Psychological Bulletin
1980, Vol. 88, No. 3, 622–637

Regression Toward the Mean and the Study of Change

John R. Nesselroade
College of Human Development
Pennsylvania State University

Stephen M. Stigler
and Paul B. Baltes
Center for Advanced Study
in the Behavioral Sciences
Stanford, California

Among many threats to the representation and assessment of change in behavioral research are effects associated with the phenomenon of regression toward the mean. This concept has a long history, but its definition and interpretation have remained unclear. In the present article, regression effects in longitudinal sequences of observations are examined by formulating expectations for later observations conditioned on an initial selection score value. The expectations are developed for several variations of classical test theory and autocorrelation models. From this perspective, expectations based on the general concept of regression are seen not only to depart from those depicted in the psychometric lore but to vary considerably from one underlying model to another, particularly as one moves from the two-occasion to a multiple-occasion measurement framework. In some cases "unrelenting" regression toward the mean occurs. In others, scores may initially regress and then show egression from the mean. Still other patterns are expected for some models. In general, it is important to understand that regression toward the mean is not an ubiquitous phenomenon, nor does it always continue across occasions. It is necessary to specify the characteristics of model assumptions to understand the when, how, and extent of regression toward the mean. Past interpretations have been incomplete and to an extent incorrect because they focused largely on a limited circumstance: two-occasions of measurement and simplexlike correlation matrices.

Researchers interested in the study of change, particularly developmental psychologists, have been confronted repeatedly by a variety of warnings and injunctions that psychometricians and methodologists have raised vis-à-vis the analysis of change data. On one hand, developmental psychology is generally defined as dealing with the study of change and interindividual differences and similarities in change (e.g., Baltes, Reese, & Nesselroade, 1977; Wohlwill, 1973), thereby dramatizing the need for methods and procedures for measuring change. On the other hand, however, it is concluded by psychometricians (e.g., Cronbach & Furby, 1970) that attempts to assess change directly are so beset by measurement, design, and interpretation problems that researchers should depend primarily on less direct alternatives. Their proposals include, for example, using multivariate conceptions of change and the analysis of covariance matrices and variances corrected for error or rephrasing research questions so that indirect methods of assessing changes (e.g., examining differences between posttreatment means of randomly constituted experimental and control groups) can be used. One of the

Development and preparation of this manuscript was partially supported by National Institute on Aging Grant 5-R01-AG004403 to P. B. Baltes and S. L. Willis. The work of the second author was partially supported by National Science Foundation Grants BNS 76-22943 A02 and SOC 78-01668.

S. M. Stigler is now at the Department of Statistics, University of Chicago. P. B. Baltes is now at the Max Planck Institute of Human Development and Education, 1000 Berlin 33, West Germany.

Requests for reprints should be sent to John R. Nesselroade, College of Human Development, Pennsylvania State University, University Park Pennsylvania 16802.

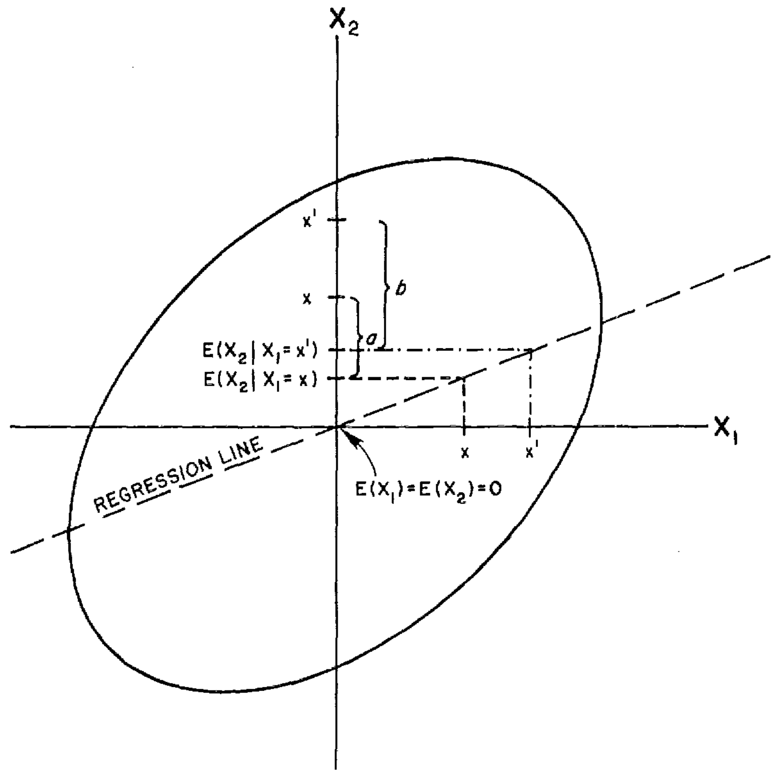

Figure 1. Hypothetical bivariate distribution [$\text{corr}(X_1, X_2) < 1.00$] showing regression toward the mean effect. [For a given value of X_1, e.g., x, $E(X_2|X_1 = x)$ is closer to $E(X_2)$ in standard deviation units than x is to $E(X_1)$ in standard deviation units. Difference between x and $E(X_2|X_1 = x)$ labeled a is less than difference between x' and $E(X_2|X_1 = x')$ labeled b.]

most often identified culprits in discussions of change measurement problems is the phenomenon known as regression toward the mean. Failure to account properly for regression toward the mean has been invoked as a devastating confound in such areas of study as compensatory education (Campbell & Erlebacher, 1970), business (Hotelling, 1933), general education and psychology (Thorndike, 1942), behavior genetics (Humphreys, 1978), and developmental research (Furby, 1973).

The present article is both a logical extension and, to some extent, a reformulation of earlier articles, especially those by Baltes, Nesselroade, Schaie, and Labouvie (1972), Campbell and Stanley (1963), Clarke, Clarke, and Brown (1960), Furby (1973), and Lord (1963). We contend that regression toward the mean is a concept that is too readily invoked, often in a simplistic way, in discussion of change and that there still remains a great need to spell out more carefully not only what it is but what it is not. In so doing we identify those conditions under which one should be on the alert for regression effects and also those conditions under which regression effects are minimized, nonexistent, or manifested in ways that are at variance with traditional expectations.

Two features of the present article are particularly noteworthy. One, it is shown how the discussion of regression toward the mean is modified by considering multiple occasions of measurement. In the past, much of the discussion has centered on either two-variable or one-variable-measured-on-two-occasions situations. We believe that these situations maximize design and analysis problems associated with regression toward the mean. Second, careful attention is paid to the nature of the underlying interoccasion correlation structure of scores. It seems to us that much past discussion of regression effects has not sufficiently considered the role of different patterns of autocorrelation and has often assumed that simplexlike correlation matrices are the only germane pattern. It is shown that the form of regression toward the mean changes significantly as alternative patterns of cross-occasion correlations are assumed.

Past Literature

The statistical concept of regression is over a century old—indeed, it was introduced under the name "reversion" by Galton in 1877. Yet despite the relative antiquity of the concept, a frequent failure to grasp its implications has been and continues to be the root of errors in scientific inference. In this section, regression toward the mean is briefly defined as it is known to psychologists, particularly developmental psychologists. In a later section, a more precise mathematical formulation is used.

Meaning of the Regression Concept

In past literature regression toward the mean has been defined in relation to both the *identification* of the phenomenon and its *explanation*. In the present context, regression toward the mean is considered primarily in the setting of at least two occasions of observation involving the same measurement variable and the same experimental units. Such a restriction permits the use of the term *change*.

In the developmental literature, Furby's (1973) article is perhaps the latest comprehensive effort to deal with regression toward the mean. Furby, for example, defined the regression effect in the following manner (our notation): For a given score on X_1 (e.g., x'), the corresponding mean score on X_2 [e.g., $E(X_2|X_1 = x')$] is closer to $E(X_2)$ in standard deviation units than x' is to $E(X_1)$ in standard deviation units. She concluded that such effects will be observed when the correlation is less than perfect, for whatever reasons. In addition, the greater the deviation from the mean score [e.g., $x' - E(X_1)$], the greater the regression toward the mean. Figure 1 is the kind that is typically used to demonstrate the phenomenon of regression.

In principle, the preceding statement of statistical regression toward the mean can be correct and useful but only within restricted boundaries that for some issues are inappropriate. For example, the definition of regression involves the implicit, if not explicit, standardization of score distributions, thus invoking immediate constraints and signifying that some choice in the form of data representation already has been made by the investigator. Furthermore, as is shown later, Furby's statement is correct for the two-occasion situation, but generalization to multiple occasions involves additional qualifications.

Explanations of the Regression Phenomenon

A number of explanations for the phenomenon of regression can be found in the literature (e.g., Furby, 1973). They range from simplistic formulations to alternatives that attempt to account for regression by specifying the nature of its antecedents.

Regression is often accounted for by the essentially tautological statement that it is due to the lack of perfect correlation between the two sets of scores. From a strictly causal point of view, it would be as useful to say that the lack of perfect correlation is due to regression toward the mean. In neither case is one gaining any knowledge over and above that provided by the definition of regression. Lack of perfect correlation and regression toward the mean are essentially equivalent.

Two major lines of reasoning have been used to explain why a lack of perfect correlation and regression toward the mean occur. One is the assumed nature of measurement errors. This argument, couched in classical test theory and the two-occasion situation, depends on the assumption that errors of measurement are uncorrelated over time and that at the first occasion individuals in extreme scoring groups (high or low) have a higher likelihood for positive (high) or negative (low) error components in their observed scores. Since error of measurement is assumed to be uncorrelated across occasions, the expectation is that extreme scoring groups will have a 50:50 ratio of positive and negative error on other measurement occasions. The scores of previously established extreme groups will, on other occasions, include on the average the same distribution of positive and negative error components. Therefore, in terms of observed scores at the second occasion, the extreme groups are "moving" closer to the overall mean. This interpretation assumes that the movement is not in the true scores but is fallaciously produced in the observed scores by error of measurement. The magnitude of the phenomenon would be greater the less accurate (reliable) were a given set of measurements. In principle, if one is willing to accept the basic tenets of measurement theory that are involved, this rationale is correct. As is shown later, however, the argument applies only to the two-occasion situation.

A second class of explanations for the lack of perfect correlation and associated regression has to do with identifying causal factors operating between occasions. For example, Furby (1973) and Clarke et al. (1960) recognized that extreme scores (in addition to and independent of error of measurement components) in a distribution must reflect, at least in part, relatively rare combinations of antecedent events. It is further argued that rare combinations of events tend not to be maintained over time, and one therefore arrives at the expectation that individuals exposed at one time to rare combinations of antecedent events will tend not to be so exposed at a later time and consequently will show a change toward the overall mean. This argument is in many respects a variation on the error of measurement view, although it deals with antecedents of change rather than its measurement. Moreover, the argument is a potential but not a necessary explanation as alternative conceptions of causal factors are possible. For example, applications of causal factors do not necessarily involve random events, and they may lead to egression, as is discussed later.

Basic Definition and Notation

The remainder of the discussion is focused on the regression toward the mean concept as it pertains to the study of phenomena defined by the repeated (at least two occasions) assessment of a particular variable. The primary objective is to identify distinct aspects of data that bear directly on the anticipation and evaluation of regression effects, to formulate models that reflect these characteristics, and to develop expectations in relation to them. To provide an explicit basis for these subsequent developments, a statistical formulation of the general regression concept is presented first.

Consider a sequence of test scores, recorded as measurements of the same attribute of a single individual on a series of occasions. We denote this sequence by the letters $X_1, X_2, \ldots, X_n$. Suppose, for this discussion, that the tests have been scaled so that, for the population of individuals under study, the standard deviations of the scores on one occasion are the same as those on another occasion. To simplify the notation for the discussion, this common standard deviation is taken as unity and test scores as deviations from the population mean. Thus if we consider our individual as randomly selected from the population, his or her score on Occasion i will have expectation $E(X_i) = 0$ and standard deviation $\sigma(X_i) = 1$, for each Occasion i.

The following notation is also used: The covariance of X_i and X_j is written as $\text{cov}(X_i, X_j)$, and the correlation of X_i and X_j is written $\text{corr}(X_i, X_j)$, where $\text{corr}(X_i, X_j)$ equals $\text{cov}(X_i, X_j)/\sigma(X_i)\sigma(X_j)$. Since $\sigma(X_i) = \sigma(X_j) = 1$, we have $\text{corr}(X_i, X_j) = \text{cov}(X_i, X_j)$ here, but not all variables discussed later have standard deviation 1, and it is well to maintain the distinction between $\text{corr}(X_i, X_j)$ and $\text{cov}(X_i, X_j)$.

The expressions to be derived are general, but, for purposes of illustration, the examples are chosen from the traditional psychometric framework of measurement. Thus all of the examples we consider may be thought of as cases in which the scaled test score X_i is derived from a raw test score Y_i by the relation $X_i = Y_i/\sigma(Y_i)$. The raw test score Y_i is measured as a deviation from a population mean and can in turn be broken down into two components, for example, $Y_i = S_i + e_i$. The classical interpretation of such a representation is the measurement error model in which S_i is taken as the "true score" on Occasion i and e_i the measurement error on that occasion. We often adopt this terminology because of its familiarity, but the ideas are more general than those of measurement error models. They also encompass, for example, observed scores conceived of as measured without error and consisting of permanent trait and transient state components (see, e.g., Nesselroade & Bartsch, 1977). We comment briefly later on the methodological difficulties in distinguishing between these models.

Two Occasions Versus Multiple Occasions

In this section formulas for the general situation are developed without yet considering the role of measurement error as it is treated in the psychometric literature. How does regression toward the mean extend from the two-occasion framework? Is the extension a simple one, and if not, what aspects of the stochastic structure of the test scores underlie its complexity? What considerations are needed if a proper analysis of behavior over multiple occasions is to be made? To better understand what is at issue, let us first examine how the extension from two occasions to multiple occasions might naively, and falsely, be thought to occur.

Two-Occasion Case

The phenomenon of regression toward the mean for two occasions (scores X_1 and X_2) is concerned with the conditional expectation of X_2, given that the test was administered on Occasion 1 and the score $X_1 = x$ has been recorded. Under the supposition that the scores are jointly normally distributed,[1] this conditional expectation is given by

$$E(X_2 | X_1 = x) = \operatorname{corr}(X_1, X_2) \cdot x.$$

Since $-1 \leq \operatorname{corr}(X_1, X_2) \leq 1$ always, if X_1 and X_2 are both measured in units of their standard deviations, then given $X_1 = x$, we "expect" X_2 to be closer to its mean [viz., $E(X_2) = 0$]. *Expect* may be given a more concrete meaning here by supposing that we select a very large number of individuals with scores nearly x on Occasion 1. Their average score on Occasion 2 will be very nearly $\operatorname{corr}(X_1, X_2) \cdot x$. Some will have scored lower, some higher, and some even higher than the initial score x, but the average will be very nearly $\operatorname{corr}(X_1, X_2) \cdot x$. This situation was shown in Figure 1.

Multiple-Occasion Case

Does regression toward the mean continue in the same fashion beyond the second occasion?

Expectation fallacy. Through repeated application of the formulas given for the two-occasion case, we have

$$E(X_2 | X_1 = x) = \operatorname{corr}(X_1, X_2) \cdot x,$$
$$E(X_3 | X_2 = y) = \operatorname{corr}(X_2, X_3) \cdot y,$$
$$E(X_4 | X_3 = z) = \operatorname{corr}(X_3, X_4) \cdot z,$$

and so forth.

Now although these equations are correct, they lead naturally to a mistaken conclusion in the present multiple-occasion situation. The natural misunderstanding is to picture an expected sequence of test scores, starting at an observed $X_1 = x$ and getting successively closer to the mean of zero, as if regression toward the mean implied a necessarily continuous regression beyond the second occasion, as we believe is often assumed. Intuitively, given $X_1 = x$, we might naively expect the sequence

$$\begin{aligned} X_1 &= x, \\ X_2 &= \operatorname{corr}(X_1, X_2) \cdot x, \\ X_3 &= \operatorname{corr}(X_2, X_3) \cdot X_2, \\ &= \operatorname{corr}(X_2, X_3) \cdot \operatorname{corr}(X_1, X_2) \cdot x, \\ X_4 &= \operatorname{corr}(X_3, X_4) \cdot X_3, \\ &= \operatorname{corr}(X_3, X_4) \cdot \operatorname{corr}(X_2, X_3) \\ &\qquad \cdot \operatorname{corr}(X_1, X_2) \cdot x, \end{aligned}$$

and so forth. Thus if $\operatorname{corr}(X_i, X_{i+1}) = .5$ for all i (as may well be true) and if we observe $X_1 = 2.0$, the expected sequence becomes

$$X_1 = 2, \; X_2 = 1, \; X_3 = .5, \; X_4 = .25,$$

and so forth.

[1] The assumption that all scores and errors are jointly normally distributed can be dispensed with if one replaces *expected sequence* with *best predictive sequence* throughout, and if one understands *best* in the restrictive sense of smallest mean square prediction error among all sequences in which each term is a linear function of the previous terms. Thus if one does not view this linearity assumption as restrictive, the assumption of normality is unimportant to the conclusions.

Regression toward zero (the mean) seems steady and unrelenting. *But the repeated application of the two-occasion formula is generally an error.* It would be tantamount to invoking a very restrictive and only infrequently fulfilled assumption about the stochastic structure of sequences of test scores. Technically, the assumption is that the sequence of test scores follows a Markov process (Frederiksen & Rotondo, 1979; Glass, Willson, & Gottman, 1972; Hibbs, 1974; Nelson, 1973).

The distinguishing feature of a Markov sequence of scores is that if one wishes to determine the probability distribution of a future score, given both past and present scores, then the information about the past is unnecessary—all useful information about future score distributions is summarized in the present score. In particular, if we wish to calculate $E(X_3|X_1 = x,\ X_2 = y)$ and the sequence is Markov, then it would be enough to calculate $E(X_3|X_2 = y)$, for with Markov sequences,

$$E(X_3|X_2 = y) = E(X_3|X_1 = x, X_2 = y).$$

But this, as is seen, is not generally true for stochastic sequences, and the consequences for expected sequences and regression toward the mean may be important. To see this more clearly, consider a correctly calculated expected sequence for a multiple-occasion situation.

Correct expected sequence. Given that an individual is selected as having score $X_1 = x$ on Occasion 1, what is his or her future sequence of scores expected to be? There will be two factors at work: the value of the selected score or selection criterion $X_1 = x$, and the characteristics of the population of observed scores [including the population means, $E(X_i) = 0$, and $\text{corr}(X_1, X_i)$].

To produce a correct expected sequence, the question can be asked with respect to two occasions at a time but in a different way than was done in the fallacious argument identified earlier. The task is to follow longitudinally the same selected group of individuals across occasions. This means that the selection criterion remains $X_1 = x$ throughout. We ask, then, for the expected sequence from the standpoint of the occasion on which selection occurred (the first occasion). Because selection occurred only once, we do not need to reask the question on each occasion. Then the expected sequence is simply

$$E(X_2|X_1 = x),\ E(X_3|X_1 = x),\ E(X_4|X_1 = x),$$

and so forth, and all of these can be immediately given from our knowledge of the two-occasion situation: The expected sequence, beginning with Occasion 1 and given $X_1 = x$, is

$$x;\ \text{corr}(X_1, X_2)\cdot x;\ \text{corr}(X_1, X_3)\cdot x;\ \text{corr}(X_1, X_4)\cdot x;$$

and so forth. It can be proved mathematically that, for example,

$$E[X_3|X_1 = x, \text{ and } X_2 = \text{corr}(X_1, X_2)\cdot x] = E(X_3|X_1 = x),$$

but by keeping in mind that selection is only performed once, on the first occasion, we see that this more laborious computation is not needed.[2] Thus we have the following comparison between the expected sequences of scores, given that an individual is selected as having score x on Occasion 1.

Correct expected sequence:

x
$\text{corr}(X_1, X_2)\cdot x$,
$\text{corr}(X_1, X_3)\cdot x$,
$\text{corr}(X_1, X_4)\cdot x$, and so forth.

Incorrect expected sequence:

x
$\text{corr}(X_1, X_2)\cdot x$,
$\text{corr}(X_2, X_3)\cdot\text{corr}(X_1, X_2)\cdot x$,
$\text{corr}(X_3, X_4)\cdot\text{corr}(X_2, X_3)\cdot\text{corr}(X_1, X_2)\cdot x$, and so forth.

It is easily seen, since $|\text{corr}(X_i, X_{i+1})| \leq 1$, that the incorrect sequence necessarily re-

[2] Comparison of this expression with the earlier fallacious argument is instructive. The earlier argument, which would have followed for a Markov sequence, would have had $E[X_3|X_1 = x, \text{ and } X_2 = \text{corr}(X_1, X_2)\cdot x] = E[X_3|X_2 = \text{corr}(X_1, X_2)\cdot x]$. The right-hand side of the equation is what we would have for the expected score on Occasion 3 if we selected an individual on Occasion 2 with score $\text{corr}(X_1, X_2)\cdot x$ from a previously unselected population. The fact that $E[X_3|X_2 = \text{corr}(X_1, X_2)\cdot x]$ and $E(X_3|X_1 = x)$ are not generally equal is a reflection of the fact that we only select once (on Occasion 1), and the expected score must be based on this condition and not some subsequent one.

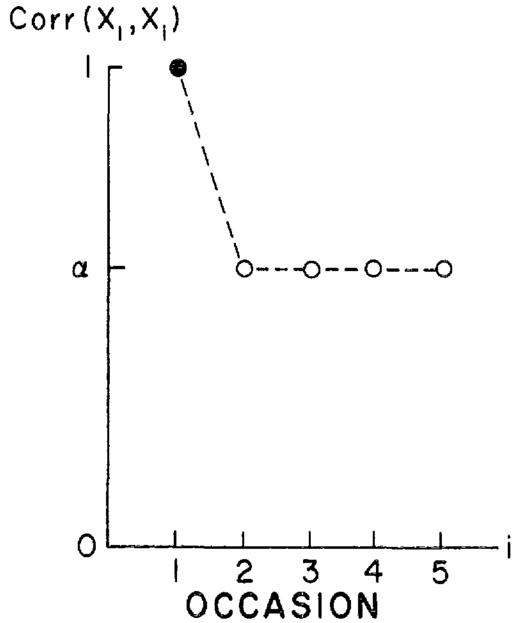

Figure 2. Autocorrelation sequence (scaled expected score sequence) for the classical measurement error model. [Here $\alpha = \sigma^2(S)$.]

gresses toward the mean of zero, as noted in the earlier discussion of the expectation fallacy. Clearly the pattern of the autocorrelations $\text{corr}(X_1, X_i)$ will determine the pattern of the expected sequence. It is not necessarily one of continual, unrelenting regression toward the mean. If, for example, the autocorrelation sequence $\text{corr}(X_1, X_i)$, $i > 1$, is constant [e.g., $\text{corr}(X_1, X_2) = \text{corr}(X_1, X_3) = \cdots = \text{corr}(X_1, X_n) = .5$], we would expect regression from Occasion 1 to Occasion 2 but not thereafter. The expected sequence of scores of a group of individuals selected as having score $X_1 = x$ on Occasion 1 is then x, $.5x$, $.5x$, $.5x$, $\ldots$, $.5x$.

One specific example of this particular regression pattern is in the classical measurement error model. Suppose that over n occasions of measurement, the test scores of a randomly selected individual, $X_1, X_2, \ldots, X_n$, each consist of a true score S that remains unchanged and a measurement error e_i, such that $S, e_1, e_2, \ldots, e_n$ are mutually independent (and hence uncorrelated). Then $X_i = S + e_i$, $i = 1, \ldots, n$, and as before we suppose that $\sigma^2(X_i) = 1$ and $E(S) = 0 = E(e_i)$. Now, $\sigma^2(X_i) = \sigma^2(S) + \sigma^2(e_i)$ [because $\text{cov}(S, e_i) = 0$], and since $\text{cov}(e_i, e_j) = 0$ for $i \neq j$, we have $\text{cov}(X_i, X_j) = \sigma^2(S)$, and therefore $\text{corr}(X_i, X_j) = \sigma^2(S)$ for $i \neq j$. Let us denote $\alpha = \sigma^2(S)$. Then $0 \leq \alpha \leq 1$ [since $\alpha + \sigma^2(e_i) = 1$], and given $X_1 = x$ we have as the correct expected sequence:

$$x, \alpha x, \alpha x, \alpha x, \ldots, \alpha x.$$

Thus, as shown in Figure 2, after the first retest, no further regression is expected; the effect on changes in the expected score of selecting the individual as having score $X_1 = x$ is dissipated after the second occasion. Only when all individuals have the same true score is $\alpha = \sigma^2(S) = 0$ and the regression complete; only when there is no measurement error is $\alpha = 1$ and is there no regression even from the first to second occasion. The incorrect expected sequence for this example would have been x, αx, $\alpha^2 x$, $\alpha^3 x$, and so forth.

We emphasize that the "correct" expected sequence is computed assuming that a group of individuals each has score $X_1 = x$ on Occasion 1. If additional selection is exercised further down the line, further regression may occur but not always toward the overall mean of zero. For example, if of all individuals selected as having scores $X_1 = x$ on Occasion 1, *one* is selected after the second test as having score $X_2 = 0$, we would expect his or her score on the third occasion to regress toward the mean of the initially selected population [viz., $E(X_2 | X_1 = x) = \alpha x$] and thus away from zero. In fact,

$$E(X_3 | X_1 = x \text{ and } X_2 = 0) = \frac{\alpha}{(\alpha + 1)} \cdot x$$

(as long as $\alpha < 1.$). As stated earlier, one way to interpret the correct expected sequence is to think of a very large number of individuals from the even much larger population as being selected as having score x on Occasion 1. Then their average scores on the successive occasions will nearly agree with the expected sequence.

Also in the context of this example, the often held belief that measurement error (or a transient state component) necessarily produces a regression effect that makes it impossible or difficult to measure change properly is not correct, at least not with multiple-occasion data. In this example we see that after the second occasion, the troublesome effect of selecting the individual based on the initial score $X_i = x$ (the regression effect) is spent. Lacking a differential effect from Occasions k to $k+1$, the expected change in score between these occasions is zero. That is,

$$\begin{aligned} E(X_{k+1} - X_k | X_1 = x) &= x[\operatorname{corr}(X_1, X_{k+1}) \\ &\quad - \operatorname{corr}(X_1, X_k)] \\ &= x[\sigma^2(S) - \sigma^2(S)] \\ &= 0 \end{aligned}$$

for this model. Notwithstanding (a) the existence of a possibly large measurement error and (b) a screening or selection at the first occasion, the effect of regression toward the mean is gone after Occasion 2 in terms of expected *changes* in scaled scores.[3] The *level* of the expected scaled score remains at the constant value $E(X_k | X_1 = x) = x\sigma^2(S)$, but any change in aggregates of such selected scores that cannot be attributed to random fluctuation must be attributed to change in true score and not to a regression effect. Because the effect of measurement error on regression toward the mean does not extend beyond the second occasion, an easy design control is apparent. If individuals are selected at the first occasion, a proper representation of change unconfounded with regression effects is achieved by beginning the charting of change functions at the second occasion. In other words, the first occasion serves as criterion for selecting extreme groups, and the second occasion serves as initiation point for representing change.

The previous observations concerning the disappearance of measurement-error-related regression effects are more widely applicable. If we consider the most general model we have discussed, that in which the raw scores $Y_i = S_i + e_i$ and the true scores S_i change either stochastically or deterministically over time, then the expected change between future occasions $E(X_{k+1} - X_k | X_1 = x)$ is proportional to $x \cdot \operatorname{cov}(S_{k+1} - S_k, S_i)$ and depends neither on measurement errors nor on other factors that are uncorrelated over time, both with each other and with any nontransitory true score component. (As presented here, this conclusion depends on the variances of the unscaled scores remaining constant across occasions.)

Nature of the Autocorrelational System

Thus far we have shown that designs involving more than two occasions lead to different anticipations regarding regression toward the mean. For example, if (as previously described) $\operatorname{corr}(X_1, X_2) = \operatorname{corr}(X_1, X_3) = \cdots = \operatorname{corr}(X_1, X_n)$, then there is no further regression toward the mean after the second occasion. Extreme groups identified by their scores on Occasion 1 will exhibit regression on Occasion 2, but (as the earlier example shows) this pattern need not continue. Of course, if an extreme group is redefined on Occasion 2 based on Occasion 2 scores, this group will exhibit regression on Occasion 3. But the original extreme group will not continue to regress in this example.

Obviously, a great variety of autocorrelational patterns can be conceived of, and each could be examined concerning its implications for the development of regression expectations. To display a range of possibilities within our space limitations, we focus on three prototypical autocorrelational systems: (a) constant autocorrelations, (b) autocorrelations that decrease with increasing separation of observations in a sequence, and (c) autocorrelations that increase with increasing separation of observations. A limited set of more complex patterns is briefly considered later.

Constant Autocorrelations

The statement that regression ceases after Occasion 2 is restricted to the situation in which the autocorrelations between occasions, $\operatorname{corr}(X_1, X_i)$, for $i = 2, \ldots, n$, are constant at whatever value. Note also that the anticipation of no further regression toward the

---

[3] In the present example, Occasions 1 and 2 are used as selection and postselection occasions. This need not be the case. Should other occasions from a longer observational sequence be used as selection and postselection occasions, for example, Occasion 1 and Occasion 4, the regression toward the mean effect shown earlier for Occasion 1 and Occasion 2 would hold for the pair of occasions being considered. In other words, the regression toward the mean effect described is not specific to the first pair of occasions in an observation series. It is specific to the first pair of occasions defining the selection and postselection events.

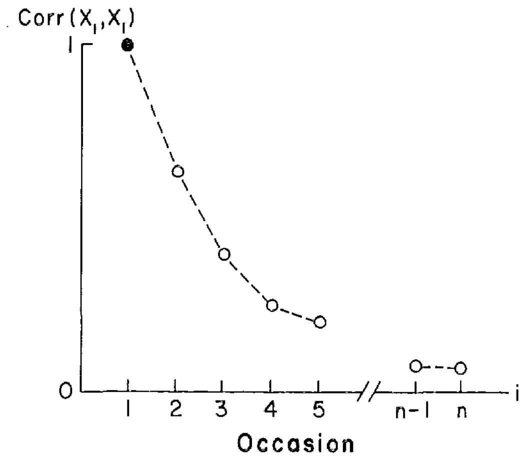

Figure 3. Decreasing autocorrelation sequence. (For $X_1 = 1.00$, this is the expected score sequence. If $X_1 = x$, the expected score sequence is this sequence multiplied by x.)

mean beyond the second occasion is independent of the magnitude of measurement error (reliability). Measurement error is only one of several reasons why the autocorrelation might be less than one.

Assuming constant autocorrelations is not always reasonable. Since $\text{cov}(X_1, X_k - X_{k-1}) = \text{corr}(X_1, X_k) - \text{corr}(X_1, X_{k-1})$, it is equivalent to assuming that the initial score X_1 is uncorrelated with all future score changes in the scaled test scores $X_k - X_{k-1}$. Now in developmental research it is often found that score changes are correlated with initial scores and thus that the magnitude of interoccasion correlations changes with the interoccasion distance. However, constant autocorrelations are possible. Consider a situation in which true trait scores are invariant with occasion-specific state variability, resulting from treatments applied randomly to subgroups of individuals, separately at each occasion (Baltes, Nesselroade, & Cornelius, 1978). Such a situation is, of course, similar to that in the error of measurement model, although in this instance the treatment is substantive.

Let us examine the role of the autocorrelation sequence in further detail by considering examples of two additional patterns of interoccasion correlation:[4] for decreasing, $\text{corr}(X_1, X_2) > \text{corr}(X_1, X_3) > \cdots > \text{corr}(X_1, X_n)$; for increasing, $\text{corr}(X_1, X_2) < \text{corr}(X_1, X_3) < \cdots < \text{corr}(X_1, X_n)$.

Note that these two models of autocorrelations are extreme cases selected for didactic purposes. More complicated models are briefly mentioned later. Furthermore, at this point we are proceeding as if these two autocorrelational models exist. In a later section, we introduce possible underlying measurement models that are associated with each of the autocorrelational patterns.

Decreasing Autocorrelation: Continual Regression

A second major type of interoccasion correlation is one in which the autocorrelation sequence $\text{corr}(X_1, X_i)$ decreases with i, that is, in which the interoccasion correlations decrease with increasing distance between occasions. The classical case of such a matrix of correlations is the simplex (Guttman, 1954; Humphreys, 1960; Kenny, 1979). This case is illustrated in Figure 3.

We assume that individuals are selected for study based on their scores on Occasion 1, and thus the selected individuals have expected score sequences $X_1 = x, x\,\text{corr}(X_1, X_2), \ldots, x\,\text{corr}(X_1, X_n)$; thus the expected score sequence is just a scaled version of the autocorrelation sequence, scaled by the initial score x. Thus the behavior of the autocorrelation sequence determines that of the expected score sequence, and a decreasing sequence of autocorrelations leads to a continuing regression toward the mean.

In our view, many developmental psychologists believe that this arrangement is, in principle, the most common one in empirical research. Because of the frequent appearance of the simplex pattern in research and perhaps partly because of the subtle reinforcement of the notion that comes from the expectation fallacy discussed earlier, there has been a tendency to consider this situation as ubiquitous. Yet, as the present development shows, it is ubiquitous only to the extent that decreasing autocorrelations $\text{corr}(X_1, X_i)$ are, and, as is demonstrated, this need not be the case.

---

[4] For reasons of simplicity, we restrict attention to sequences with positive autocorrelations.

Increasing Autocorrelations: Egression

The third major situation that can be encountered is that in which the autocorrelations $\text{corr}(X_1, X_i)$ increase with i, that is, in which interoccasion correlations increase as the distance between occasions increases.

In such a situation (see Figure 4 for an illustration), we would find the following outcomes. Focusing on the use of Occasion 1 as a predictor for all subsequent occasions, we see that the magnitude of regression decreases continually when compared with the effect from Occasion 1 to Occasion 2.

The score sequence expected for an individual with initial score $X_1 = x$ (viz., the autocorrelation sequence scaled by multiplying by x) would decrease from Occasion 1 to Occasion 2, but after that (looking from the second occasion on), it would exhibit *egression* from the mean. The future expected interoccasion change between Occasions k and $k+1$ of an individual selected with initial score $X_1 = x$, evaluated from the standpoint of the time of selection (Occasion 1), is away from the mean. In terms of our mathematical formulation, we would have

$$E(X_{k+1} - X_k | X_1 = x) > 0,$$

if $x > 0$ (and the reverse inequality if $x < 0$).

Now, this situation may seem unlikely, but it is possible, and there are a number of situations in which it might actually be anticipated. In data reported by Humphreys and Parsons (1979), cross-correlation of listening and intellectual composite scores at Grades 5, 7, 9, and 11 show increased values over time if one corrects for within-occasion correlated error components. In developmental research in general, a pattern of increasing autocorrelations would occur if there is a system of determinants that leads to accumulating advantage and disadvantage as a function of initial standing. Baltes et al. (1978) have simulated one such outcome pattern, and we present a mathematical formulation of other examples later. For the present, the important point to note is that such a situation could occur and that it would correspond to expected behavior contrary to the continuing regression toward the mean that is often thought to be ubiquitous. Here, after the initial two-occasion regression between Occasions 1 and 2, all future expected changes are away from the mean.

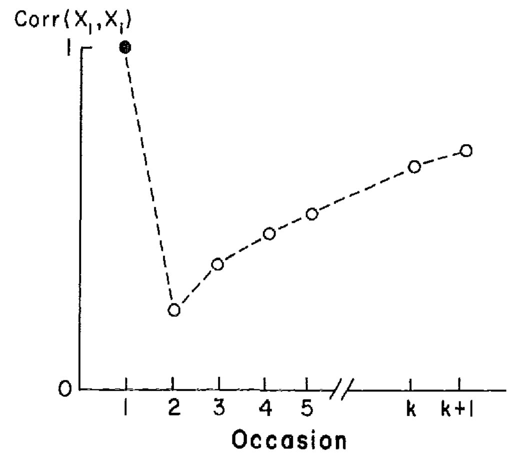

Figure 4. Increasing autocorrelation sequence. (For $X_1 = 1.00$, this is the expected score sequence. If $X_1 = x$, the expected score sequence is this sequence multiplied by x.)

Complex Systems of Autocorrelations

The cases considered thus far all display patterns of marked regularity. In each case the autocorrelations are simply ordered—constant, decreasing, or increasing—and more complex situations are clearly possible. Causal explanations for patterns in the autocorrelation sequence have not yet entered into our statistical discussion, but to understand how more complex systems relate to regression toward the mean, we consider a model that exhibits such behavior, as shown in Figure 5.

Recognizing that this is but one of several alternatives that could be considered, let us suppose than an occasion-specific intervention is imposed on an ongoing system. Without speculating on the explanatory origin of the intervention, we might suppose it has the effect of adding a component C to the true score on the occasion in which the intervention takes place and that this component C is uncorrelated with any other true score or measurement error. For purposes of illustration, suppose that, except for the intervention, the observed scores follow the classical measurement model and that the intervention

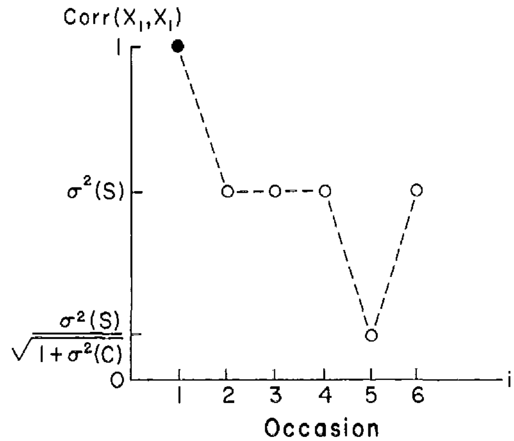

Figure 5. One possible complex autocorrelation sequence with an occasion-specific intervention on Occasion 5.

takes place on Occasion 5. Then the unscaled observed scores are

$$Y_i = S + e_i, \qquad \text{for } i = 1, 2, 3, 4, 6, 7, \ldots, n;$$
$$Y_i = S + C + e_i, \text{ for } i = 5;$$

furthermore, $\text{cov}(Y_i, Y_k) = \sigma^2(S)$, $\sigma^2(Y_i) = \sigma^2(S) + \sigma^2(e_i)$ for $i \neq 5$, and $\sigma^2(Y_5) = \sigma^2(S) + \sigma^2(C) + \sigma^2(e_i)$.

For the scaled scores $X_i = Y_i/\sigma(Y_i)$, if we suppose as before that $\sigma^2(S) + \sigma^2(e_i) = 1$,

$$\text{corr}(X_1, X_i) = \begin{cases} \sigma^2(S) & i = 2, 3, 4, 6, \ldots, n; \\ \dfrac{\sigma^2(S)}{[1 + \sigma^2(C)]^{\frac{1}{2}}} & i = 5. \end{cases}$$

Thus the autocorrelation sequence makes an isolated decrease at $i = 5$. This signals that the corresponding expected score sequence, after remaining constant from Occasions 2 through 4, regresses toward the mean from Occasions 4 to 5. Subsequently, however, it egresses to its initial value.

Some More Complex Models

Discussion thus far has focused on isolating different potential patterns that might be followed by the expected score sequence. The models that have been introduced for purposes of illustration have been relatively simple and perhaps unrealistic for many applications.

In any real application, a proper analysis of test data would involve a design strategy that incorporated planning, test evaluation, and model identification and estimation, and it is not the purpose of the present article to enter into such an extensive discussion. It may be useful, however, to present a few more complicated illustrations of possible models, to show both the directions such a strategy might take and how our development of the regression (or egression) of expected sequences fits in with such situations. In addition, by showing how similar patterns can be expected to occur with very different models, the difficulties of model identification and the importance of sound experimental design are examined.

All of the models to be discussed are described within the framework introduced earlier, where the raw scores $Y_i = S_i + e_i$, the S_is are true scores, and the e_is are measurement errors (assumed to be uncorrelated with each other and with the true scores). Of course, a situation with no measurement error could be included by assuming $\sigma^2(e_i) = 0$.

Autoregression models. The first class of examples exhibits the pattern of decreasing autocorrelations associated with a simplex correlation matrix (Guttman, 1954; Humphreys, 1960); that is, $\text{corr}(X_i, X_{i+j})$ decreases as j increases. In particular, we first examine a situation in which the true scores evolve in time, perhaps as a result of a changing environment or biology, perhaps as a result of experimenter intervention. Specifically, consider a model in which the true score has no permanent component but in which past changes of environments or interventions have an effect (albeit a decaying one) on present test scores. Consider then the case in which the true scores S_i follow a first-order autoregressive process (Glass et al., 1972; Nelson, 1973). That is, suppose the first score S_1 has $E(S_1) = 0$ and $\sigma^2(S_1) = 1$ and that the remaining scores develop sequentially; given S_1, $S_2 = \Theta S_1 + C_2$; given S_2, $S_3 = \Theta S_2 + C_3$; given S_i, $S_{i+1} = \Theta S_i + C_{i+1}$. C_i is viewed as an intervention effect or environmental change specific to the ith occasion. So that the true scores (S_i) have the same scale, we suppose that $0 < \Theta \leq 1$, that $\sigma^2(C_i) = 1 - \Theta^2$, that $E(C_i) = 0$, and that S_1, the C_is, and the e_is are mutually independent (and hence uncorrelated). By substituting repeatedly, we see that the ith true score can be written

$$S_i = C_i + \Theta C_{i-1} + \Theta^2 C_{i-2} + \cdots + \Theta^{i-1} S_1.$$

An easy calculation shows that here,

$$\mathrm{corr}(S_i, S_{i+j}) = \Theta^j,\ j \geq 0,$$

and in particular,

$$\mathrm{corr}(S_1, S_i) = \Theta^{i-1},\ i \geq 1.$$

Then

$$\mathrm{corr}(X_1, X_i) = \frac{\Theta^{i-1}}{[1 + \sigma^2(e_i)]},\ i > 1,$$

and the expected sequence (after Occasion 1) is a multiple of Θx, $\Theta^2 x$, $\Theta^3 x$, and so forth. The sequence is illustrated in Panel a of Figure 6. We can see here that in the special case of no measurement error $[\sigma^2(e_i) = 0]$, the "incorrect" sequence is the "correct" one, and there is continued regression toward the mean. But such cases need not exhibit regression all the way to the mean eventually, as the following variation on this model shows.

Suppose the unscaled scores are

$$Y_i = S + S_i + e_i,$$

where S is independent of the S_is and e_is, and the S_is behave exactly as did the S_is described earlier; that is, they follow a first-order autoregressive model. Let $X_i = Y_i/\sigma(Y_i)$, where $\sigma^2(Y_i) = \sigma^2(S) + 1 + \sigma^2(e_i)$ here. Then $\mathrm{cov}(Y_i, Y_{i+j}) = \sigma^2(S) + \Theta^j$ and

$$\mathrm{corr}(X_i, X_{i+j}) = \frac{\sigma^2(S) + \Theta^j}{\sigma^2(S) + 1 + \sigma^2(e_i)}$$

$$\mathrm{corr}(X_1, X_i) = \frac{\sigma^2(S) + \Theta^{i-1}}{\sigma^2(S) + 1 + \sigma^2(e_i)},\quad i > 1.$$

Here the expected sequence, given $X_1 = x$, decreases from x toward

$$\frac{\sigma^2(S)}{\sigma^2(S) + 1 + \sigma^2(e_i)} \cdot x,$$

thus manifesting regression toward but not to the mean, as shown in Panel b of Figure 6. We suppose here that $\Theta > 0$, but the model also makes mathematical sense if $\Theta < 0$. Such a value of Θ would correspond to negative correlation between successive scores, and the autocorrelation sequence (and thus the expected score sequence) would decrease in damped oscillations about zero, as in Panel c in Figure 6.

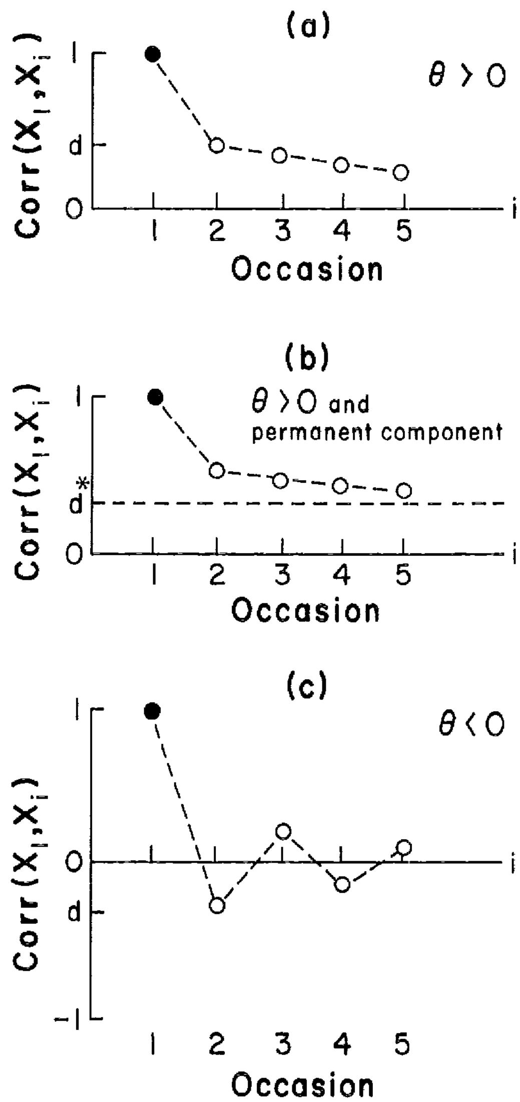

Figure 6. Expected sequences for autoregression models. {Here $d = \Theta \cdot [1 + \sigma^2(e_i)]^{-1}$ and $d^* = \sigma^2(S) \cdot [\sigma^2(S) + 1 + \sigma^2(e_i)]^{-1}$.}

Accumulating advantage models. We next consider a class of models that exhibits the second nonconstant autocorrelation pattern discussed earlier, namely, increasing autocorrelations (and thus egression from the mean after the second occasion). Earlier examples reflected a variety of characteristics, including the persistence of initial nontransitory true

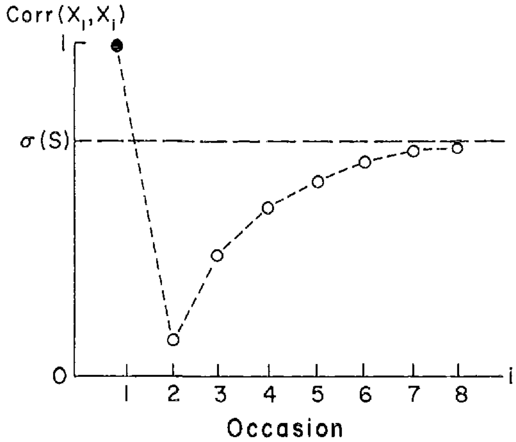

Figure 7. Autocorrelation sequence for the accumulating advantage model.

score components and, in the case of the autoregressive model, serially correlated true scores. But although the true scores in the autoregressive model are serially correlated (and high true scores on one occasion tend to be followed by high true scores on the succeeding occasion), they are regressive in that this component of the score tends itself to regress toward its mean. That model does not represent a situation of accumulating advantage, in which "the rich get richer and the poor get poorer," but rather a situation in which opportunities are available independently of one's ability and past gains are "taxed" and past losses are partly compensated on each occasion. An autoregressive model possesses a memory of the past, but it fades with time. In the next example, this is not so: Advantage does truly accumulate, and the increasing pattern may occur.

We suppose in this example that an individual initially has true score $S_1 = S$ and that after each occasion his or her true score increases by an amount $a \cdot S$, where $a > 0$. (i.e., The true score increases by aS if $S > 0$; if $S < 0$ it decreases by $a|S|$. The rich get richer and the poor get poorer.) Then the true score on Occasion i is $S_i = S + (i-1)aS = [1 + (i-1)a]S$, and the raw test score on Occasion i is

$$Y_i = S_i + e_i,$$
$$= [1 + (i-1)a]S + e_i,$$

where S and e_i are independent (and hence uncorrelated) and $\sigma^2(e_i)$ is constant (e.g., the same test is administered on each occasion).[5] Here

$$\text{cov}(S_1, S_i) = [1 + (i-1)a]\sigma^2(S)$$

and

$$\sigma^2(S_i) = [1 + (i-1)a]^2\sigma^2(S).$$

Suppose, to simplify calculations, that $\sigma^2(S) + \sigma^2(e_i) = 1$, so $\sigma^2(Y_1) = 1$. Then if $X_i = Y_i/\sigma(Y_i)$,

$$\text{corr}(X_1, X_i) = \frac{[1 + (i-1)a]\sigma^2(S)}{\{[1 + (i-1)a]^2\sigma^2(S) + \sigma^2(e_i)\}^{\frac{1}{2}}},$$
$$= \frac{\sigma^2(S)}{\{\sigma^2(S) + [1 + (i-1)a]^{-2}\sigma^2(e_i)\}^{\frac{1}{2}}}.$$

Since $a > 0$, $[1 + (i-1)a]^{-2}$ decreases toward zero as i increases, and corr (X_1, X_i) increases toward $\sigma(S)$, as shown in Figure 7.

The result indicates that the accumulation of S_i makes e_i a progressively less important component of the scaled score X_i, since $\sigma^2(S)$ increases while $\sigma^2(e_i)$ stays constant. The reason $\text{corr}(X_1, X_i)$ does not increase toward one is that e_1 does not become a less important component of X_1; we may write the limit of $\text{corr}(X_1, X_i)$ as $\sigma(S) = [1 - \sigma^2(e_1)]^{\frac{1}{2}}$, in fact.

In this accumulating advantage model, the true scores evolve deterministically, and there is egression from the mean. To show that deterministic evolution need not produce increasing autocorrelations and to show how all ordinal patterns (constant, increasing, and decreasing) may be encompassed in a single model, let us consider a variant of the accumulating advantage model. Suppose

$$S_i = \Theta^i S,$$

and so

$$Y_i = \Theta^i S + e_i,$$

where S might be thought of as an initial true score and Θ is a change parameter. If $0 < \Theta < 1$, then a positive true score decays

[5] For the model given, changes in true score are functionally related to the initial base: $S_i - S_{i-1} = aS$. However, the same pattern of autocorrelations can be obtained from models in which true score changes are only positively correlated with S, for example, $S_i - S_{i-1} = aS + U_i$, where the U_i are independent.

geometrically with time; if $\Theta > 1$, it grows exponentially with time; the reverse is true for negative true scores. The case $\Theta = 1$ is just the classical measurement model (Figure 2).

Then an easy calculation gives

$$\mathrm{cov}(Y_1, Y_i) = \Theta^{i+1}\cdot\sigma^2(S)$$
$$\sigma^2(Y_i) = \Theta^{2i}\sigma^2(S) + \sigma^2(e),$$

and for $i > 1$,

$$\begin{aligned}\mathrm{corr}(X_1, X_i) &= \mathrm{corr}(Y_1, Y_i),\\ &= \frac{\Theta^{i+1}\sigma^2(S)}{\{[\Theta^2\sigma^2(S) + \sigma^2(e)][\Theta^{2i}\sigma^2(S) + \sigma^2(e)]\}^{\frac{1}{2}}},\\ &= \frac{\Theta\sigma^2(S)}{\{[\Theta^2\sigma^2(S) + \sigma^2(e)][\sigma^2(S) + \Theta^{-2i}\sigma^2(e)]\}^{\frac{1}{2}}}.\end{aligned}$$

Clearly, then, if $\Theta > 1$, $\mathrm{corr}(X_1, X_i)$ increases toward $\Theta\sigma(S)/[\Theta^2\sigma^2(S) + \sigma^2(e)]^{\frac{1}{2}}$ (see Figure 4), and if $0 < \Theta < 1$, $\mathrm{corr}(X_1, X_i)$ decreases toward zero (see Figure 3). If $\Theta = 1$, the sequence is stationary, as we have seen (Figure 2). This model illustrates that the decreasing pattern can arise from different suppositions, as, for example, from the progressive (and deterministic) decay of an initial true score or from a stochastic disappearance of initial conditions, as in the autoregressive model. Additional explanations could be formulated, but these should suffice to illustrate the range of phenomena of concern.

Discussion and Conclusions

Summary

We have tried to present a systematic examination of regression toward the mean with explicit attention paid to certain important design and model characteristics that impinge directly on the interpretation of the regression phenomenon.

We remind the reader that the mathematical statements presented earlier concerning the nature of regression effects on different circumstances rest, as do all such developments, on certain assumptions and conventions. The sequences of expected score values developed stemmed from consideration of individuals initially selected as having score $X_1 = x$. As mentioned earlier, repeated or subsequent selection would result in different expected outcomes than the ones presented here. We also assumed that errors (e_i) are independent of true scores (S_i) and of previous observed scores (X_{i-1}). Some of the models we discussed permitted X_i to depend on S_{i-1} by allowing S_i to depend on S_{i-1}. What if the previous test score (X_{i-1}) influenced the present true score, as might be the case if people are told their scores? Then S_i and e_{i-1} can be correlated, since S_i and $S_{i-1} + e_{i-1}$ are. The effect of this and other departures from the assumptions, where plausible, needs to be examined, but that is beyond the scope of the present discussion.

One feature of primary significance in our treatment of regression effects is the focus on observation sequences that extend over more than two occasions of measurement. Historically, discussions of the effects of regression toward the mean are based largely on two measurements (either two variables measured on one occasion or one variable measured on two occasions). We believe that research on change processes will be best served by theoretical concepts and empirical inquiry extending beyond the two-occasion case and thus have developed a multiple-occasion perspective on the nature of regression effects. Examination of regression effects in a broader temporal perspective shows that although traditional conceptions are essentially accurate in predicting what will occur between Occasions 1 and 2, proper expectations for scores at a third and later occasions are not so evident. We return to some design implications of this fact.

A second important feature of our presentation is an explicit concern with the underlying processes (deterministic or stochastic) assumed to represent the observed data and how variations in the nature of the scores lead to different expected regression effects. Expected score sequences were presented for a variety of models that might be assumed to represent the kinds of data obtained from experimental designs common to social and behavioral science research. Included were classical measurement theory models, both with and without intervention effects superimposed on true scores, and autoregression models, both with and without a permanent true score component.

To emphasize matters of traditional concern to behavior change researchers, the mathematical formulation was developed so that the underlying structure of a sequence of measurements could be characterized by the nature of patterns of correlations between scores on Occasion 1 and on subsequent occasions [$\text{corr}(X_1, X_i)$, $i = 2, 3, \ldots, n$]. Varying the nature of these autocorrelation patterns and extending the sequence of observations to more than two occasions of measurement provided the variety of expected regression effects noted.

There are several possible patterns for $\text{corr}(X_1, X_i)$. For didactic purposes, we concentrated on three in particular to illustrate a range of possible regression effects. Included were the situations in which (a) $\text{corr}(X_1, X_i)$ decreased as i increased (as observations became more removed in the sequence), (b) $\text{corr}(X_1, X_i)$ increased as i increased, and (c) corr (X_1, X_i) remained stable as i increased. For all three of these general cases, some regression toward the mean is expected between Occasions 1 and 2. Only for the first situation, however, is further regression expected. For the second situation, the expected score sequence from Occasions 2 to 3 and later actually shows egression from the mean. For the third situation, the expected regression effects are dissipated after Occasion 2, and barring other influences, no further change is expected.

Implications for Design and Control

By understanding more clearly the nature of the regression phenomenon in multiple-occasion data in relation to other influences affecting scores, one may engineer designs to control for the expected regression effects. For example, under the classical measurement error model, regression effects are expended in one measurement occasion after selection has occurred. As a consequence, an easy control arrangement is to use Occasion 1 as a selection point and to begin charting change for the groups selected at the second occasion. Intervention effects, however, depending on when in the sequence after selection the intervention is introduced, may lead to sequences that involve regression toward the mean and then egression from the mean. The point is that when the underlying structure of the score sequence (the model) has been identified, the regression effects are predictable, and thus initial selection, intervention, and measurement can be more usefully arranged in the design of studies focusing on the assessment of change.

Students of behavioral change should not think of regression toward the mean as a univocal phenomenon with straightforward, unalterable effects. Rather, our examination reveals that if one is to anticipate and take into account the effects of regression in analyzing and evaluating change data, explicit consideration must be given to the nature of the process assumed to generate those data. Designs can be arranged to deal more effectively with the confounding of substantive change by expected regression effects, if the underlying model can be specified and the lengths of the observation sequences adjusted.

Statistically, whether the models can be identified and estimated depends on the design. Typically, one could analyze collections of individual's histories and not simply compare means of treatment and control groups (even randomly constituted ones). The latter course would often be valid but would not make the most efficient use of the data; real effects might be masked by averaging or might remain undetected because of inability to take account of the correlation to estimate better the variability of means.

Finally, developmentalists are encouraged to think beyond the confines of one or two approaches (e.g., use of classical test theory model, two occasions of measurement, and simplexlike correlation patterns) as they attempt to elaborate further the nature of change across the life span. We believe that developmental researchers should continue to develop and formulate research designs that exhibit the unique characteristics and strengths of their fields. In the case of developmental psychologists, for example, this involves the study of multiple-occasion change. Recognizing, for instance, the significance of measurement error in a two-occasion situation is unquestionably important. However, letting concepts based on two-occasion models become a stumbling block to developmental research would be unfortunate.

References

Baltes, P. B., Nesselroade, J. R., & Cornelius, S. W. Multivariate antecedents of structural change in development: A simulation of cumulative environmental patterns. *Multivariate Behavioral Research*, 1978, *13*, 127–152.

Baltes, P. B., Nesselroade, J. R., Schaie, K. W., & Labouvie, E. W. On the dilemma of regression effects in examining ability-level-related differentials in ontogenetic patterns of intelligence. *Developmental Psychology*, 1972, *6*, 78–84.

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. *Life-span developmental psychology: Introduction to research methods.* Monterey, Calif.: Brooks/Cole, 1977.

Campbell, D. T., & Erlebacher, A. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), *Disadvantaged child* (Vol. 3). New York: Brunner/Mazel, 1970.

Campbell, D. T., & Stanley, J. E. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching.* Chicago: Rand McNally, 1963.

Clarke, D. B., Clarke, A. M., & Brown, R. I. Regression to the mean—A confused concept. *British Journal of Psychology*, 1960, *51*, 105–117.

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

Frederiksen, C. H., & Rotondo, J. Time series models and the study of longitudinal change. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development.* New York: Academic Press, 1979.

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

Galton, F. Typical laws of heredity. *Nature*, 1877, *15*, 492–495, 512–514, 532–533.

Glass, G. V, Willson, V. L., & Gottman, J. M. *Design and analysis of time-series experiments.* Boulder: University of Colorado, Laboratory of Educational Research, 1972.

Guttman, L. A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences.* Glencoe, Ill.: Free Press, 1954.

Hibbs, D. A., Jr. Problems of statistical estimation and causal inference in time-series regression models. In H. A. Costner (Ed.), *Sociological methodology, 1973–1974.* San Francisco: Jossey-Bass, 1974.

Hotelling, H. Review of *The triumph of mediocrity in business* by H. Secrist. *Journal of the American Statistical Association*, 1933, *23*, 463–465.

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

Humphreys, L. G. To understand regression from parent to offspring, think statistically. *Psychological Bulletin*, 1978, *85*, 1317–1322.

Humphreys, L. G., & Parsons, C. K. A simplex process model for describing differences between cross-lagged correlations. *Psychological Bulletin*, 1979, *86*, 325–334.

Kenny, D. A. *Correlation and causality.* New York: Wiley-Interscience, 1979.

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

Nelson, C. R. *Applied time series analysis.* San Francisco: Holden-Day, 1973.

Nesselroade, J. R., & Bartsch, T. W. Multivariate experimental perspectives on the construct validity of the trait-state distinction. In R. B. Cattell & R. M. Dreger (Eds.), *Handbook of modern personality theory.* Washington, D.C.: Hemisphere, 1977.

Thorndike, R. L. Regression fallacies in the matched groups experiment. *Psychometrika*, 1942, *7*, 85–102.

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

Received November 19, 1979 ■