasing in t_I, and like straight-line growth the correlation strongly depends on the choice of t_I. Unlike straight-line growth the correlation is no longer symmetric about the zero value, which for this example is $t_I = 6$.

Reference notes

Mathematical results for the form of $\rho_{\xi(t)\Delta}$ are obtained in Rogosa and Willett (1985b) for straight-line growth, exponential growth, and the simplex model (Eqs. 9, 16, and 13, respectively). In terms of the notation and parameters of

TABLE 5-7. Correlation between Change and Initial Status for Exponential Growth Example in Figure 5-2

t_I	3	4	5	6	7	8	9	10
$\rho_{\xi(t_I)\Delta}$	−.84	−.67	−.37	0	.31	.50	.53	.68

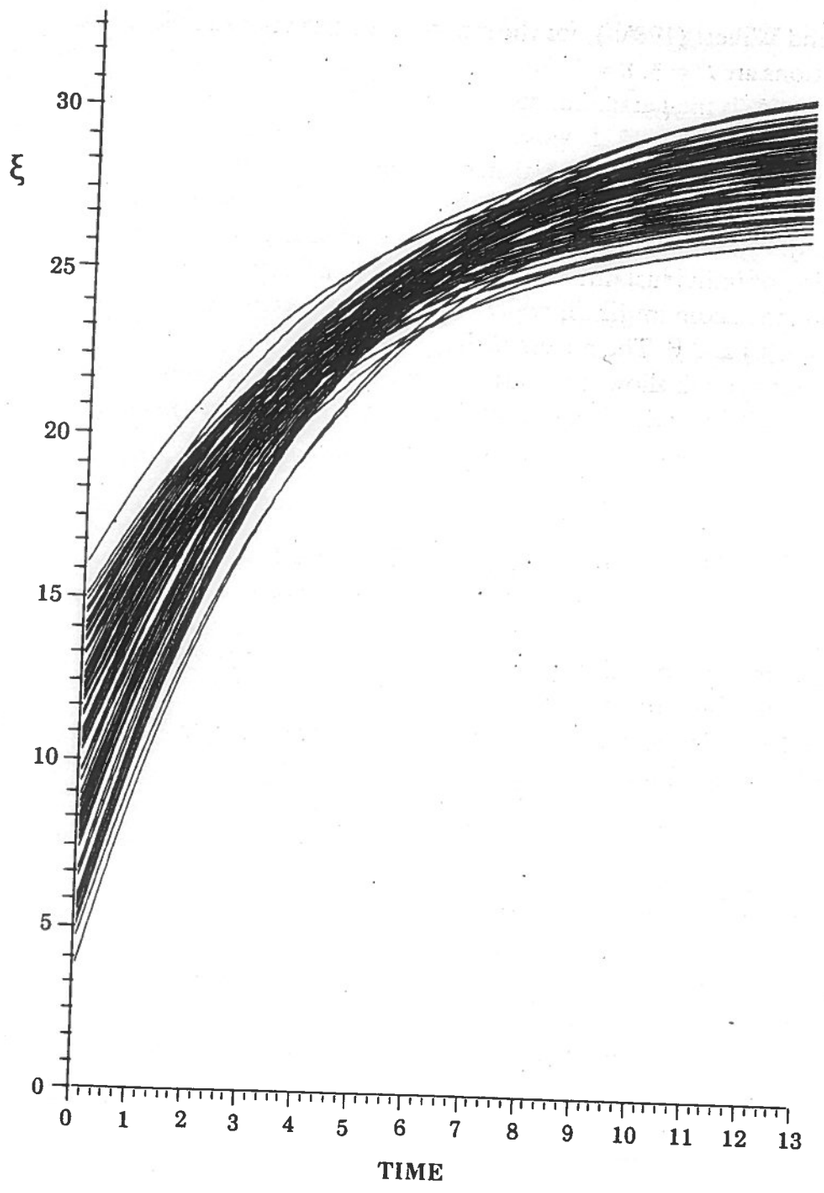

FIGURE 5-2. An illustrative collection of exponential growth curves in ξ following Equation 5.2, with $\gamma_p = \gamma$.

Rogosa and Willett (1985b), for the straight-line growth example the parameter specifications are $t^{\circ} = 3$, $\kappa = 3$. For the exponential growth example in Figure 5-2 and Table 5-7, the parameter specifications are $t^{\circ} = 6$; $\gamma_p = \gamma = .23$; $\mu_\lambda = 30$; $\sigma^2_\lambda = 1.4$, and $\sigma^2_{\xi(6)} = .437$. Rogosa and Willett (1985b) also obtain the form of the regression of change on initial status. Rogosa et al. (1982, pp. 734–735) examine the bias of the correlation between observed change and observed initial status. Blomqvist (1977, Eq. 3.2) using straight-line growth and a linear representation of individual differences in growth as a function of initial status (Eq. 3.1), obtains maximum likelihood estimates of the elements of the covariance matrix of $\xi(0)$ and θ. The results of Rogosa and Willett (1985b, Section 2) for straight-line growth allow the construction of maximum-likelihood estimates of the correlation between change and initial status or the regression of change on initial status for t_I other than $t_I = 0$.

MYTH 5: YOU CAN'T AVOID REGRESSION TOWARD THE MEAN

Typical statements of this myth are Furby (1973, p. 172), "Regression toward the mean is ubiquitous in developmental psychological research" and Lord (1963, p. 24), "The regression effect is one of the two main reasons why studies of growth may become confusing or confused." What is nearly ubiquitous about regression toward the mean is the absence of explicit, defensible definitions of the phenomenon. That is, regression toward the mean is often talked about but rarely explicitly stated. Intuitively, regression toward the mean says that on the average you are going to be closer to the mean at time 2 than you were at time 1. The few formal statements of regression toward the mean in the literature define it in standard deviation units: for example, Furby (1973, p. 174) and Nesselroade, Stigler, and Baltes (1980, p. 623). Thus, in the population, regression toward the mean for true scores at times t_1 and t_2 is said to occur when

$$\frac{E[\xi(t_2) \mid \xi(t_1) = C] - \mu_{\xi(t_2)}}{\sigma_{\xi(t_2)}} < \frac{C - \mu_{\xi(t_1)}}{\sigma_{\xi(t_1)}}. \tag{5.3}$$

Because this inequality is satisfied whenever $\rho_{\xi(t_1)\xi(t_2)} < 1$, regression toward the mean is thought to be unavoidable. The formulation in Eq. (5.3) is best thought of as a harmless mathematical tautology and one which provides little insight for the study of change.

A more realistic definition of regression toward the mean uses the actual metric of ξ to express closer to the mean at time 2 than at time 1. The alternative formulation of regression toward the mean is

$$E[\xi(t_2) \mid \xi(t_1) = C] - \mu_{\xi(t_2)} < C - \mu_{\xi(t_1)}. \tag{5.4}$$

Only if $\sigma_{\xi(t_1)}$ and $\sigma_{\xi(t_2)}$ are constrained to be equal, as is done in Lord (1963, p. 21) and in Furby (1973, p. 173), is Eq. (5.4) equivalent to Eq. (5.3). Most important, Eq. (5.4) is satisfied only when $\rho_{\xi(t_1)\Delta} < 0$ (where $\Delta = \Delta(t_1, t_2)$). So, for the formulation in Eq. (5.4), regression toward the mean is not ubiquitous; regression toward the mean pertains only when the correlation between change and initial status is negative. Myth 4 discusses conditions for this to hold.

The formulation in (5.4) corresponds to the original notion of Galton (1886) much more closely than does Eq. (5.3). Specifically, Galton would indicate no regression toward the mean if the time 2 on time 1 regression coefficient $\beta_{\xi(t_2)\xi(t_1)}$ is greater than or equal to one. This is equivalent to $\rho_{\xi(t_1)\Delta} \geqslant 0$, for which the inequality in Eq. (5.4) is not satisfied. By expressing the severity of the regression effect as the ratio

$$\frac{E[\xi(t_2) \mid \xi(t_1) = C] - \mu_{\xi(t_2)}}{C - \mu_{\xi(t_1)}} = \beta_{\xi(t_2)\xi(t_1)} \tag{5.5}$$

the correspondence of Eq. (5.4) to Galton's formulation is seen.

The standard textbook representation of regression toward the mean employs a picture of the time 2 on time 1 plot with an ellipse representing the bivariate data (e.g., Nesselroade et al., 1980, Figure 1). For a choice of a time 1 value C, the time 2 on time 1 regression line gives the expected value at time 2. The peculiar aspect of this standard picture is that it is always drawn to show equal variances at time 1 and time 2, making Eq. (5.4) equivalent to Eq. (5.3). An alteration of the standard picture in Figure 5-3 allows variance to increase over time. Figure 5-3 shows that the expected value is farther away from the mean at time 2 than at time 1. Thus, regression toward the mean does not hold. Another example is seen in the collection of straight-line growth curves in Rogosa et al. (1982, Figure 3).

Reference notes

Healy and Goldstein (1978), Rogosa et al. (1982, p. 735), and Rogosa and Willett (1985b, Section 2.5) provide similar discussions of regression toward the mean with reference to collections of individual growth curves. Rogosa and Willett (1985b) define explicitly the conditions for Eq. (5.4) to hold. Nesselroade et al. (1980) examine the structure of regression toward the mean for multioccasion data. In the Nesselroade et al. paper, regression toward the mean is analyzed in terms of correlation structures. Consequently, some regression toward the mean will always pertain because of the standardization involved in the correlation

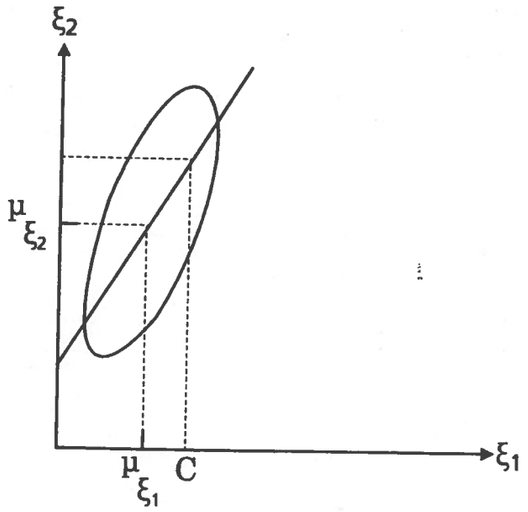

FIGURE 5-3. An illustration based on the standard depiction of regression toward the mean, in which regression toward the mean does not hold.

matrix. Nesselroade et al. use the term "egression from the mean" to describe a regression toward the mean that is less severe between t_1 and t_3 than between t_1 and t_2 (even though there is regression toward the mean between t_1 and t_3). Perhaps a better use of this term would be egression from the mean as the opposite of regression toward the mean, which would exist over the time interval $[t_1, t_2]$ if and only if the correlation between $\xi(t_1)$ and $\Delta(t_1, t_2)$ is positive.

MYTH 6: RESIDUAL CHANGE CURES WHAT AILS THE DIFFERENCE SCORE

What ails the difference score, according to the psychometric literature, is low reliability and negative correlation with initial status. The discussion of previous myths has shown such deficiencies of the difference score to be more illusory than real. Nonetheless, these concerns have motivated the use of residual change scores. In terms of true scores, residual change is a deviation of true outcome at time 2 from the regression prediction using time 1 information; using $\xi(t_1)$ as the time 1 information yields a true residual change of the form $\xi_p(t_2) - \mu_{\xi(t_2)} - \beta_{\xi_2\xi_1}[\xi_p(t_1) - \mu_{\xi(t_1)}]$. With fallible measures, the usual sample estimate of residual change is the residual from the observed-score time 2 on time 1 regression which is denoted by $\hat{R}$.

A look at the properties of $\hat{R}$ is not pretty. Bias? Yes; $\hat{R}$ may be a badly biased estimate of true residual change. Precision? Not much; the sampling variability is

rather large because $\hat{R}$ contains uncertainty both from measurement error and from finite sample size in the regression adjustment. Reliability? At best, not much better than the reliability of the difference score. Various modifications of $\hat{R}$, mainly intended to ameliorate the effects of measurement error on the regression adjustment, do little to mend its severe deficiencies.

The demonstrations in the literature of superior reliability for residual change use time 1–time 2 true-score correlations near one and equal true-score and observed-score variances across time (Linn & Slinde, 1977, Table 2). Then, the reliability of the difference score is near zero, yet the reliability of residual change (even assuming an infinite sample size for making the regression adjustment) is only negligibly better. With $\rho_{\xi_1\xi_2} = 1$, the reliability for residual change is .09 for $\rho(X) = .8$ and .05 for $\rho(X) = .9$. Outside the extreme limitations of that comparison, not even the slight advantage for the residual change score holds up. Table 5-8 presents the reliability of the residual change score for different values of t_1 and $t_2 - t_1$ using the same X_i configuration as described for Table 5-6. The reliability of residual change increases with $t_2 - t_1$ and depends strongly on the choice of t_1. Compare these entries with the reliability of the difference score of .133 for $t_2 - t_1 = 1$ and .58 for $t_2 - t_1 = 3$; this reliability does not depend on t_1 as σ^2_Δ does not change. Thus, for many t_1 values the difference score is *more* reliable than residual change. The values given in Table 5-8 are obtained from of Rogosa et al. (1982, Eq. 20), which is the squared correlation between the true residual change and $\hat{R}$; this formula inflates the actual reliability of $\hat{R}$ as all available formulas for the reliability of residual change assume an infinite sample size for the regression adjustment (i.e., $\beta_{X_2X_1}$ known).

The logical problems of the residual-change approach dwarf its technical shortcomings. Instead of addressing the relatively simple question—how much did individual p change on the attribute ξ?—residual change attempts to assess how much individual p would have changed on ξ if all individuals had started out "equal." The obvious questions is, equal on what—true initial status, observed initial status, true initial status and other background characteristics? The cor-

TABLE 5-8. Reliability of Residual Change for Straight-line Growth Example

	$t_2 - t_1$	
t_1	1	3
0	.013	.557
2	.145	.683
4	.213	.487
6	.105	.347

rect answer is unknown, and it depends on the correct specification for the prediction of change. The difficulties with residual change are analogous to those with statistical comparisons of treatment effects in nonequivalent groups. Residual change is one example of attempts to statistically adjust for preexisting differences, which the literature on the analysis of quasi-experiments has shown to be doomed to failure.

A major use of residual change measures is to detect correlates of change. Questions about correlates of change are of the type, "What kind of people are improving or gaining the most"? When the potential correlate is a variable defining membership in an experimental group, the question is whether people getting care or treatment are improving more than people who are not. Questions about correlates of change can be expressed in terms of systematic individual differences in growth. Individual differences in growth exist when parameters of individual growth curves (e.g., the θ_p) differ across individuals, (i.e., some people grow faster than others). Individual differences in growth are systematic if individual differences in a growth parameter can be linked with one or more exogenous characteristics.

A common analysis consists of correlating the observed residual change with an exogenous, individual characteristic denoted by W. Tucker, Damarin, and Messick (1966) formed estimates of the correlation between the exogenous variable and the true residual-change score. Lord (1963) presented a slightly different measure, which is equivalent to a partial correlation instead of the part correlation in Tucker *et al.*

The failure of these measures to assess systematic individual differences in growth is demonstrated by an example using the collection of straight-line growth curves illustrated in Figure 5-1. The example includes two cases. Case 1 is no systematic individual differences in growth; that is, the correlation $\rho_{W\theta}$ between the exogenous variable and rate of change is zero. Case 2 is large systematic individual differences in growth; that is, $\rho_{W\theta} = .7$. The example assumes perfect measurement of ξ and W. Table 5-9 shows values of the correlation from Tucker et al. (1966) $\rho_{[\xi(t_2)\cdot\xi(t_1)]W}$ for $\rho_{W\theta} = 0, .7$. Table 5-10 repeats the display for the partial correlation from Lord (1963) $\rho_{\xi(t_2)W\cdot\xi(t_1)}$. When there are no systematic individual differences in growth, the correlations may be large positive or large negative depending on the choice of t_1. Even large systematic individual differences in growth may result in near zero or even negative values of these correlations. Thus, neither of these correlations can be counted on to assess correlates of change.

Residual change correlations, whether partial or part correlations, are based on an adjustment for the effects of initial status. And this adjustment naturally depends on the choice of time at which initial status is measured. Thus, the attempt to purge initial status from the measure of change fails. The fatal flaw of the residual change procedures is the attempt to assess correlates of change by ignoring individual growth. Questions about systematic individual differences in growth cannot be answered without reference to individual growth. Yet these time 1–time 2 correlation procedures valiantly attempt to do so.

TABLE 5-9. Values of $\rho_{[\xi(t_2)\cdot\xi(t_1)]W}$ for Straight-line Growth Example in Figure 5-1

t_1	$\rho_{W\theta} = 0$	$\rho_{W\theta} = .7$
0	.64	.92
1	.50	.92
2	.29	.85
3	0	.70
4	−.29	.47
5	−.50	.25
6	−.64	.07
7	−.73	−.06
8	−.78	−.15

Reference notes

Rogosa et al. (1982, pp. 738–741, p. 743, Appendix) enumerate the statistical, psychometric, and logical shortcomings of the residual-change score as a measure of individual change for both two-wave and multiwave longitudinal data. Rogosa and Willett (1985b, Section 3) obtain the mathematical forms for the Tucker et al. (1966) and Lord (1963) correlations and demonstrate the failure of these procedures for the assessment of correlates of change. The values in Tables 5-9 and 5-10 were obtained from Rogosa and Willet (1985b, Eqs. 23 & 24, re-

TABLE 5-10. Values of $\rho_{\xi(t_2)W\cdot\xi(t_1)}$ for Straight-line Growth Example in Figure 5-1

t_1	$\rho_{W\theta} = 0$	$\rho_{W\theta} = .7$
0	.84	.92
1	.77	.92
2	.56	.91
3	0	.88
4	−.56	.77
5	−.77	.54
6	−.84	.18
7	−.89	−.15
8	−.90	−.37

spectively) for a collection of straight-line growth curves with parameter values $t^0 = 3$, $\kappa = 3$; for case 1, $\rho_{W\xi(t^0)} = .91$, and for Case 2 $\rho_{\theta W} = .7$, $\rho_{W\xi(t^0)} = .6$, $t^\mu = 6.5$, and $t^1 = .43$. With multiwave data, an estimate of $\rho_{\theta W}$ can be obtained by correcting the observed correlation between $\hat{\theta}$ and W for attentuation using a maximum-likelihood estimate of the reliability of $\hat{\theta}$ constructed by substituting estimates from Blomqvist (1977) into Equation 22 of Rogosa et al. (1982).

MYTH 7: ANALYSES OF COVARIANCE MATRICES INFORM ABOUT CHANGE

This myth serves as an umbrella for illustrations of the unattractiveness of three related approaches to the analysis of longitudinal data: path analysis, structural regressions, and simplex models. These three procedures all use the between-wave covariance matrix as the starting point for the statistical analysis. The main message of this myth is that the between-wave covariance matrix provides little information about change or growth. The examples illustrate this message.

Path Regressions Inform About Change?

Path analysis models for longitudinal data use the temporal ordering of the measurements to delimit the possible paths between the variables. Consider the example of a three-wave design with measures on X at times t_1, t_2, t_3. The path regressions for the unstandardized variables are

$$X_2 = \alpha_2 + \beta_1 X_1 + e_2$$
$$X_3 = \alpha_3 + \beta_2 X_2 + \beta_3 X_1 + e_3 \qquad (5.6)$$

Thus, the path analysis model includes direct paths from X_1 to X_2 and to X_3 (parameters β_1 and β_3, respectively) and from X_2 to X_3 (parameter β_2). The path coefficients are functions of the entries of the between-wave covariance matrix. An example of the use of this model is Goldstein (1979), in which X is a reading test score obtained on a nationwide British sample with measurements of ages 7, 11, and 16. Goldstein obtains the following estimates: $\hat{\beta}_1 = .841$, $\hat{\beta}_2 = 1.11$, $\hat{\beta}_3 = -.147$. The negative estimate for β_3 causes considerable discomfort, as summarized by Goldstein:

> This is difficult to interpret and may indicate that non-linear or interaction terms should be included in the model, or perhaps that the change in score between seven and 11 years is more important than the seven-year score itself. However, the addition of non-linear terms does not change this picture to any extent. (p. 139)

(Although not central to the present discussion, Goldstein's analysis employs complex transformations of the measures to straighten the X_i, X_j, scatterplots and disattenuation of the sample regression coefficients.)

Compare those path analysis results with the following simple facts. Let the true scores $\xi(t_i)$ $(i = 1, 2, 3)$ be determined by a straight-line growth curve for each individual (cf. Figure 5-1). Then the partial regression coefficients are

$$\beta_{\xi(t_3)\xi(t_1)\cdot\xi(t_2)} = \frac{t_2 - t_3}{t_2 - t_1} < 0$$
$$\beta_{\xi(t_3)\xi(t_2)\cdot\xi(t_1)} = \frac{t_3 - t_1}{t_2 - t_1} > 0 \qquad (5.7)$$

Remarkably, the parameters depend only on the times at which the observations were taken, and thus neither regression coefficient contains any information about growth! Estimates of either parameter are totally independent of the information in the data. The implications of Eq. (5.7) for the path analysis in Eq. (5.6) are devastating. The first parameter in Eq. (5.7) corresponds to β_3 in Eq. (5.6) and agrees with Goldstein's negative value of $\hat{\beta}_3$, with the magnitude affected by the data transformations and the success of the disattenuation procedures. The second parameter corresponds to β_2 and is consistent with Goldstein's positive value for $\hat{\beta}_2$. Different results for the coefficients in Eq. (5.7) will be obtained for different forms of the individual growth curve. The comparison of the path analysis with the mathematical results for straight-line growth attempts to illustrate some of the perils of summarizing the longitudinal data by the analysis of the between-wave covariance matrix of the X_i or even the $\xi(t_i)$, thereby ignoring the analysis of individual growth.

Structural Regression Models Inform about Change?

Structural regression models are a more sophisticated but equally flawed approach to the analysis of longitudinal data. These models incorporate regression relations among latent variables (i.e., $\xi(t_i)$), with measurement models relating the observed indicators (X_i) to the latent variables. Estimation of these models is based on fitting the covariance structure implied by the structural equation model to the between-wave covariance matrix of the observations. Consider the simple structural regression model shown in Figure 5-4 with one latent variable ξ observed at times t_1 and t_2 and a latent background measure, W. Each latent variable has two indicators. This model is equivalent to the model for change in alienation that appears frequently as an example in Jöreskog's papers. The path from W to ξ_2 represents the exogenous influence on change. The structural

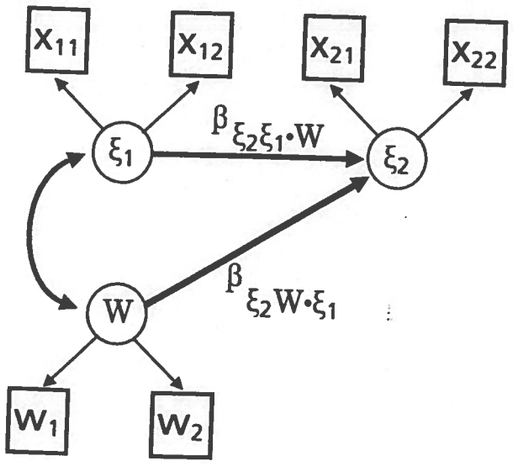

FIGURE 5-4. A depiction of the structural regression model for change in ξ with an exogenous variable, $\mathcal{W}$.

parameter for that path is the regression coefficient for the latent variable at time 2 on the exogenous variable, with the latent variable at time 1 partialed out. In Jöreskog's example, where ξ is alienation and $\mathcal{W}$ is socioeconomic status (SES), a negative estimate of this parameter is interpreted as indicating that high SES reduces alienation.

What does the structural parameter $\beta_{\xi(t_2)\mathcal{W}\cdot\xi(t_1)}$ reveal about exogenous influences on growth? Not very much. For the simple case of a collection of straight-line growth curves, this structural parameter has a complicated functional form that depends strongly on the time chosen for the initial measurement. The time span that pertains to a particular study is unknown and depends on the particular substantive problem. For a specified relation between the exogenous and variable and individual change, the structural parameter may be positive, negative, or zero, depending on the choice of time of initial status. Also, the structural parameter increases with the length of the interval between measurements. Consider two numerical examples based on the collection of growth curves in Figure 5-1: (1) large influences of the exogenous variable ($\rho_{\mathcal{W}\theta} = .7$) and (2) no relation between the exogenous variable and rate of change. Tables 5-11 shows values of the structural parameter for these two cases, with $t_2 - t_1$ of 5 units. The entries in the $\rho_{\theta\mathcal{W}} = 0$ column should be compared with the zero value of the corresponding regression coefficient $\beta_{\Delta(t,\,t+5)\mathcal{W}}$. For $\rho_{\theta\mathcal{W}} = .7$, the entries should be compared with the regression coefficient $\beta_{\Delta(t,\,t+5)\mathcal{W}} = 5\beta_{\theta\mathcal{W}} = .77$. Thus, for both cases the structural regression coefficient may badly mislead about exogenous influences on growth.

TABLE 5-11. Values of Structural Regression Parameter for Straight-line Growth Example

	$\beta_{\xi(t_I+5)\mathcal{W}\cdot\xi(t_I)}$	
t_I	$\rho_{\mathcal{W}\theta} = 0$	$\rho_{\mathcal{W}\theta} = .7$
0	.85	.70
1	1.05	.85
2	1.15	1.0
3	0.0	1.2
4	−1.15	1.3
5	−1.05	1.1
6	−.85	.35
7	−.70	−.25
8	−.55	−.5

Simplex Models Describe Most Longitudinal Data?

A third example of longitudinal analyses based on the between-wave covariance matrix is the simplex model, which specifies a first-order autoregressive process for true scores. The numerical example in this section seeks to caution against the propensity to base many analyses of longitudinal data on a simplex structure without careful consideration of the longitudinal data or of alternative growth models. Expositions of covariance structure analyses have encouraged such thinking. Moreover, Werts, Linn, and Jöreskog (1977) assert "The simplex model appears to be particularly appropriate for studies of academic growth" (p. 745). Well, maybe, maybe not.

Consider the 5 × 5 correlation matrix for observed scores X_{ip} over five occasions of observation in Table 5-12. To the eye, this correlation matrix corre-

TABLE 5-12. Observed-Score Correlation Matrix for Simplex Example

	X_1	X_2	X_3	X_4	X_5
X_1	1.000				
X_2	.746	1.000			
X_3	.727	.741	1.000		
X_4	.695	.723	.741	1.000	
X_5	.656	.695	.727	.746	1.000
Standard deviation	.787	.771	.766	.771	.787

sponds extremely well to a simplex. Correlations decrease away from the diagonals, and on each subdiagonal the correlations are nearly equal. A covariance structure analysis of the corresponding covariance matrix, using LISREL with a quasi-simplex covariance structure, is exceptionally successful. The reproduced covariance and correlation matrices are almost perfect; the root mean square residuals are .003 and .006, respectively. The median discrepancy for the 10 fitted correlations is .003. The chi-square fit statistic, which has five degrees of freedom, is 2.13 (figured for 500 observations) with a p-value of .831. So it seems LISREL is very successful in fitting a simplex model to this example.

Guttman's (1954) condition for a simplex specifies that the partial correlation between earlier and later true scores with an intervening time partialed out is zero. This is the first-order Markov assumption. Straight-line growth turns out to be maximally "unsimplex" in that this partial correlation is −1 instead of 0. (For exponential growth the partial correlation is also −1.) The example in Table 5-12 actually was generated from straight-line growth in the true scores. Thus, the example shows that a simplex covariance structure marvelously fits a covariance matrix from growth curves that are maximally unsimplex. The consequences are far from benign because even when the simplex model fits wonderfully, the results of the covariance structure analysis can badly mislead. The covariance structure analyses usually go on to compute growth statistics and reliability estimates based on the simplex model, and these growth statistics (such as the correlation between true change and true initial status), estimated from the LISREL analysis, can differ markedly from the actual values. Covariance structure analyses provide very limited information about growth, in the sense that covariance matrices arising from very different collections of growth curves can be indistinguishable. Therefore, analyses of covariance structures cannot support conclusions about growth. To reiterate my central message, analysis of the collection of growth curves cannot be ignored.

Reference notes

Rogosa and Willett (1985b, Section 3.2.2) gives mathematical results for the form of the structural regression parameter examined in "Structural Regression Models Inform about Change"? (pp. 193). In their notation the example in Table 5-11 used a collection of straight-line growth curves with parameter values $t^o = 3$, $\kappa = 3$. For Equation 27 of Rogosa and Willett with $\rho_{\theta W} = 0$, $\rho_{W\xi(t^o)} = .91$: $\sigma^2_W = 1$, $\tau = 5$, and $\sigma^2_{\xi(t^o)} = .438$. For Equation 26, with $\rho_{\theta W} = .7$: $\rho_{W\xi(t^o)} = .6$, $t^\mu = 6.5$, and $t^l = .43$. The simplex example is excerpted from the more extensive discussion in Rogosa and Willet (1985a).

MYTH 8: STABILITY COEFFICIENTS ESTIMATE

(a) the consistency over time of an individual
(b) the consistency over time of an average individual
(c) the consistency over time of individual differences
(d) none of the above
(e) some of the above

The absence or obscurity of definitions of stability, along with the proliferation of stability coefficients, results in considerable ambiguity as to what a particular stability coefficient is supposed to be estimating. Thus, it is fitting that this multiple choice myth possess a lack of clarity in the identification of the correct answer. For some stability coefficients (d) is most correct; for others (e) is more correct. Even when (e) is most appropriate, it is not always clear which of (a), (b), (c) would be identified. A coefficient corresponding to choice (a) would be based on an assessment of the heterogeneity (or lack thereof) in an individual's data over time. One procedure corresponding to choice (b) would be inferences about the average growth curve using repeated measures analysis of variance; that is, is the average growth curve flat? Regarding choice (c), correlation coefficients are often used as measures of consistency of individual differences.

Rogosa et al. (1984) formulated two kinds of questions about stability, with application to the stability of behavior. The first question—is an individual consistent over time?—is rarely investigated. Unfortunately, substantive questions about the heterogeneity of an individual's data over time or about individual differences in heterogeneity rarely are addressed.

The second question—are individual differences consistent over time?—has been the focus of most empirical investigations and the major use for the menagerie of stability coefficients. Among the methods used for assessing stability of individual differences are time 1–time 2 correlations, intraclass correlations and generalizability coefficients, repeated-measures ANOVA, path analysis regression, and structural equation models with exogenous variables. The path analysis and structural regression coefficients are described in Wheaton, Muthen, Alwin, and Summers (1977, Figures 1, 2). The intraclass correlation approach fits a correlation matrix to multiwave data with all off-diagonal elements equal. Whenever individual time trends exist in the data, the intraclass correlation model will yield poor results. An example for science education question-asking is Rosenshine (1973), in which a zero intraclass coefficient is obtained because the between-wave correlation matrix contains both big positive and big negative entries.

The most attractive approach to assessing consistency of individual differences is the indices of tracking from the biometric literature, which assess maintenance of individual differences over time. Figure 5-5 depicts collections of

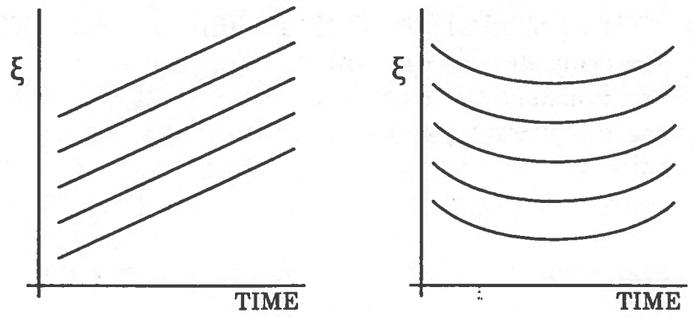

FIGURE 5-5. Two illustrations of perfect consistency of individual differences over time.

growth curves displaying perfect maintenance of individual differences over time; in Figure 5-5 individual differences are consistent across time whether the criterion is maintenance of rank order or of absolute distance. The index of tracking, γ, presented by Foulkes and Davis (1981) assesses maintenance of rank order over time; this index is the probability of two growth curves not crossing in the specified time interval. Intersections of the individual growth curves are thus evidence against tracking. No tracking is said to exist for $\gamma \geqslant .5$, the "chance level" for the probability of no crossings. As the time interval is lengthened, γ tends to decrease, as it is more difficult to maintain individual differences over a longer interval.

Data on physical growth are used to illustrate the assessment of stability of individual differences. Measurements of the height (in millimeters) of the mandibular ramus bone on a sample of 20 boys at four half-year intervals from 8.0 to 9.5 years of age are given in Goldstein (1979, Table 4.1) and have been used as an illustrative example in many papers on the analysis of growth curves. Each individual's data are very well described by a straight-line growth curve; the median squared multiple correlation for the fit of a straight line to the four observations is .95 for this sample, with upper and lower quartiles of .99 and .91. Figure 5-6 plots the 20 fitted straight-line growth curves. The estimate of the Foulkes-Davis γ index of tracking is .826, with an estimated standard error of .032. Thus, these data show strong, but not perfect, maintenance of individual differences over the 18-month interval.

Whereas the index of tracking provides a useful quantification of the consistency of individual differences, the stability coefficients widely used in the behavioral and social sciences mainly provide confusion. Numerical examples based on the collection of straight-line growth curves in Figure 5-1 are used to illustrate the properties of some of the stability coefficients. The coefficients for measurements over a time interval $[t_I, t_F]$ are:

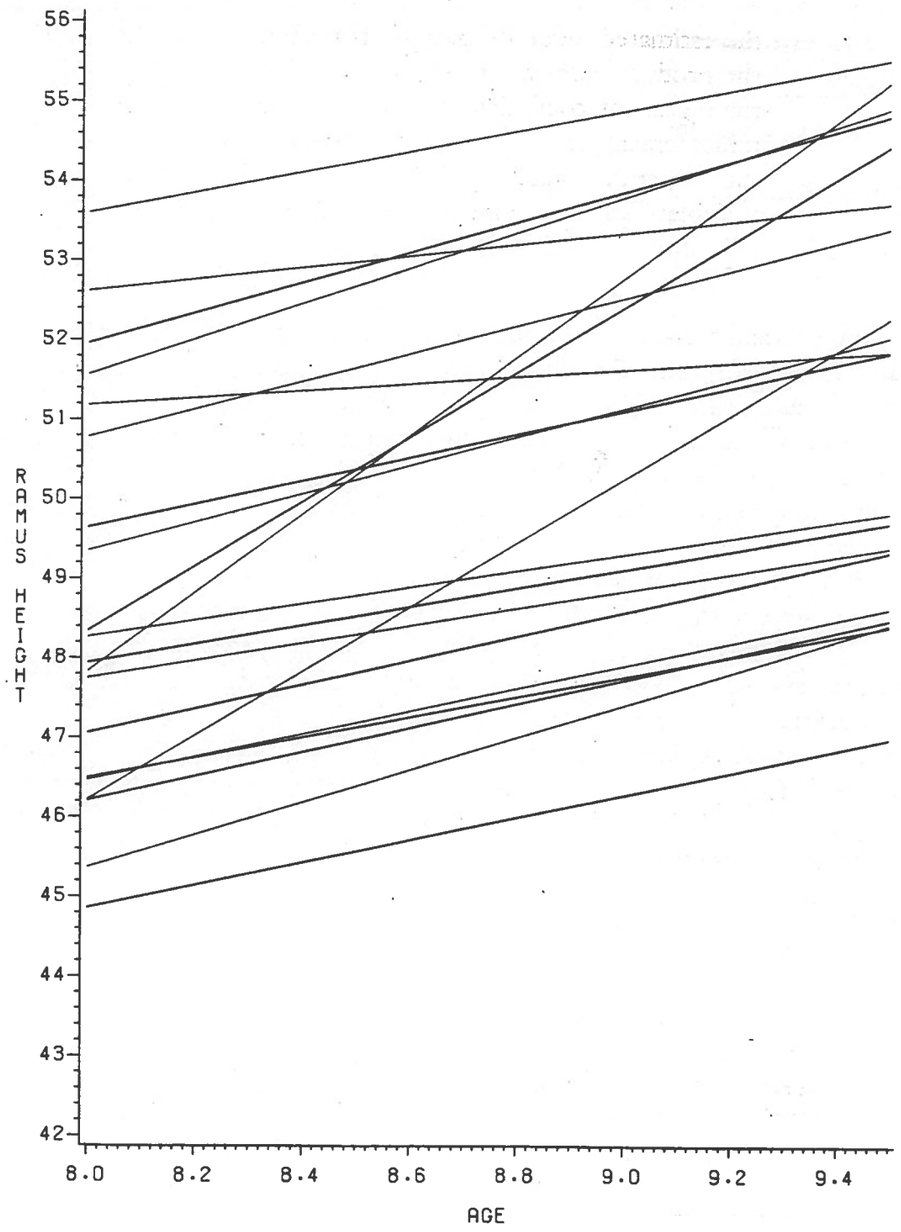

FIGURE 5-6. The fitted straight-line growth curves for the ramus data.

$\hat{\gamma}(t_I, t_F)$ the estimated index of tracking from Foulkes and Davis (1981),
$\rho_{\xi(t_I)\xi(t_F)}$ the product-moment correlation,
$\beta_{\xi(t_F)\xi(t_I)}$ the regression coefficient for later on earlier consecutive waves of measurement proposed by Heise (1969),
$\beta_{\xi(t_F)\xi(t_I)\cdot\mathbb{W}}$ the structural regression coefficient for later on earlier latent variables, with an exogenous variable partialed out, used by Wheaton et al. (1977).

Tables 5-13 and 5-14 are structured to show the effects of different $[t_I, t_F]$ intervals on the coefficients. The values of all coefficients except $\hat{\gamma}$ are determined by formulas using the population moments of the collection of growth curves; only $\hat{\gamma}$ is based on the $\xi(t_i)$ values for the 15 growth curves and has an estimated standard error less than .05 for all time intervals in the tables. All coefficients are computed in terms of true scores; only $\hat{\gamma}$ will be relatively unaffected by errors of measurement.

The coefficients differ among themselves for a given $[t_I, t_F]$ interval and differ, often in strange ways, over different intervals. Using the criterion $\hat{\gamma}$ - 2[s.e.($\hat{\gamma}$)] > .50, tracking exists for $[t_I, 7]$ in Table 5-13 for $t_I \geqslant 3$, and for $[0, t_F]$ in Table 5-14, tracking exists for $t_F \leqslant 4$. None of the other stability coefficients has an easily interpretable scale. In fact, for the same degree of consistency of individual differences (as assessed by $\hat{\gamma}$) the other coefficients vary wildly. Table 5-15 displays two sets of $[t_I, t_F]$ intervals with matching on the values of $\hat{\gamma}$. For the intervals [0,4] and [3,7] individual differences track according to the Foulkes-Davis $\hat{\gamma}$, yet the regression coefficients are small, or negative for [0,4] and much larger for [3,7]. The second set of intervals [5,7] and [0,3] show stronger tracking and similar discordance in the regression coefficients.

TABLE 5-13. Stability Coefficients for Straight-line Growth over the Interval $[t_I, 7]$

				$\beta_{\xi(7)\xi(t_I)\cdot\mathbb{W}}$	
t_I	$\hat{\gamma}(t_I,7)$	$\rho_{\xi(t_I)\xi(7)}$	$\beta_{\xi(7)\xi(t_I)}$	$\rho_{\mathbb{W}\theta}=0$	$\rho_{\mathbb{W}\theta}=.7$
0	.47	−.14	−.17	−.98	−.09
1	.53	.06	.08	−1.1	−.06
2	.58	.32	.50	−.93	−.007
3	.69	.60	1.00	1.0	−.12
4	.78	.82	1.30	2.16	.40
5	.88	.94	1.31	1.71	.83
6	.97	.99	1.17	1.28	1.10

TABLE 5-14. Stability Coefficients for Straight-line Growth over the Interval $[0, t_F]$

				$\beta_{\xi(t_F)\xi(0)\cdot\mathbb{W}}$	
t_F	$\hat{\gamma}(0, t_F)$	$\rho_{\xi(0)\xi(t_F)}$	$\beta_{\xi(t_F)\xi(0)}$	$\rho_{\mathbb{W}\theta}=0$	$\rho_{\mathbb{W}\theta}=.7$
1	.93	.98	.83	.71	.85
2	.89	.89	.67	.42	.70
3	.86	.71	.50	.13	.55
4	.69	.45	.33	−.16	.40
5	.59	.12	.17	−.45	.25
6	.49	.001	0	−.75	.11
7	.47	−.14	−.16	−1.04	−.04

TABLE 5-15. Comparisons of Stability Coefficients for Intervals Having the Same Tracking Index

				$\beta_{\xi(t_F)\xi(t_I)\cdot\mathbb{W}}$	
$[t_I, t_F]$	$\hat{\gamma}(t_I, t_F)$	$\rho_{\xi(t_I)\xi(t_F)}$	$\beta_{\xi(t_F)\xi(t_I)}$	$\rho_{\mathbb{W}\theta}=0$	$\rho_{\mathbb{W}\theta}=.7$
[0, 4]	.69	.60	.33	−.16	.40
[3, 7]	.69	.45	1.0	1.0	.12
[5, 7]	.88	.94	1.3	1.71	.83
[0, 3]	.86	.71	.50	.13	.55

Reference notes

Wohlwill (1973, Chap. 12) provides a lucid discussion and illustration of research questions about stability arising in developmental research. Foulkes and Davis (1981) and McMahan (1981) propose indices of tracking to assess consistency of individual differences. Rogosa and Willett (1983a) provide empirical comparisons of the two indices. Rogosa et al. (1984) formulate research questions about the stability of behavior; they also develop and illustrate statistical procedures for the assessment of stability. The parameter values displayed in the tables are obtained from results in Rogosa and Willett (1985b).

MYTH 9: CASUAL ANALYSES SUPPORT CAUSAL INFERENCES ABOUT RECIPROCAL EFFECTS

The best-known procedure associated with Myth 9 is cross-lagged correlation. A remarkable statement of the myth is provided by Crano and Mellon (1978): "With the introduction of the cross-lagged panel correlation method . . . , causal inferences based on correlational data obtained in longitudinal studies can be

made and enjoy the same logical status as those derived in the more standard experimental settings" (p. 41). In other words, the use of cross-lagged correlation dispenses with the need for experiments, statistical models or careful data analysis; a quick comparison of a few correlation coefficients is all that is required to study reciprocal effects. Well, I suppose that would be wonderful if it were true.

The important thing to keep in mind is that questions about reciprocal effects are very, very complex and difficult. A hierarchy of research questions about longitudinal data might start with describing how a single attribute—say, aggression—changes over time. A next step would be questions about individual differences in change of aggression over time, especially correlates of change in aggression. Only after such questions are well understood does it seem reasonable to address a question about feedback or reciprocal effects, such as how change in aggression relates to change in exposure to TV violence or, does TV violence cause aggressive behavior? Despite the complexity of research questions about reciprocal effects, empirical research has attempted to answer the oversimplified question, does X cause Y or does Y cause X? by casually comparing a couple of correlations.

The mathematical and numerical demonstrations of the failures of cross-lagged correlation in Rogosa (1980) had the following simple, limited structure. Start with a basic path-analysis regression model for two variables, X and Y, measured at times 1 and 2 (the popular two-wave, two-variable panel design)

$$X_2 = \beta_0 + \beta_1 X_1 + \gamma_2 Y_1 + u,$$
$$Y_2 = \gamma_0 + \beta_2 X_1 + \gamma_1 Y_1 + v. \qquad (5.8)$$

In the context of the statistical model in Eq. (5.8) the parameters β_1 and γ_1 represent the influence of a variable on itself over time. The parameters β_2 and γ_2 represent the lagged, reciprocal causal effects between X and Y; thus, the relative magnitudes of β_2 and γ_2 indicate the nature of the reciprocal causal effects. In Rogosa (1980) combinations of β_2 and γ_2 values are compared with the results of the method of cross-lagged correlation. Three examples from Rogosa (1980) are shown in Figure 5-7. In the first frame, the cross-lagged correlations are equal (.63), which indicates the conclusion of "spuriousness," no direct causal influences between X and Y, even though the model Eq. (5.8) stipulated that the effect from X to Y ($\beta_2 = .42$) is twice the effect from Y to X ($\gamma_2 = .21$). In the second frame the model stipulates lagged influences of equal magnitude, yet cross-lagged correlation identifies X as the causal winner. In the third frame the model stipulates an effect from Y to X nearly double the effect from X to Y. Yet the attribution of causal predominance by cross-lagged correlation is the opposite—X would be chosen the causal winner. These examples are simplified

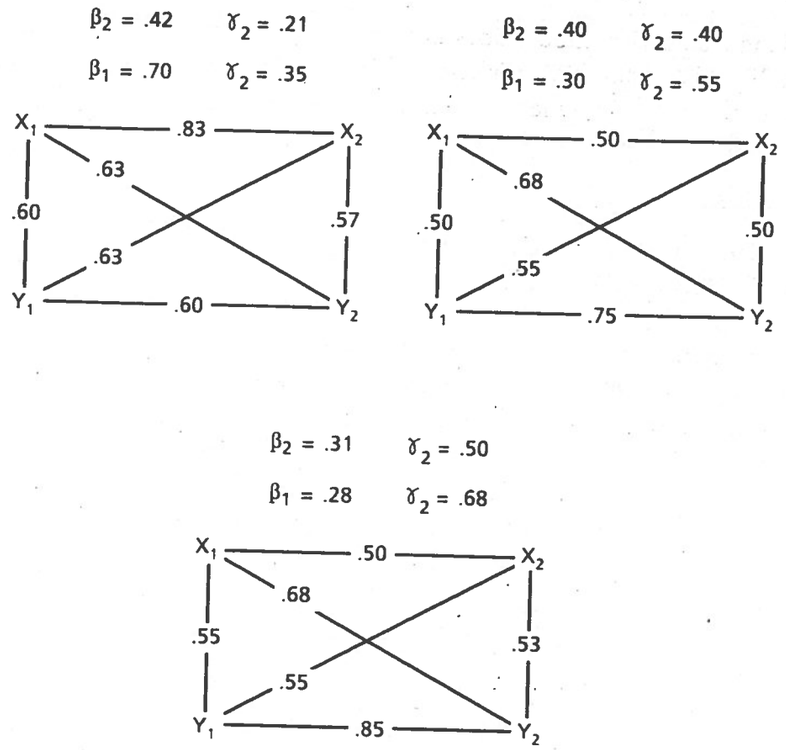

FIGURE 5-7. Numerical illustrations of misleading cross-lagged correlations in two-wave, two-variable panel data.

by the assumption of equal variances for X and Y; when variances change over time, equations in Rogosa (1980) show that the comparison of the cross-lagged correlations is even more unsatisfactory.

The major (and perhaps only) virtue of the path analysis model Eq. (5.8) is the identification of specific parameters believed to represent the reciprocal effects. If this model of the reciprocal influences between X and Y were valid, then estimation of β_2 and γ_2 would inform about reciprocal effects. Perhaps the best way to think about (Eq. 5.8) and the related structural regression models is that these comprise a simple statistical model for reciprocal effects which, however, may be a far from satisfactory scientific model of the psychological (etc.) process.

The real moral about the analysis of reciprocal effects is that you can't estimate something without first defining it, and statistical models are a good way of defining the key parameters. But this does not imply that all statistical models are sensible. The progress that has been made, especially in the use of structural equation modeling, is to move from no model at all to some statistical model.

But having a statistical model does not mean it is an adequate scientific model. Regrettably, the seductive simplicity of cross-lagged correlation has inhibited serious work on the complex question of reciprocal effects.

Reference notes

Rogosa (1980) was only one in a tradition of papers, starting with Goldberger (1971) and Heise (1970), sharply critical of cross-lagged correlation. Even Cook and Campbell (1979, chap. 7) are unenthusiastic about the usefulness of cross-lagged correlation, yet most advocates and users of this procedure remain undaunted. Rogosa (1980) exposits a number of simple statistical models for reciprocal effects between two variables—structural regression models, continuous-time feedback models, and multiple-time series models. The mathematical results in Rogosa (1980) demonstrate the inability of the method of cross-lagged correlation to recover the structure of the reciprocal effects specified by these models. Results and numerical examples are presented for two-wave and multiwave data. Rogosa (1985) provides a nontechnical overview and extensive references on approaches to the analysis of reciprocal effects.

DISCUSSION

The message of the myths, which is carried through into my work on statistical methods for longitudinal data, is that models for collections of growth curves are the proper basis for the statistical analysis of longitudinal data. The nature of research questions about growth and development makes these models a natural, if not essential, starting point. What I tried to do with these myths was to indicate some of the beliefs that have impeded doing good longitudinal research. The myths have served either to make the analysis of change appear prohibitively difficult or to direct research in unproductive directions. Rather simple approaches work well with longitudinal data, and much progress can be made using straightforward descriptive analysis of individual trajectories followed by statistical estimation procedures for collections of growth curves. Although only a small number of observations often are available in empirical research, the resulting difficulties in statistical estimation arising from these limited longitudinal designs should not alter the research questions or the proper statistical models.

The nine myths discussed in this chapter are not exhaustive. Two additional candidates deserve some mention. The first could be stated as "The average growth curve informs about individual growth." This myth dominated practice in psychological learning experiments, although Estes (1956) demonstrated that the learning curve obtained from averaging individual responses at each trial was

equivalent to the average of the individual learning curves only for special forms of the learning curve. This myth has also impeded studies of physical maturation (Bock, 1979). Another setting for this myth is the analysis of longitudinal data with a hierarchical or multilevel structure (Rogosa, 1979, pp. 168–174). A second candidate myth is that "Standardizing longitudinal data can be useful." An inexplicable champion of this myth is Goldstein (1983). Standardization renders impossible useful analyses of longitudinal data by removing essential information about individual growth and individual differences in growth. A related, but complex, issue is the effects of different metrics/transformations of X on the longitudinal analysis.

The myths speak against what I call the "Avoid Change at Any Cost Academy of Longitudinal Research," which recommends analyses that try to draw complex conclusions about change over time without any examination of individual growth. That doctrine appears counterproductive, as these myths and my technical papers so demonstrate. The doctrine of this Academy is sometimes justified by overinterpretations of the often-quoted last sentence of Cronbach and Furby (1970, p. 79): "Investigators who ask questions regarding gain [difference] scores would ordinarily be better advised to frame their questions another way." This statement could be regarded as a meta-myth. The factual basis for their conclusion is the shortcomings of the estimate of the amount of change from only two observations. But such facts do not support abandoning the framing of research questions about growth and change in a natural way. The suggested surrender to uninformative regression and residual-change analyses is to be much lamented; the proper lesson to draw from difficulties with the difference score is that richer longitudinal designs and the application of appropriate statistical models for the longitudinal data are needed.

An appropriate question to be raised at this point is, where do we go from here? The myths serve more to discredit popular analysis procedures than to prescribe replacements. This function is important in the sense of "first things first"; the groundwork for new approaches requires some appreciation of the flaws of past and current thinking.

Statistical methods respond to (well-formulated) research questions. Naturally, there is no *single* statistical procedure for the analysis of longitudinal data; different research questions dictate different data structures and thus different statistical models and methods. Although at present, the "toolkit" of dependable methods for the analysis of longitudinal data is not complete, I do believe that the natural approach of statistical modeling of individual time trajectories (promoted in this chapter and in my technical papers) serves well as the common basis for the development of statistical methods. To follow on this theme of the linking of research questions and useful statistical methods, I close this chapter with an organization of seven research topics (questions) commonly addressed with longitudinal data. The parenthetical listing of Myths under each topic in-

dicates relevant portions of this chapter, but no attempt is made here to survey the available statistical procedures and literature.

1. *Individual and group growth (Myths 1, 2, 5, 6).* A basic type of question in longitudinal research concerns description of the form and amount of change. Such questions may be posed for an individual case or for the average of a group or subgroup of cases. Interest centers on the estimation of the individual (or group) growth curve, the heterogeneity (individual differences) in the individual growth curves, and the statistical and psychometric properties of these estimates.

2. *Correlates and predictors of change (Myths 6, 7).* Questions about systematic individual differences in growth are a natural sequel to the description of individual growth. A typical research question is given by "What kind of persons learn [grow] fastest?" (Cronbach & Furby, 1970, p. 77). The key quantities are the associations between parameters of the individual growth curves and the correlate(s) of change, which may be exogenous individual characteristics (e.g., gender, IQ) or the initial status on the attribute measured over time.

3. *Stability over time (Myth 8).* Questions about consistency over time are a natural complement to questions about change. In the behavioral sciences literature many different research questions fall under the heading of "stability." Two key topics are the consistency over time of an individual and the consistency of individual differences over time.

4. *Comparing experimental groups.* The comparison of change across experimental groups is a standard, well-developed area of statistical methodology employing some form of repeated measures analysis of variance. When the effects of each treatment can be assumed identical for all members within each group (no individual differences in response to treatment), comparison of the parameters of the group growth curves yields inferences about the "treatment effects."

5. *Comparing nonexperimental groups (Myths 1, 6).* The comparison of change among nonexperimental or nonequivalent groups has been a central topic in the methodology for the evaluation of social programs. The practical or political difficulties of random assignment of individuals to treatment are sometimes overwhelming in a field trial of a program. Yet the question of the relative efficacies of each program/treatment remains. The extensive literature on this topic is dominated by the application of statistical adjustment procedures (analysis of covariance and relatives) to very meager (pretest-posttest) longitudinal data.

6. *Analysis of reciprocal effects (Myth 9).* As discussed in Myth 9, questions about reciprocal effects are common and complex. Clearly, considerable empirical research on simpler longitudinal questions should precede attempts to assess reciprocal effects. Despite the complexity of these questions, empirical research has attempted to answer the oversimplified question, "Does X cause Y or does Y cause X?" from meager longitudinal data by casually comparing a couple of correlations (or structural regression coefficients).

7. *Growth in multiple measures.* All questions about growth in a single attribute have extensions to multiple attributes. Natural questions include relative strengths and weaknesses in individual and group growth and associations of rates of growth across multiple attributes.

REFERENCES

Anderson, J. E. (1939). The limitations of infant and preschool tests in the measurement of intelligence. *Journal of Psychology, 8,* 351–379.

Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in the measurement of change* (pp. 3–20). Madison, WI: University of Wisconsin Press.

Blomqvist, N. (1977). On the relation between change and initial value. *Journal of the American Statistical Association, 72,* 746–749.

Bloom, B. S. (1964). *Stability and change in human characteristics.* New York: Wiley.

Bond, L. (1979). On the base-free measure of change proposed by Tucker, Damarin, and Messick. *Psychometrika, 44,* 351–355.

Coleman, J. S. (1968). The mathematical study of change. In H. M. Blalock & A. B. Blalock (Eds.), *Methodology in social research* (pp. 428–478). New York: McGraw-Hill.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis for field settings.* Boston: Houghton Mifflin.

Crano, W. D., & Mellon, P. M. (1978). Causal influence of teachers' expectations on children's academic performance: A cross-lagged panel analysis. *Journal of Educational Psychology, 70,* 39–49.

Cronbach L. J., & Furby, L. (1970). How should we measure "change"—or should we? *Psychological Bulletin, 74,* 68–80.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin, 53,* 134–140.

Foulkes, M. A., & Davis, C. E. (1981). An index of tracking for longitudinal data. *Biometrics, 37,* 439–446.

Furby, L. (1973). Interpreting regression toward the mean in development research. *Developmental Psychology, 8,* 172–179.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute, 15,* 246–263.

Goldberger, A. S. (1971). Econometrics and psychometrics: A survey of communalities. *Psychometrika, 36,* 83–105.

Goldstein, H. (1979). *The design and analysis of longitudinal studies.* London: Academic Press.

Goldstein, H. (1983). Measuring changes in educational attainment over time: Problems and possibilities. *Journal of Educational Measurement, 20,* 369–378.

Guttman, L. A. (1954). A new approach to factor analysis: The radex. In P. F. Lazarsfeld

(Ed.), *Mathematical thinking in the social sciences.* New York: Columbia University Press.

Healy, M. J. R., & Goldstein, H. (1978). Regression to the mean. *Annals of Human Biology, 5,* 277–280.

Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review, 34,* 93–101.

Heise, D. R. (1970). Causal inference from panel data. In E. F. Borgatta & G. W. Bohrnstedt (Eds.), *Sociological methodology, 1970.* San Francisco: Jossey-Bass.

Humphreys L. G. (1960). Investigations of the simplex. *Psychometrika, 25,* 313–323.

Jöreskog, K. G. (1979). Analyzing psychological data by structural analysis of covariance matrices. In K. G. Jöreskog & D. Sorböm (Eds.), *Advances in factor analysis and structural equation models.* Cambridge, MA: Abt Books.

Lacey, J. I., & Lacey, B. C. (1962). The law of initial value in the longitudinal study of autonomic constitution: Reproducibility of autonomic responses and response patterns over a four year interval. In W. M. Wolf (Ed.), Rythmic functions in the living system. *Annals of the New York Academy of Sciences, 98,* 1257–1290.

Linn, R. L., & Slinde, J. A. (1977). The determination of the significance of change between pre- and post-testing periods. *Review of Educational Research, 47,* 121–150.

Lord, F. M. (1956). The measurement of growth. *Educational and Psychological Measurement, 16,* 421–437.

Lord, F. M. (1958). Further problems in the measurement of growth. *Educational and Psychological Measurement, 18,* 437–454.

Lord, F. M. (1963). Elementary models for measuring change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 21–38). Madison, WI: University of Wisconsin Press.

McMahan, C. A. (1981). An index of tracking. *Biometrics, 37,* 447–455.

Nesselroade, J. R., Stigler, S. M., & Baltes, P. B. (1980). Regression toward the mean and the study of change. *Psychological Bulletin, 88,* 622–637.

Nielson, F., & Rosenfeld, R. A. (1981). Substantive interpretations of differential equation models. *American Sociological Review, 46,* 159–174.

O'Connor, E. F. (1972). Extending classical test theory to the measurement of change. *Review of Educational Research, 42,* 73–98.

Rogosa, D. R. (1980). A critique of cross-lagged correlation. *Psychological Bulletin, 88,* 245–258.

Rogosa, D. R. (1985). Analysis of reciprocal effects. In T. Husen & N. Postlethwaite (Eds.), *International encyclopedia of education* (pp. 4221–4225). London: Pergamon Press.

Rogosa, D. R. (1979). Time and time again: Some analysis problems in longitudinal research. In C. E. Bidwell & D. M. Windham, (Eds.), *The analysis of educational productivity, volume II: Issues in microanalysis* (pp. 153–201). Boston MA: Ballinger Press.

Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 90,* 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,* 795–796.

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. (1985a). Satisfying a simplex structure is simpler than it should be. *Journal of Educational Statistics, 10,* 99–107.

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

Rosenshine, B. (1973). The smallest meaningful sample of classroom transactions. *Journal of Research in Science Teaching, 10,* 221–226.

Salemi, M. K., & Tauchen, G. E. (1982). Estimation of nonlinear learning models. *Journal of the American Statistical Association, 77,* 725–731.

Tucker, L. R., Damarin, F., & Messick, S. A. (1966). A base-free measure of change. *Psychometrika, 31,* 457–473.

Tuma, N. B., & Hannan, M. T. (1984). *Social dynamics: Models and Methods.* New York: Academic Press.

Werts, C. E., Linn, R. L., & Joreskog, K. G. (1977). A simplex model for analyzing academic growth. *Educational and Psychological Measurement, 37,* 745–756.

Wheaton, B., Muthen, B., Alwin, D., & Summers, G. (1977). Assessing reliability and stability in panel models with multiple indicators. In D. R. Heise (Ed.), *Sociological methodology 1977* (pp. 84–136). San Francisco: Jossey-Bass.

Wilder, J. (1957). The law of initial value in neurology and psychiatry. *Journal of Nervous and Mental Disease, 125,* 73–86.

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

interval. In W. M. Wolf (Ed.), Rhythmic functions in the living system. *Annals of the New York Academy of Science, 98,* 1257–1290.
Linn, R. L., & Slinde, J. A. (1977). The determination of the significance of change between pre- and post-testing periods. *Review of Educational Research, 47,* 121–150.
Lord, F. M. (1956). The measurement of growth. *Educational and Psychological Measurement, 16,* 421–437.
Lord, F. M. (1958). Further problems in the measurements of growth. *Educational and Psychological Measurement, 18,* 437–454.
Lord, F. M. (1963). Elementary models for measuring change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 21–38). Madison, WI: University of Wisconsin Press.
McMahan, C. A. (1981). An index of tracking. *Biometrics, 37,* 447–455.
Nesselroade, J. R., Stigler, S. M., & Baltes, P. B. (1980). Regression toward the mean and the study of change. *Psychological Bulletin, 88,* 622–637.
Nielson, F., & Rosenfeld, R. A. (1981). Substantive interpretations of differential equation models. *American Sociological Review, 46,* 159–174.
O'Connor, E. F. (1972). Extending classical test theory to the measurement of change. *Review of Educational Research, 42,* 73–98.
Rogosa, D. R. (1979). Time and time again: Some analysis problems in longitudinal research. In C. E. Bidwell & D. M. Windham (Eds.), *The analysis of educational productivity. Vol. II: Issues in microanalysis* (pp. 153–201). Boston, MA: Ballinger Press.
Rogosa, D. R. (1980). A critique of cross-lagged correlation. *Psychological Bulletin, 88,* 245–258.
Rogosa, D. R. (1985). Analysis of reciprocal effects. In T. Husen & N. Postlethwaite (Eds.), *International encyclopedia of education* (pp. 4221–4225). London: Pergamon Press.
Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 90,* 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,* 795–796.
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. (1985a). Satisfying a simplex structure is simpler than it should be. *Journal of Educational Statistics, 10,* 99–107.
Rogosa, D. R., & Willett, J. B. (1985b). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50,* 203–228.
Rosenshine, B. (1973). The smallest meaningful sample of classroom transactions. *Journal of Research in Science Teaching, 10,* 221–226.
Salemi, M. K., & Tauchen, G. E. (1982). Estimation of nonlinear learning models. *Journal of the American Statistical Association, 77,* 725–731.
Tucker, L. R., Damarin, F., & Messick, S. A. (1966). A base-free measure of change. *Psychometrika, 31,* 457–473.
Tuma, N. B., & Hannan, M. T. (1984). *Social dynamics: Models and methods.* New York: Academic Press.
Werts, C. E., Linn, R. L., & Jöreskog, K. G. (1977). A simplex model for analyzing academic growth. *Educational and Psychological Measurement, 37,* 745–756.
Wheaton, B., Muthen, B., Alwin, D., & Summers, G. (1977). Assessing reliability and stability in panel models with multiple indicators. In D. R. Heise (Ed.), *Sociological methodology, 1977* (pp. 84–136). San Francisco: Jossey-Bass.
Wilder, J. (1957). The law of initial value in neurology and psychiatry. *Journal of Nervous and Mental Disease, 125,* 73–86.
Wohlwill, J. F. (1973). *The study of behavioral development.* New York: Academic Press.

SUPPLEMENTAL QUESTIONS AND ANSWERS

1. SHOULD LISREL (STRUCTURAL EQUATION MODELING) ANALYSES ALWAYS BE AVOIDED?

In general, my answer is yes. Analyses of relationships among variables are fundamentally inadequate and askew, because such analyses do not address the individual level processes that generate the data. My analyses and results in making this argument are less ambitious than the heroic efforts of David Freedman (1987, 1991) who takes on these modeling issues in real-life research settings. In his critique of path analysis applications, Freedman (1987) makes an important appeal for more serious (rather than casual) attention to model building: "My opinion is that investigators need to think more about the underlying social processes . . .". and he argues that "as if by experiment" conclusions "must depend on a theory of how the data came to be generated." Continuing this theme, Freedman (1991) promotes the value of "shoe leather" science (close examination of the phenomena) as contrasted with the social science practice of (causal) inferences based on regression models for distant information (e.g., survey data, archival data).

For the longitudinal research setting, my answer is emphatically yes if the goal is to address longitudinal research questions like those listed previously. Longitudinal research examples have been prominent in expositions and illustrations of structural equation methods, and claims for the usefulness of structural equation methods are common—for example, according to Alwin (1988), structural equation methods "are perhaps most useful in longitudinal research designs where the research questions involve the descriptive analysis of change and its explanation" (p. 74). But the facts are that the parameters estimated in the standard structural equation model applications have little or no relevance to parameters of interest (i.e., those defined by useful longitudinal research questions). The main problems with the use of structural equation models is not in the details of those estimation procedures, but in the meaninglessness of the parameters being estimated. To supplement the brief exposition in Myth 7, here I give some detailed data examples of the inadequacies of the standard structural equation models approach, continuing the results and examples in Rogosa (1987, 1993). The expository strategy is to create an example of longitudinal data with simple and known structure and then see what results would be indicated by the standard structural equation modeling analyses.

The data example was chosen to be small and manageable. From a population of individuals, a data set of 40 cases, each observed at three time points, is drawn. For each individual the true observations fall on a straight-line growth curve (as in Figure 1.1). So for each case there is a longitudinal record with the times of observation having values {1, 3, 5}; in addition there is a background exogenous variable for each individual. Shown in Exhibit 1 are the values of the true scores, denoted by $\xi(t_i)$, and the exogenous variable W. The mean rate of change in the population is 5, with individual rates of change ranging between 0 and 9. The rate of change has zero correlation with the background variable, W.

EXHIBIT 1: STRAIGHT-LINE GROWTH DATA, LISREL EXAMPLE

CASE	ξ(1)	ξ(3)	ξ(5)	W
1	37.56	49.29	61.02	15.97
2	45.65	51.58	57.51	15.38
3	40.94	52.88	64.82	11.48
4	47.36	55.45	63.54	16.89
5	52.71	62.70	72.70	19.18
6	30.45	46.34	62.23	11.82
7	43.65	58.37	73.09	15.33
8	41.16	49.26	57.37	13.21
9	44.15	52.00	59.84	13.09
10	38.16	46.59	55.03	10.32
11	37.68	39.87	42.06	10.26
12	45.30	54.38	63.47	15.60
13	39.37	48.15	56.94	13.90
14	36.66	43.75	50.84	13.53
15	53.40	62.32	71.23	14.45
16	59.35	62.80	66.25	20.16
17	53.14	64.35	75.56	16.11
18	44.90	58.82	72.75	15.06
19	41.79	59.44	77.09	18.33
20	38.25	48.98	59.71	13.77
21	47.24	60.79	74.34	15.88
22	53.57	67.71	81.84	18.25
23	35.54	43.51	51.48	10.15
24	37.54	50.25	62.95	9.46
25	37.07	49.71	62.35	15.81
26	32.40	44.69	56.98	11.60
27	45.22	62.08	78.94	14.08
28	35.67	47.42	59.17	12.19
29	38.30	51.13	63.97	14.07
30	52.61	55.52	58.42	16.68
31	38.36	48.49	58.62	15.07
32	45.14	51.44	57.73	13.94
33	53.82	64.27	74.73	20.40
34	49.46	61.42	73.39	16.00
35	56.29	59.04	61.80	17.47
36	49.59	57.58	65.57	17.30
37	41.45	59.43	77.41	15.86
38	47.42	57.42	67.43	18.95
39	57.00	65.73	74.47	18.90
40	41.06	43.54	46.03	13.79

Data Description.

	MEAN	MEDIAN	STDEV
ξ(1)	44.16	43.90	7.24
ξ(3)	54.21	53.63	7.24
ξ(5)	64.27	63.21	9.24
W	14.99	15.20	2.803

Correlations

	ξ(1)	ξ(3)	ξ(5)
ξ(3)	0.842		
ξ(5)	0.536	0.907	
W	0.766	0.765	0.598

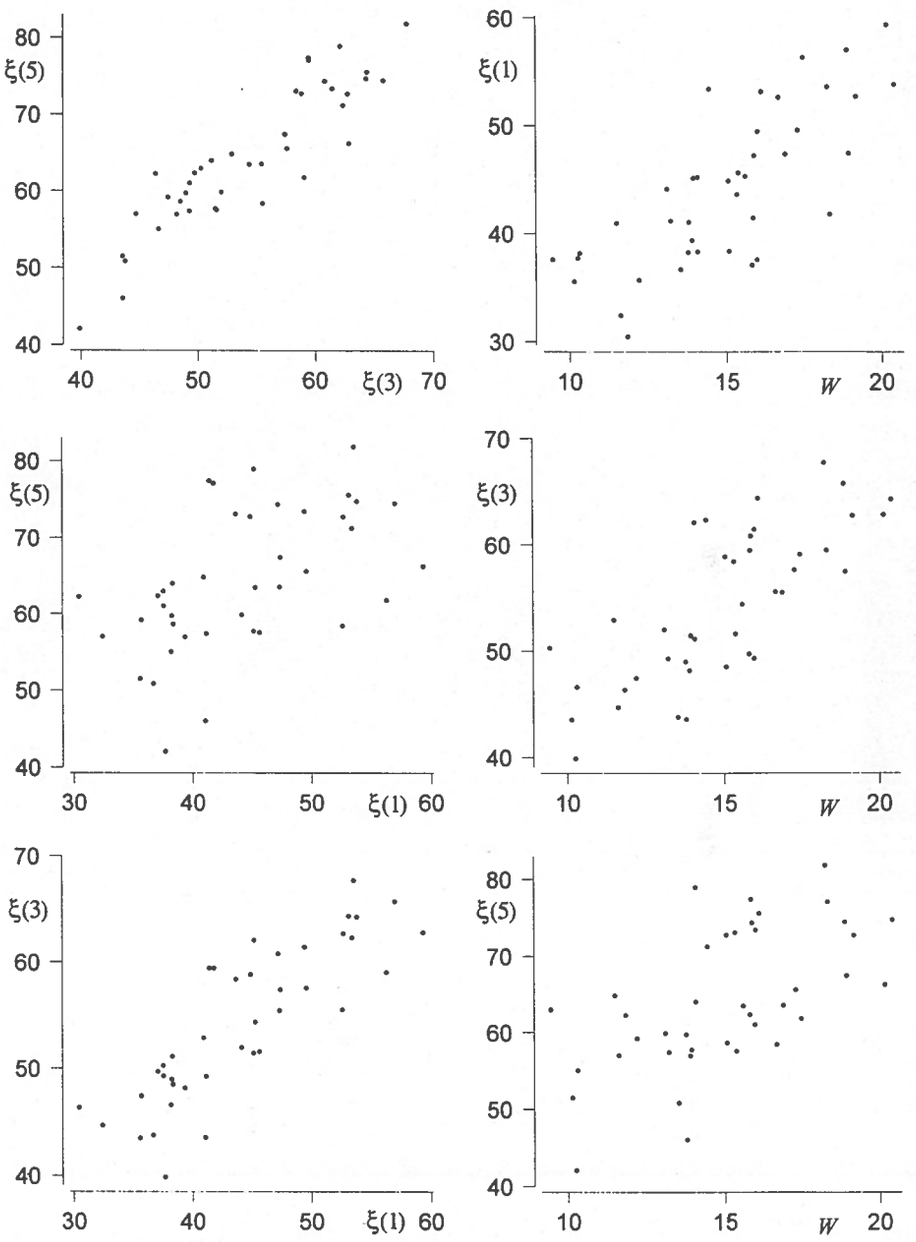

Exhibit 1: Scatterplots between the ξ-values and scatterplots between *W* and the ξ-values.

The scatterplots displayed in Exhibit 1 show that although the $\xi(t_i)$ are generated from straight-line individual growth curves, the between-variable scatterplots appear rather ordinary. That is, the standard view of between-variable relationships would not cause any concerns for a between-variable regression analysis.

Causal Influences on Change: Three-Waves, Single Variable

In this illustration we revisit the substantive setting of the first section of Myth 7 (relating to the discussion of the Goldstein example). With three observations on each individual, what can be learned about individual change, individual differences in change, and so forth? To supplement the argument in Myth 7 that the three-wave path analysis is uninformative, the data example in Exhibit 1 is used to illustrate the results in Equation 1.7. In terms of true scores, the pictorial form of the structural regression model is shown in Figure 1.8.

The regression for $\xi(t_3)$ matches exactly the theoretical results from Equation 1.7—$\beta_3 = (3 - 5)/(3 - 1)$ and $\beta_2 = (5 - 1)/(3 - 1)$—with squared multiple correlation of 1.0. The "structural coefficients" contain no information from the data, nonetheless about causal effects. So what can be learned about change from such an analysis? Annotated MINITAB output for the regression is

```
The regression equation is
ξ(5) = −0.000003 − 1.00 ξ(1) + 2.00 ξ(3)

Predictor          Coef      Stdev      t-ratio   p
Constant    −0.00000309    0.00000            *   *
ξ(1)          −1.00000       0.00            *   *
ξ(3)           2.00000       0.00            *   *

s = 0          R-sq = 100.0%      R-sq(adj) = 100.0%
```

Rogosa (1993, Equation 6 and Figure 4) gives the theoretical result for the corresponding path analysis regression on fallible observed scores. For this data example observed scores were generated by adding measurement error having variance 10; resulting reliabilities for the scores at times {1, 3, 5} are {.84, .84,

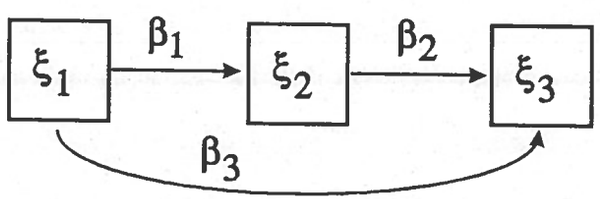

FIG. 1.8. Representation of three-wave structural regression model.

.90}. The path analysis regression for $X(5)$ for these 40 cases produces the fit: $X(5) = 5.054 - .1212\,X(1) + 1.19\,X(3)$ with squared multiple correlation .552. Also in Rogosa (1993) are results and illustrations of similar failures for the structural regression model approach when the underlying individual growth curves are not simple straight line, but exponential growth to an asymptote (as in Equation 1.2).

Exogenous Variable and Change: Two Waves

To illustrate the second part of Myth 7, we use the data example from Exhibit 1 to illustrate the misleading consequences of basing an analysis of the standard structural model shown in Figure 1.4—two waves with an exogenous variable. In the population from which the example data are drawn there is *no* association between the background variable W and individual rate of change θ; $\rho_{W\theta} = 0$. Table 1.11 gives numerical theoretical results that the structural (causal) regression coefficients may be large positive or large negative even when $\rho_{W\theta} = 0$. The results of the structural regression using the example data set are shown below. When $\xi(3)$ is used as the initial value the structural coefficient for the influence of W on change is significant with a negative value and when $\xi(1)$ is used as the initial value the structural coefficient for the influence of W on change is significant with a positive value.

1. ξ(3) as the Initial Value

```
The regression equation is
ξ(5) = 0.68 − 0.757 W + 1.38 ξ(3)

Predictor       Coef     Stdev     t-ratio      p
constant       0.683     4.555        0.15  0.882
W           −0.7570    0.3329       −2.27  0.029
ξ(3)         1.3822    0.1290       10.72  0.000

s = 3.752     R-sq = 84.4%     R-sq(adj) = 83.5%
```

2. ξ(1) as the Initial Value

```
The regression equation is
ξ(5) = 31.2 + 1.50 W + 0.239 ξ(1)

Predictor       Coef     Stdev     t-ratio      p
Constant      31.213     7.546        4.14  0.000
W             1.5004    0.6678        2.25  0.031
ξ(1)          0.2392    0.2587        0.92  0.361
```

2. WHAT ARE USEFUL DATA ANALYSIS APPROACHES?

Data analysis strategies and methods follow directly from the modeling approach that is the basis for the original Myths chapter. The unifying theme is that all questions can be addressed by models and methods that start with the individual unit trajectories. Thus, useful methods for the analysis of longitudinal data take as the starting point a model for the individual history.

The simplest instance of this type of model for a quantitative outcome is a straight-line growth curve for each individual. Fitting such a model to the individual's data points can be thought of as using the model to smooth the data in order to derive an attribute for the individual, such as the rate of improvement or decline in that measure. The power of this approach is the straightforward way in which such analyses can be built up for complex settings (e.g., comparing groups, assessments of stability, and so forth) without losing firm contact with the data. The examples given here illustrate analyses directed toward the first three questions described in the previous listing of research questions: (a) individual and group growth: description and estimation of the form and amount of change; (b) correlates and predictors of change: systematic individual differences in growth such as the question "What kind of persons learn (grow) fastest?"; (c) stability over time: consistency over time of individuals and individual differences. (Note that here we are limited to questions about quantitative outcomes, such as functional abilities or blood pressure.)

In this exposition, sketches of analyses are presented for two examples of actual longitudinal panel data: (1) the Ramus data (briefly treated as part of Myth 8), which consist of four longitudinal observations on each of 20 individuals with no exogenous measure and (2) the North Carolina data, which consist of eight longitudinal observations on each of 277 individuals and with an exogenous ability measure. These analyses methods typically work well for four or more longitudinal observations on each individual (although some missing data can be accommodated). Three longitudinal observations is an absolute minimum for the statistical procedures. In the subsequent question (Question 3: "What can be done with meager time-1, time-2 data?"), the situation of just two longitudinal observations is discussed; artificial time-1, time-2 data is used to illustrate the limited, but useful, descriptive analyses that can be supported when just two observations are available (the traditional measurement of change pre-test, post-test setting). In addition, these methods are contrasted with the (misleading) traditional measurement of change analysis procedures that are based on between-variable relations (see also Myth 6).

Data Structures

The simplest structure of the longitudinal panel data is illustrated by the display of the first four cases from the North Carolina achievement data (Williamson,

Applebaum, & Epanchin, 1991), shown below (total 277 females). Each individual has a row of data; the first column contains the verbal ability score, which is used as the exogenous background measure, W. The multiple longitudinal observations follow: eight waves of achievement test scores in math (grades 1–8).

W	T→	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0
120		380	377	460	472	495	566	637	628
95		362	382	392	475	475	543	601	576
99		387	405	438	418	484	533	570	589
101		342	368	408	422	470	543	493	589

Model and Parameters of Interest

Form of the Individual Growth Curve

The simplest model, which serves as a basis for these data analysis examples, is the straight-line growth model,

$$\xi_p(t) = \xi_p(0) + \theta_p t,$$

where $\xi_p(t)$ is the true score of person p at time t and θ_p is the constant rate of change for person p. The straight-line growth model is useful for heuristic reasons because of its simplicity, as it yields a simple index for individual rate of progress. In addition, in applications, straight-line growth serves as a useful approximation to actual growth processes (see Hui & Berger 1983, p. 753; Rogosa & Willett 1985b, p. 205). Moreover, when observations at only a few time points are available (e.g., $T = 4$) the data may justify the estimation of nothing more complicated than a constant-rate-of-change model. Although many uses of straight-line growth curves can be justified, nonlinear growth functions may be crucially important in many applications, and methods for straight-line growth curves should be thought of as a first approximation toward the use of more complex growth models.

Individual Change

Thus, estimates of θ_p provide a simple index for individual rate of learning. The parameter θ_p is closely related to the amount of true change; for example, in two-wave (or pre-post) data, true change is equal to $\theta_p(t_2 - t_1)$. Growth curves for different individuals have different values of rate of change θ_p and level $\xi_p(0)$. When describing the learning of a group of individuals, the distribution, over individuals, of empirical rates of learning is informative. The first two moments of the rate of change are μ_θ and σ^2_θ. Similarly, we may want to describe the variability in level of performance at each time, $\xi_p(t)$. As it turns out, the variance of $\xi_p(t)$ has a functional dependence on time, and investigation of the form of this function leads naturally to the definition of a "centering" point and a scaling factor associated with the time scale. These have been denoted t°

and κ, respectively. (See Rogosa and Willett, 1985b.) Both t° and κ are properties of the particular collection of straight-line growth curves.

Correlation Between Change and Initial Status

Another quantity of central importance is the correlation between change, θ, and initial status, $\xi(t_1)$, where t_1 indicates initial time of measurement, and this was discussed in Myth 4. The correlation is used to investigate whether those with lowest initial status make the most progress (negative value) or those with the highest initial status make the most progress (positive value). As discussed in Rogosa and Willett (1985b), the choice of t_1 is of critical importance because $\rho_{\xi(t_1)\theta}$ is functionally dependent on time. (The definitions of t° and κ also arise naturally from an investigation of this dependence; see Rogosa and Willett, 1985b.)

Consistency of Individual Differences

As discussed in Myth 8, the index γ was proposed by Foulkes and Davis (1981) as an index of tracking, and is defined as the probability that two randomly chosen growth curves do not intersect. High values of γ indicate high consistency of individual differences over time. Another way of interpreting γ is to note that high values of γ denote "the maintenance over time of relative ranking within the response distribution" (Foulkes & Davis, 1981, p. 439). Thus γ indicates the stability of individual differences. If a collection of individual growth curves have a high estimated value of γ, that indicates that individuals that started out relatively high maintain that advantage and individuals starting out low retain that disadvantage (regardless of the overall growth rate).

Systematic Individual Differences

To address additional research questions about systematic individual differences in growth (i.e., correlates of change) longitudinal data sets often include measurements on one or more exogenous characteristic which are denoted here by W (e.g., home environment). This terminology derives from the structure of the inquiry that: "Individual differences in growth exist when different individuals have different values of θ_p. Systematic individual differences in growth exist when individual differences in a growth parameter such as θ_p can be linked with one or more W's" (Rogosa and Willett 1985b p. 205). A model for individual differences in growth is needed for investigating systematic individual differences in growth. For the purpose of this exposition, W is regarded as measured without error; in practice, with fallible measurement interest would normally be on relations between θ_p and the true score underlying W. The relation of W to the slope parameter, summarized by the conditional expectation $E(\theta \mid W)$, is stated here as the simplest possible straight-line regression.

$$E(\theta \mid W) = \mu_\theta + \beta_{\theta W}(W - \mu_W).$$

This equation for $E(\theta \mid W)$ is an example of a "between-unit" model. A similar relation can be stated for the intercept in the equation for $\xi_p(t)$. In the case where there is no measured exogenous variable, this between-unit model is $E(\theta) = \mu_\theta$.

A common procedure in the literature is to correlate the value of the background demographic variable or curricular variable with performance at a given time. That is, the cross-sectional correlation is computed, sometimes for every occasion in time. For example, with a background variable, W, correlations of the test score with W at various grades would be computed, and from these correlations conclusions about learning are attempted. Rogosa and Willett (1985b) have shown that such cross-sectional correlations are not useful for this purpose. To illustrate, consider a situation where the correlation between true rate of change and the background variable is zero. Then the correlation between the true test score, $\xi(t)$, and the demographic variable, W, at any one slice in time could be big or small. Consequently, this correlation really doesn't inform about systematic individual differences in learning. The reverse is true also. Consider a demographic variable for which this correlation is large. Regardless, the correlation between the background variable and a test score at a specific time can be positive, zero, or negative depending upon the time chosen for the cross-sectional correlation. Obviously, no useful conclusions about learning can be drawn from the cross-sectional correlations.

Data Analysis and Parameter Estimation

Since 1981, I have used various versions of a program we call TIMEPATH (originally developed with the assistance of John Willett and Gary Williamson, current version written with Ghassan Ghandour) for the analysis of quantitative longitudinal panel data. In this program, ordinary least-squares regression is used to estimate the growth curve model from the longitudinal data for each individual. As the empirical rate of change can be treated as an attribute of an individual (just like a measurement on X or W), the obtained slopes for each individual regression can be profitably used for various descriptive analyses. Such descriptive analyses may, in many situations, be more important and informative than the formal parameter estimation.

To estimate many of the parameters discussed above, maximum likelihood estimates derived from the results in Blomqvist (1977) are used. In the current program (Rogosa & Ghandour, 1989), standard errors for these parameter estimates and confidence intervals for the parameters are obtained by bootstrap resampling methods in which rows (individual units) are resampled. In Tables 1.17 and 1.18, the reported standard errors are just the standard deviation over 4,000 bootstrap replications, and the endpoints of the reported 90% confidence intervals are just the 5% and 95% values of the empirical distributions from the resampling (i.e., the 200th values from the maximum and minimum values). More sophisticated and more accurate confidence intervals could be constructed using the methods in Efron and Tibshirani (1993), but these simpler intervals

were chosen for the purposes of this exposition. In Rogosa and Saner (in press), equivalences obtained from the application of newer computational programs based on hierarchical linear model methodology (especially the HLM program of Bryk & Raudenbush; Bryk & Raudenbush, 1987) for these data are illustrated and some shortcomings discussed. More technical detail on estimation can be found in Rogosa and Saner (in press). When present, missing longitudinal observations are treated (deliberately) in a very simple manner—the individual growth curves are fit to the data that are present, and the overall SSE from the individual fits is just weighted according to the observations present.

Descriptive Analyses of Growth Rates

The most basic step in the analysis is the fitting of a straight-line growth curve (the regression of X on t for each p) by ordinary least squares. The estimates of slope, squared multiple correlation, and other properties of the straight-line fit including diagnostics can be displayed and summarized (see for example the output in Table 1.16). Estimation of the straight-line growth model allows comparisons of rates of change across individuals. Stem-and-leaf diagrams, boxplots, and the five-number summaries of the empirical rates are useful ways to describe both typical rates of learning and the heterogeneity across individuals (see, for example, Figure 1.9). Using the estimated $\hat{\theta}_p$ values for each individual, plots representing relations of change with initial status (see Figure 1.10) and relations with the exogenous measure W (see Figure 1.12) are especially useful for diagnostic examination of the corresponding correlation parameter estimates.

Parameter Estimates

Tables 1.17 and 1.18 present a collection of parameter estimates based on the growth curve model. As a first step, parameters of interest are "typical" rates of change μ_θ or median(θ), and a measure of heterogeneity σ^2_θ, the variance of the θ_p. Point and interval estimates are provided by the bootstrap resampling. The estimate of the reliability $\rho(\hat{\theta})$ is simply the estimate of σ^2_θ divided by the observed variance of the $\hat{\theta}_p$. Our statistical procedures provide a maximum likelihood estimate of the correlation between true rate of change and true initial status $\rho_{\xi(t_1)\theta}$; the data examples show one negative value and one positive value for this correlation. A good estimate of this correlation is made possible by the availability of multiple (e.g., four or more) longitudinal observations; a pervasive problem in the pre-test, post-test dominated measurement of change literature was that when only two observations were available, the only estimate was the correlation between observed change and observed initial status which may have large, usually negative, bias (see Rogosa et al., 1982). Systematic individual differences in growth are indicated in these analyses by the quantity $\rho_{\theta W}$ or by $\beta_{\theta W}$; for example, nonzero values of $\beta_{\theta W}$ indicate that W is a predictor of growth. Maximum likelihood estimates of these parameters are simply obtained by disattenuating the observed relations by use of the estimate of $\rho(\hat{\theta})$; point estimates along with bootstrap standard error and confidence intervals are given for the North Carolina data in Table 1.18, which shows strong relations of θ with the verbal ability measure.

For the consistency of individual differences, as discussed in Myth 8, the index γ was proposed by Foulkes and Davis (1981) as an index of tracking, and is defined as the probability that two randomly chosen growth curves do not intersect. High values of γ indicate high consistency of individual differences over time. The probability of no intersection is estimated from a count of the number of intersections that each individual trajectory has with the other individuals; for each individual $\hat{\gamma}_p$ is one minus the number of intersections over $n - 1$. Individuals with a low value of $\hat{\gamma}_p$ are those whose relative standing changes considerably over the time period. The proportion of no intersections is accumulated over the n individuals to produce an overall $\hat{\gamma}$ estimate. The standard error can be obtained from a jackknife approximation given by Foulkes and Davis (1981) or by using bootstrap resampling. The value of $\hat{\gamma}$ in both Tables 1.17 and 1.18 indicate reasonably strong tracking.

Ramus Data

The first data example consists of four longitudinal observations on each of 20 cases. The measurement is the height of the mandibular ramus bone (in mm) for boys each measured at 8, 8.5, 9, 9.5 years of age. These data, which have been used by a number of authors, can be found in Table 4.1 of Goldstein (1979a). These data are small enough that it is convenient to present extended output. Fitting a straight line to each individual's observations yields output from the TIMEPATH program that is shown in part in Table 1.16. In Table 1.16, the columns are the `ID` number for the case, `Rate` the estimated rate of change (slope of the straight-line growth curve), `R_sqr` the squared multiple correlation for the straight-line fit, `D_Rsq` the increase in squared multiple correlation resulting from fitting a quadratic form (useful for detecting cases with large curvature), and the final columns contain the longitudinal observations.

This output provides the raw information and is a very first step in describing individual change. It can be seen that these individual histories correspond closely to the straight-line model by examining the individual R^2 (or the corresponding mean-square residuals) from each fit. For these data the median R^2 is .95, with only two of the 20 values below .91. For the individual rates of change both numerical summaries such as that below and graphical descriptive summaries as in Figure 1.9 are useful:

Rate of Change

	N	MEAN	MEDIAN	STDEV	MIN	MAX	Q1	Q3
$\hat{\theta}$	20	1.866	1.500	1.165	0.460	4.960	1.205	2.010

TABLE 1.16
TIMEPATH Individual Fit Output for Ramus Data

ID	*Rate*	*R_sqr*	*D_Rsq*	*T→*	*8.00*	*8.50*	*9.00*	*9.50*
1	1.180	94.2	1.2		47.80	48.80	49.00	49.70
2	1.280	97.9	.6		46.40	47.30	47.70	48.40
3	1.520	98.6	.3		46.30	46.80	47.80	48.50
4	1.420	92.4	7.4		45.10	45.30	46.10	47.20
5	1.100	95.3	3.9		47.60	48.50	48.90	49.30
6	.740	91.5	3.1		52.50	53.20	53.30	53.70
7	2.240	90.5	9.2		51.20	53.00	54.30	54.50
8	1.800	74.2	22.2		49.80	50.00	50.30	52.70
9	4.080	98.8	.4		48.10	50.80	52.30	54.40
10	2.040	90.6	4.3		45.00	47.00	47.30	48.30
11	.460	99.0	.8		51.20	51.40	51.60	51.90
12	4.960	94.5	2.5		48.50	49.20	53.00	55.50
13	1.920	98.0	1.9		52.10	52.80	53.70	55.00
14	1.040	98.8	.6		48.20	48.90	49.30	49.80
15	1.480	99.6	.4		49.60	50.40	51.20	51.80
16	1.760	98.8	1.0		50.70	51.70	52.70	53.30
17	1.520	96.9	3.0		47.20	47.70	48.40	49.50
18	1.300	87.1	12.4		53.30	54.60	55.10	55.30
19	1.440	90.9	8.8		46.20	47.50	48.10	48.40
20	4.040	92.2	.7		46.30	47.60	51.30	51.80

The display of the individual rates of change in Figure 1.9 shows three individuals "improving" considerably faster than the others. The most complete descriptive view is given by a plot of the fitted growth curves which is shown in Figure 1.6. That plot is used to illustrate the high stability of individual differences among these individuals (below estimate of γ is .826). The observed correlation between X_1 and $\hat{\theta}$ is -0.188; the corresponding scatterplot is given in Figure 1.10.

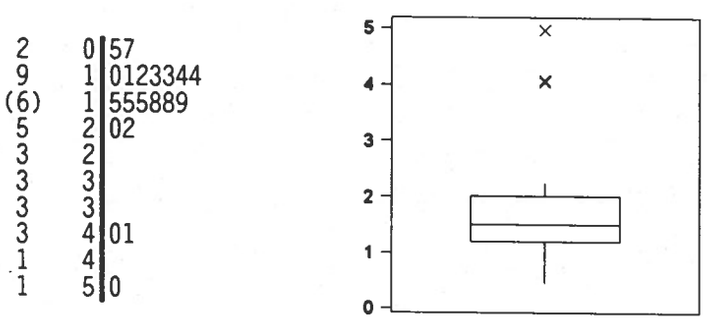

FIG. 1.9. Graphical description of individual growth rates for ramus data: (a) stem-and-leaf display; (b) boxplot.

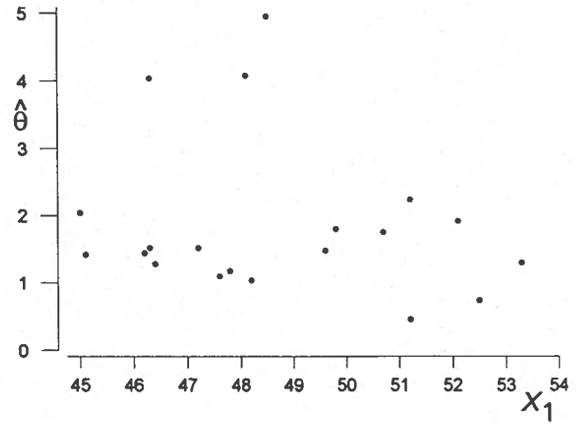

FIG. 1.10. Scatterplot of empirical rate vs. observed initial status for ramus data.

Estimation and inferences for model parameters obtained from the maximum-likelihood estimation and bootstrap resampling procedures are shown in Table 1.17. Of note is the high estimated reliability of the rate of change; the standard error of measurement for an individual rate is .39. But even with considerable accuracy in assessing individual change, with only a small number of cases the between-person moments (variance components, correlations) have considerable uncertainty as can be seen from the rather wide confidence intervals.

Other quantities that can be estimated from the growth curve modeling include the reliabilities of the observed measures at each of the times of observation. For these ramus data the reliability estimates are: {.970, .969, .971, .975}. Bootstrap standard errors for these estimates are between .01 and .015.

TABLE 1.17
Parameter and Variance Component Estimation for Ramus Data

Estimate of	*Point*	*s.e.*	*90% CI*
Median(θ)	1.48	.14	(1.30, 1.80)
μ_θ	1.85	.252	(1.474, 2.298)
σ^2_θ	1.203	.500	(.336, 1.971)
$\rho(\hat{\theta})$	.886	.086	(.725, 928)
$\rho_{\theta\xi(8)}$	−.196	.168	(−.439, .098)
γ	.826	.065	(.668, .879)

Path Analysis Controls

The analyses briefly described above do provide some information about change. Contrast those results with a standard path analysis of these 4-wave data using standard multiple regression methods shown in Figure 1.11. Using all plausible "causal paths", values of the path coefficients are shown; only the coefficients for the lag-1 paths are statistically significant. (Refitting using just the significant paths changes little.) The real question is, What in the world does this analysis reveal about change (or any conceivable longitudinal research question)?

North Carolina Data

The second data analysis example is real education data previously analyzed using the maximum-likelihood estimates in TIMEPATH in an excellent expository paper by Williamson et al. (1991). Descriptions of the individual trajectories and rates of change would use the same displays as in the Ramus data. Again, these data conform well to the straight-line growth model; the median value of R^2 for the 277 individual fits is .963. A brief numerical description of the individual rates of change is

Rate of Change

	N	MEAN	MEDIAN	STDEV	MIN	MAX	Q1	Q3
$\hat{\theta}$	277	36.45	36.39	7.472	9.71	64.24	31.46	41.02

One reason to examine this data example is the existence of the exogenous variable, the verbal ability measure, which allows questions about correlates of change to be addressed. The initial descriptive information would be the correlation between $\hat{\theta}$ and W, which is .624, and the corresponding scatterplot is shown in Figure 1.12.

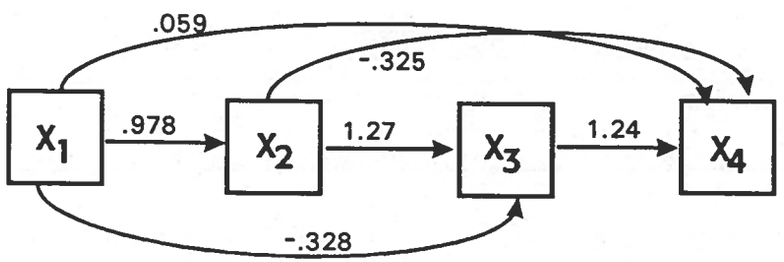

FIG. 1.11. Four-wave path analysis results for ramus data.

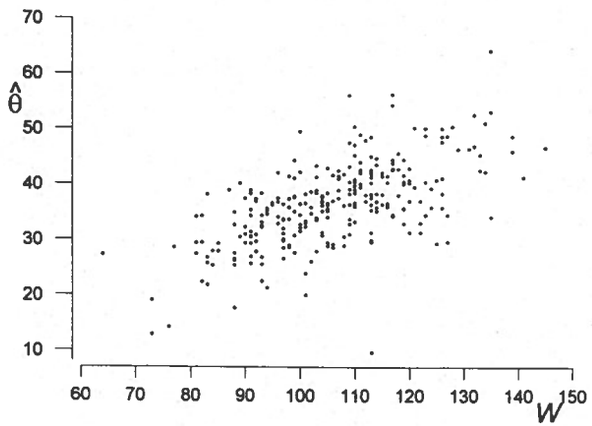

FIG. 1.12. Scatterplot of observed rate of change vs. background ability measure for North Carolina data.

The parameter estimation summarized in Table 1.18 reveals that these data permit rather accurate assessment of rates of change; the reliability estimate is high, and the standard error of measurement of $\hat{\theta}$ is 3.1. Of considerable note in these data is the large positive value of .651 for the estimate of the correlation between true rate of change and true initial status. Also of note is that the estimates show a strong relation with W, the verbal ability measure. (Entries in Table 1.18 correspond to point estimates reported in Table 3 of Williamson et al., 1991). Note that with the larger sample of 277 cases, greater accuracy (smaller standard errors, narrower confidence intervals) for the estimation of the between-person moments is obtained.

TABLE 1.18
Parameter and Variance Component Estimation for North Carolina Data

Estimate of	*Point*	*s.e.*	*90% CI*
μ_θ	36.45	.448	(35.72, 37.19)
Median(θ)	36.39	.327	(35.86, 36.95)
σ^2_θ	46.23	5.95	(36.67, 56.12)
$\rho(\hat{\theta})$	.828	.019	(.792, .854)
$\rho_{\theta\xi(1)}$	.651	.090	(.513, .809)
γ	.721	.0174	(.689, .746)
$\beta_{\theta w}$	.336	.028	(.291, .382)
$\rho_{\theta w}$	.686	.045	(.609, .754)

3. WHAT CAN BE DONE WITH (MEAGER) TIME-1, TIME-2 DATA?

The data used for illustration here consist of observations on 40 cases, and shown below are the values for observations at time 1, X_1, at time 2, X_2, and also values for an exogenous background variable W. For purposes of exposition the longitudinal observations might be reading achievement scores of elementary school children, and the background variable W might be some measure of home environment, leading to obvious substantive questions such as, Do students with "better" home environments (books in the home, etc.) make better progress or improvement in reading?

Some Descriptive Analyses of Individual Change with Two Observations

The estimate of the amount of change for each individual is simply the observed amount of improvement: $D = X_2 - X_1$. The display of the data given in Table 1.19 has this difference score appended in the first column. Descriptive analyses of D, such as those illustrated below have value, and are essentially the best one

TABLE 1.19
Data and Difference Scores for Two-Wave Example

Case	*D*	X_1	X_2	*W*	*Case*	*D*	X_1	X_2	*W*
1	21.93	37.52	59.45	15.97	21	30.90	45.57	76.47	15.88
2	16.52	45.13	61.65	15.38	22	31.35	50.79	82.14	18.25
3	31.00	35.15	66.15	11.48	23	4.96	36.56	41.52	10.15
4	20.44	44.13	64.57	16.89	24	18.02	39.48	57.50	9.46
5	17.75	52.74	70.49	19.18	25	26.36	38.34	64.69	15.81
6	33.86	30.43	64.29	11.82	26	21.72	32.57	54.29	11.60
7	22.18	45.86	68.04	15.33	27	39.74	44.18	83.92	14.08
8	14.95	41.09	56.04	13.21	28	24.97	32.79	57.76	12.19
9	10.80	45.60	56.39	13.09	29	24.91	38.61	63.52	14.07
10	11.79	41.64	53.43	10.32	30	4.46	54.90	59.36	16.68
11	2.12	40.55	42.67	10.26	31	16.79	37.42	54.22	15.07
12	17.71	43.60	61.30	15.60	32	11.51	43.19	54.71	13.94
13	16.49	40.33	56.82	13.90	33	18.79	57.07	75.86	20.40
14	19.51	36.47	55.98	13.53	34	24.57	52.40	76.97	16.00
15	22.33	50.94	73.27	14.45	35	6.74	53.35	60.09	17.47
16	10.08	56.39	66.47	20.16	36	16.56	47.21	63.77	17.30
17	24.16	54.82	78.98	16.11	37	42.52	37.53	80.05	15.86
18	22.97	46.23	69.21	15.06	38	21.08	47.89	68.97	18.95
19	39.50	40.34	79.84	18.33	39	23.16	58.79	81.95	18.90
20	21.81	39.78	61.59	13.77	40	2.69	39.98	42.67	13.79

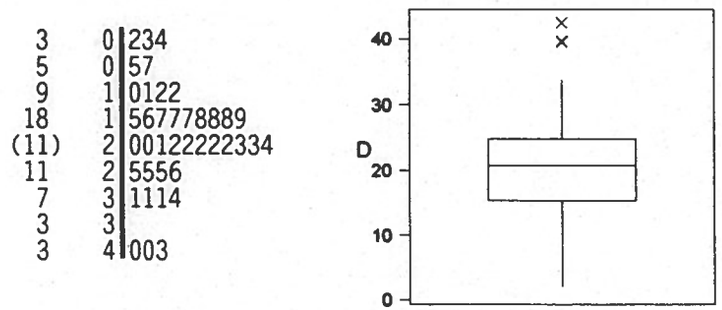

FIG. 1.13. Graphical description of individual difference scores for two-wave example data: (a) stem-and-leaf display; (b) boxplot.

can do with these limited longitudinal data. If the individual history is merely two observations, the difference score is essentially (with unit time) the slope of the straight-line growth curve, and fitting a line to two points can yield disappointing statistical or psychometric properties. But the core problem is the meager data, not the summary. These limitations and statistical difficulties do preclude the estimation of variance components, etc., as was done in the two previous examples. Nonetheless, we can at least do simple descriptive summaries that address some of the longitudinal research questions. The data analysis setting is certainly not as hopeless as might be concluded from the (deliberate) overstatement in Motto 1 of Rogosa et al. (1982): "Two-waves are better than one, but not much."

Analyses of Individual Change

Below we have both a quantitative summary of the observed data and the amount of change; graphical summaries of the individual change are shown in Figure 1.13. For the "average" individual there was notable improvement of about 20 points. But clearly there also appear to be large individual differences in change, with some individuals gaining more than 40 points and others less than 5 points.

Observed Data

	N	MEAN	MEDIAN	STDEV	MIN	MAX	Q1	Q3
X_1	40	43.93	43.40	7.29	30.43	58.79	38.40	50.07
X_2	40	64.18	63.65	10.92	41.52	83.92	56.50	72.57
W	40	14.992	15.200	2.803	9.462	20.399	13.288	16.837

Change

	N	MEAN	MEDIAN	STDEV	MIN	MAX	Q1	Q3
D	40	20.25	20.76	9.71	2.12	42.52	15.34	24.82

One question that follows is whether these individual differences are linked to quantities such as initial status or the exogenous variable. The correlation between observed change and observed initial status (correlation between D and X_1) is $-.199$; this estimate is biased with a somewhat complicated form (as discussed in, for example, Rogosa, et al. 1982, Equation 11). The corresponding scatterplot is shown in Figure 1.14.

To describe relations between change and exogenous variables, start with the scatterplot shown in Figure 1.15. The sample correlation between W and D is 0.158, and the corresponding regression analysis for predicting individual change from W yields:

```
The regression equation is D = 12.0 + 0.549 W
Predictor      Coef    Stdev   t-ratio      p
Constant     12.015    8.460      1.42  0.164
W            0.5488   0.5549      0.99  0.329
s = 9.713      R-sq = 2.5%     R-sq(adj) = 0.0%
```

Apparently, there is little or no relation between individual change and the background variable W. Observed correlation is near zero, with test statistic of .99.

More generally, even with the meager two-wave data, valuable descriptions based on individual change can be built up to address more complex research questions and settings. For example, in two-group (or more) experimental studies, comparison of the difference scores across the experimental groups is equivalent to repeated measures analysis of variance. A thorough exposition of this equivalence to repeated measures ANOVA for the two-group, pre-test, post-test

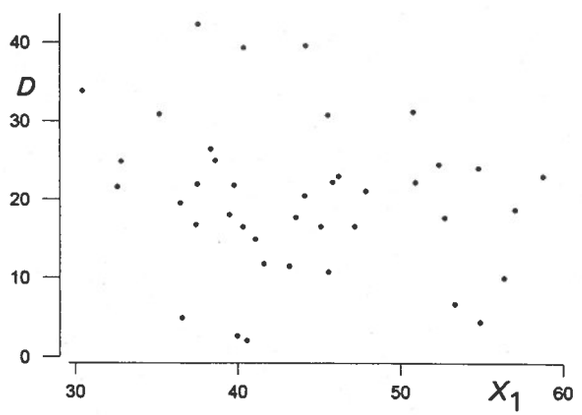

FIG. 1.14. Scatterplot of observed change vs. observed initial status for two-wave example data.

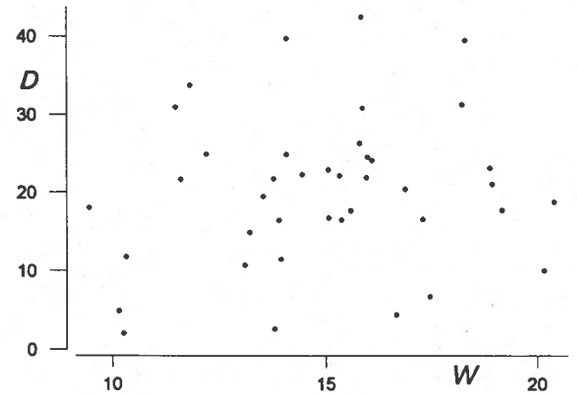

FIG. 1.15. Scatterplot of observed change vs. background variable for two-wave example data.

data structure is given in Brogan and Kutner (1980). Similar strategies are useful in nonexperimental, comparative settings. For quasiexperiments, a question that can and should be addressed in these nonexperimental studies is, Which group changed more (or declined less)? Of course, causal attribution (e.g., to the program or intervention) cannot be made, and, most important, attempts at statistical adjustments, explicit or implicit, to draw an "as if by experiment" conclusion are doomed. The difference score (or a measure of change obtained from richer data) is not the core problem. Approaches that start with assessments of individual change are far better than standard ANCOVA methods, or more esoteric adjustment procedures such as standardized change scores and the like.

Artificial Data with Known Structure

These exemplar time-1, time-2 data were produced from an underlying structure with known parameter values. The 40 individual cases were drawn from a population with specified characteristics (methods for constructing such artificial data are described in Question 5). True change, Δ, has a Uniform distribution with lower and upper values of 4 and 36; thus true change has mean 20 and variance 85.333. True status at time 1, ξ_1, has population mean 45 and variance 53.33 and for time 2 ξ_2 has population mean 65 and variance 96.0. Measurement error has mean 0 and variance 10 at each time point. Consequently, in the population the reliability of change, D, is .810, and the reliabilities of X_1 and of X_2 are .842 and .906, respectively. The correlation between true change and true initial status is $-.316$. The exogenous variable W was constructed to have no association with true change; $\rho_{\Delta W} = 0$.

Comparisons with Traditional Measurement of Change Analyses

The starting point for traditional analyses in the past measurement of change literature is not the description of individual change, but instead, the description of between-variable relations, most notably the time-2, time-1 scatterplot. The time-2 versus time-1 (X_2 versus X_1) scatterplot shown in Figure 1.16 has correlation .491. Also examined would be the between-variable relations with the exogenous variable in Figure 1.16, which show noticeable association between W and the longitudinal observations. However, Rogosa and Willett (1985b) have demonstrated that such cross-sectional associations do not provide useful information about correlates of change or progress. The correlation matrix provides the usual summary of these between-variable relations:

	X_1	X_2
X_1		
X_2	0.491	
W	0.716	0.619

It is unclear what is revealed about change from these between-variable relations.

To investigate the importance of the exogenous variable, it would be typical to carry out a regression analysis predicting X_2 from X_1 and W. The results from this regression, shown below, produce a highly significant coefficient for W (t-value 2.98) and thus would lead to exactly the wrong conclusion about W—even though in the structure of the data there is zero correlation between individual

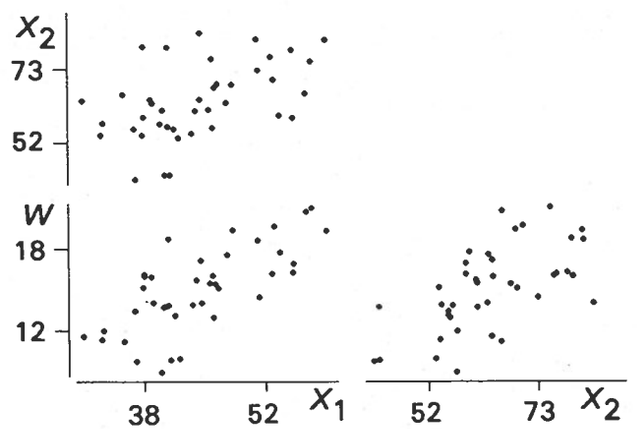

FIG. 1.16. Traditional scatterplots for two-wave example data arranged as time 2 vs. time 1 (top) and background variable vs. time 1 and time 2 (bottom).

change and W, the measurement of change analysis would flag W as an important predictor of change!

The regression equation is $X_2 = 25.7 + 0.145\, X_1 + 2.14\, W$

Predictor	Coef	Stdev	t-ratio	p
Constant	25.690	8.821	2.91	0.006
X_1	0.1447	0.2759	0.52	0.603
W	2.1431	0.7181	2.98	0.005

s = 8.770 R-sq = 38.8% R-sq(adj) = 35.5%

Similarly, construction of the standard residual change score (see Myth 6) yields a correlation with W of .307 (double the sample correlation between W and D).

4. IS REGRESSION TOWARD THE MEAN REALLY AN IMPEDIMENT TO ASSESSING CHANGE?

The answer is no, but the material in Myth 5 would benefit from some augmentation. In that and other treatments of regression toward the mean, my focus has been on the importance of clear, explicit definition in the common references to "regression toward the mean" and then showing the consequences of that definition. My previous discussion has followed the literature in examining time-1, time-2 regression toward the mean, either in terms of perfectly measured $\xi(t_i)$ or in terms of the fallible X_i. And while those facts (reviewed below) are useful, the important additional message is that discussion of time-1, time-2 regression toward the mean, to some extent, misses the point—interest in assessing change should be on the estimate of change, such as the estimate of the amount of change Δ_p or of the rate of change θ_p.

The conventional definition of time-1, time-2 regression toward the mean in standard deviation units is uninteresting, whether it be in terms of X_1 and X_2 or in terms of $\xi(t_1)$ and $\xi(t_2)$, the latter in Equation 1.3. As is shown there, the condition for regression toward the mean to hold is for the time-1, time-2 correlation to be less than 1. This tautology (discussed following Equation 1.3) has the following derivation: starting with Equation 1.3, substitute $E[\xi(t_2) \mid \xi(t_1) = C] = \mu_{\xi(t_2)} + \rho_{\xi(t_1)\xi(t_2)} [\sigma_{\xi(t_2)}/\sigma_{\xi(t_1)}] (C - \mu_{\xi(t_1)})$ and then simplify to obtain the condition $\rho_{\xi(t_1)\xi(t_2)} < 1$.

One illustration that this statement of regression toward the mean is not important is provided by the sample collection of 15 growth curves depicted in Figure 1.17. This set of individual trajectories has values that satisfy the definition of time-1, time-2 regression toward the mean. Using, say, $t_1 = 3$, $t_2 = 7$, the population correlation $\rho_{\xi(3)\xi(7)} = .894$, and thus Equation 1.3 is satisfied—regression toward the mean "holds." Yet the correlation between $\xi(3)$ and the amount (or rate) of change is .707, which implies that rates of improvement for

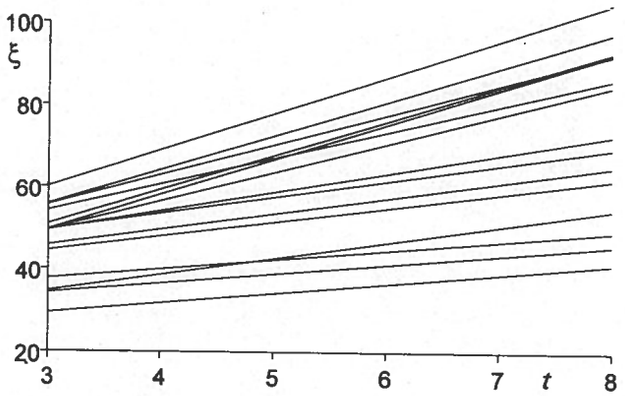

FIG. 1.17. Collection of 15 straight-line growth trajectories for regression toward the mean illustration.

those high on $\xi(3)$ are larger than the rates of improvement for those low on $\xi(3)$. Intuitively, this is just the opposite of what is meant by the phrase regression toward the mean.

For illustration, values for $\xi(3)$, $\xi(7)$ and the corresponding fallible observations, the X-values are shown in Table 1.20. The population means at times 3 and 7 are 40 and 60, respectively. (The X-values are constructed by adding measurement error with variance 25 to the ξ-values yielding reliabilities of .631 for $t = 3$ and .895 for $t = 7$.) For the ξ-values, 13 of the 15 cases are further from the mean at $t = 7$ than at $t = 3$; only cases 7 and 12 are closer to the mean at $t = 7$ than at $t = 3$. For example, individual 3 is 2.4 units from the mean at $t = 3$ and 10.5 units away at $t = 7$.

TABLE 1.20
True and Fallible Two-Wave Data for Regression Toward the Mean Example

	ξ-values		*X-values*	
	$t = 3$	$t = 7$	$t = 3$	$t = 7$
1	41.93	57.27	40.11	54.88
2	27.02	36.49	29.05	40.23
3	42.37	70.52	41.15	73.72
4	47.82	73.36	52.09	72.19
5	41.02	54.89	41.97	62.37
6	51.06	86.68	50.67	82.19
7	44.81	63.33	47.36	69.74
8	47.38	80.69	48.70	83.13
9	30.68	46.52	35.04	51.35
10	31.93	41.15	31.24	49.51
11	42.53	76.04	51.05	69.02
12	45.50	61.40	51.97	53.52
13	35.43	44.59	29.40	31.12
14	40.91	75.20	40.05	78.89
15	48.28	77.50	56.36	77.66

The more appropriate statement of time-1, time-2 regression toward the mean in terms of the actual metric is given in Equation 1.4 and is equivalent to $\rho_{\xi(t_1)\theta} < 0$ or $\beta_{\xi(t_2)\xi(t_1)} < 1$. This condition is simply a statement about the collection of individual growth curves and has no relevance to the ability to assess change.

Regression in Terms of the Estimate of Change

Consideration of time-1, time-2 regression toward the mean had led me previously to believe that regression toward the mean was irrelevant to assessing change and was a concept/concern best ignored and forgotten. But that's not fully correct. The important point that is missed in presenting the facts above is that the quantity of interest is the estimate of change: of the amount of change Δ or the rate of change θ. This is where there is some constructive role for talking in terms of a regression toward the mean. The time-2 observation is not a replication of the time-1 observation whenever there is systematic individual change. Where regression toward the mean is a relevant concept is that an estimate of Δ or the rate of change θ from fallible data has more variability than if measurement were perfect; that is variance$(\hat{\theta})$ > variance(θ). Thus an extreme value of $\hat{\theta}_p$ is likely to be larger than what would be obtained if measurement were perfect. This concern is a good justification for empirical Bayes estimates of individual growth rates (e.g., Fearn, 1975). For this concern about the error in $\hat{\theta}_p$, the traditional and sensible warning about selecting extremes based on fallible scores still pertains—i.e., selecting students with observed low rates of change for an intervention runs the risk of overstating the value of the intervention as their actual rates of improvement are likely greater than that observed.

5. HOW CAN LONGITUDINAL DATA EXAMPLES WITH KNOWN STRUCTURE BE CREATED?

To create examples of longitudinal panel data with known structure, we can use the basic relations and properties of collections of growth curves. The procedures illustrated here are for data based on underlying straight-line growth curves.

Simulation Procedure

Start by choosing the center for the time metric by specifying t° where $t^\circ = -\sigma_{\xi(0)\theta}/\sigma^2_\theta$. Then for the parameters of the straight-line growth model $\xi_p(t) = \xi_p(t^\circ) + \theta_p(t - t^\circ)$ specify the parameter distributions over individuals of the

uncorrelated random variables $\xi(t^\circ)$ and θ (e.g., each distribution Gaussian or each distribution Uniform) to generate these parameter values for each p. By doing so, the scale for the time metric $\kappa = \sigma_{\xi(t^\circ)}/\sigma_\theta$ is specified. By then stating the discrete values of the times of observation $\{t_i\} = t_1, \ldots, t_T$, we then have values for the $\xi_p(t_i)$ for $p = 1, \ldots, n$. The exogenous characteristic W is generated with specified mean and variance, specifying the two correlations $\rho_{W\xi(t^\circ)}$ and $\rho_{W\theta}$ (under the constraint $(\rho_{W\xi(t^\circ)})^2 + (\rho_{W\theta})^2 \leq 1$). The final step is to create the fallible observables by the addition of measurement error to the $\xi_p(t_i)$ according to the classical test theory model: $X_p(t_i) = \xi_p(t_i) + \varepsilon_i$ for $p = 1, \ldots, n$.

Consequences for Second Moments

The choices of the values above determine the population values of the familiar second moments of $\xi_p(t_i)$ or $X_p(t_i)$ for the artificial data. In practice, these values of these quantities—variances, correlations, etc.—are often chosen first (say to correspond to values familiar from empirical research or common sense), and then solutions (explicitly or by trial and error) for the corresponding values for the simulation procedure above are obtained. The relations that provide values of these second moments for the $\xi_p(t_i)$ are
variance

$$\sigma^2_{\xi(t)} = \sigma^2_{\xi(t^\circ)} + ((t - t^\circ)/\kappa)^2\sigma^2_{\xi(t^\circ)},$$

covariance (also yields correlation, using above)

$$\sigma_{\xi(t_1)\xi(t_2)} = (t_1 - t^\circ)(t_2 - t^\circ)\sigma^2_\theta + \sigma^2_{\xi(t^\circ)},$$

correlation between change and initial status

$$\rho_{\theta\xi(t)} = \frac{t - t^\circ}{[\kappa^2 + (t - t^\circ)^2]^{1/2}},$$

correlation with exogenous variable, W

$$\rho_{W\xi(t)} = \frac{(t - t^\circ)\rho_{W\theta} + \kappa\rho_{W\xi(t^\circ)}}{[\kappa^2 + (t + t^\circ)^2]^{1/2}}$$

Technical Specifications for Exhibit 1

In terms of the model parameters, the values for the artificial data in Exhibit 1 are $t^\circ = 2$; $\sigma^2_\theta = 5.333$; $\sigma^2_{\xi(t^\circ)} = 48$; for $\theta \sim U[1, 9]$, $\xi(t^\circ) \sim U[38, 62]$. Population mean rate of change is 5, and values of the population correlation coefficients among the $\xi(t_i)$ for observation times $t_1 = 1$, $t_2 = 3$, $t_3 = 5$ are $\rho_{\xi(1)\xi(3)} = .80$, $\rho_{\xi(1)\xi(5)} = .447$, $\rho_{\xi(3)\xi(5)} = .894$. Furthermore, for the fallible measure X with $\text{var}(\varepsilon) = 10$, the population correlations are $\rho_{X(1)X(3)} = .674$, $\rho_{X(1)X(5)} = .391$, $\rho_{X(3)X(5)} = .781$.

ACKNOWLEDGMENTS

I wish to thank Ghassan Ghandour and Haggai Kupermintz for computational and editorial assistance and Gary Williamson for providing the North Carolina data. Programs described in this chapter can be obtained by writing to David Rogosa at rag@leland.stanford.edu.

ADDITIONAL REFERENCES

Alwin, D. F. (1988). Structural equation models in research on human development and aging. In K. W. Schaie, R. T. Campbell, W. M. Meredith, & S. M. Rawlings (Eds.), *Methodological issues in aging research* (pp. 71–170). New York: Springer.

Brogan, D. R., & Kutner, M. H. (1980). Comparative analyses of pretest-postest research designs. *American Statistician, 34*, 229–232.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*, 147–158.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* New York: Chapman & Hall.

Fern, T. (1975). A Bayesian approach to growth curves. *Biometrika, 62*, 89–100.

Freedman, D. A. (1987). As others see us: A case study in path analysis. *Journal of Educational Statistics, 12*, 101–128.

Freedman, D. A. (1991). Statistical models and shoe leather. In P. Marsden (Ed.), *Sociological Methodology 1991* (pp. 291–313). Washington, DC: American Sociological Association.

Goldstein, H. (1979a). *The design and analysis of longitudinal studies.* London: Academic Press.

Goldstein, H. (1979b). Some models for analysing longitudinal data on educational attainment. *Journal of the Royal Statistical Society A, 142*, 407–442.

Holland, P. W. (1988). Causal inference, path analysis and recursive structural equation models. In C. Clogg (Ed.), *Sociological Methodology 1988* (pp. 449–484). Washington, DC: American Sociological Association.

Hui, S. L., & Berger, J. O. (1983). Empirical Bayes estimation of rates in longitudinal studies. *Journal of the American Statistical Association, 78*, 753–760.

Rogosa, D. R. (1987). Casual models do not support scientific conclusions: A comment in support of Freedman. *Journal of Educational Statistics, 12*, 185–195.

Rogosa, D. R. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. M. Meredith, and S. C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 171–209). New York: Springer.

Rogosa, D. R. (1991). A longitudinal approach to ATI research: Models for individual growth and models for individual differences in response to intervention. In R. E. Snow and D. E. Wiley (Eds.), *Improving inquiry in social science: A volume in honor of Lee J. Cronbach* (pp. 221–248). Hillsdale, NJ: Lawrence Erlbaum.

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

Rogosa, D. R., & Ghandour, G. A. (1989). TIMEPATH: Statistical analysis of individual trajectories. Stanford University.