Anthony S. Bryk, Huron Institute Herbert I. Weisberg, Huron Institute

1.0 Introduction

In recent years the federal government has mounted several large-scale evaluations of the effectiveness of various educational programs. For a variety of practical, ethical, and political reasons, strict adherence to the canons of experimental research design has generally been impossible. Our interest in the general problem of analyzing nonexperimental data has grown out of our involvement in two major evaluations: Head Start Planned Variation and Follow Through. Both of these studies were quasi-experiments involving the comparison of several curricular models, but with no random assignment of subjects to treatments.

In the last few years there has been much controversy over how to draw valid inferences from such quasi-experiments, and, indeed, whether it is even possible to do so. Tucker, Damarin, and Messick (1966) have suggested residual analysis. Cronbach and Furby (1970) and Werts and Linn (1970) have argued that the General Linear Model and, in particular, analysis of covariance, is the only appropriate approach. More recently, Smith (1972) and Kenny (1975) have suggested that gain scores may in fact be the answer. In addition, several other approaches have been suggested in this context, including direct and indirect standardization (Wiley, 1971), and matching (Rubin, 1973). We too have contributed to the confusion with our value-added analysis (Bryk and Weisberg, 1974, 1976).

On the other hand such writers as Lord (1967) and Campbell and Erlebacher (1970) have pointed to problems in the use of such adjustment strategies. These problems have led some (c.f. Gilbert, Light and Mosteller [1975] and Riecken and Boruch [1974]) to push hard for the implementation of randomization if at all possible.

In trying to understand the extent to which various analysis strategies could provide unbiased estimates of treatment effects, we became frustrated by the absence in the literature of realistic models which focus on the process by which the data are generated. Some educational researchers seem to have valuable intuition about the processes generating quasi-experimental data, but they do not express these ideas in the form of a coherent model. Most mathematically-oriented methodologists on the other hand, seem to accept uncritically the general linear model which has been so useful in many non-educational settings. In thinking about quasi-experimental data, we gradually reached the conclusion that three main factors must be considered: (1) the nature of individual growth; (2) the process by which individuals are assigned to treatment groups; and (3) the nature of the treatment effect.

In education as in many other domains, we are typically examining systems which are fundamentally dynamic and adaptive. Individuals are growing, and changing in interaction with one another and the environment even in the absence of any external intervention. Such phenomena are not explicitly considered in the linear model, but are rather treated as unmeasured nuisance variables.

Second, explicit consideration of the process by which individuals are assigned to groups is almost non-existent in the literature (Kenny [1975] is an exception). Most analysts begin by assuming the convenient fiction that there exist two distinct populations from which the program and control groups are randomly selected. Each population is characterized by a multivariate (usually normal) distribution on the variables of interest. The bias reduction properties of various strategies are examined in terms of the parameters of these two distributions (see for example, Cochran, and Rubin, [1974].

An alternative way to conceptualize this situation is to imagine that we are sampling from a single population of individuals and then assigning them in a non-random fashion to a program and control group. We prefer this latter fiction because it allows us to focus explicitly on the mechanism of assigning subjects to programs.

Third, the treatment effect is traditionally modeled as a constant increment for all subjects, although consideration of "aptitude-treatment interactions" permits some flexibility. We propose to consider models in which the treatment effect may be a function of the individual's own status or even the group context in which he or she is developing.

Our purpose in this paper is to assess the ability of various data analysis strategies to remove biases which result from nonrandom assignment of subjects to groups. To do this, we examine each technique in terms of a mathematical model which represents somewhat realistically the way quasi-experimental data arise. The remainder of this paper consists of four sections.

First, we discuss briefly the various analysis strategies which have been suggested for analyzing educational quasi-experiments. A set of sixteen strategies whose properties are to be explored is presented.

In section 3 we present in detail the mathematical model upon which this investigation is based. Explicitly represented in this general model are specific models of the three main processes mentioned above: individual growth, assignment of subjects to groups, and the treatment effect.

In section 4 we develop some analytical results under the assumption of a relatively simple model. Some of these results relate to growth systems in general and may have broad application in studying developmental processes. Others relate more specifically to the problem of eliminating bias in the analysis of quasi-experimental data.

In section 5 we discuss the computer simulation program based on the mathematical model described in section 3. This program has been used to explore the properties of strategies in more complex situations. Results of some simulation runs are presented.

2.0 Analysis Strategies

In this section we discuss briefly the set of analysis strategies whose properties are to be studied. For simplicity, we consider only the situation where one treatment group is being compared to one control group. We also assume that the information available to us consists of the pre-test, the post-test, and the age for each individual. Some methods make use of other "covariates" besides pre-test and age. We are including these two because (1) they are generally quite important in educational contexts and (2) the inclusion of other variables would add tremendous complexity.

Most techniques for analyzing quasiexperimental data fall under the general heading of <u>linear adjustments</u>. These adjustments have the following general form:

 $\hat{\tau} = \overline{Y}_{2p} - \overline{Y}_{1p} - \beta (\overline{Y}_{1p} - \overline{Y}_{1c}) - \gamma (\overline{a}_{1p} - \overline{a}_{1c})$

where

 τ =true treatment effect

 $\overline{Y}_{2p}(\overline{Y}_{2c}) = \text{post-test} \text{ mean for program}$ (control) group

- $\overline{Y}_{1p}(\overline{Y}_{1c})$ =pre-test mean for program (control) group
- $\overline{a}_{1p}(\overline{a}_{1c})$ =mean age at pre-test time for program (control) group

The coefficients β and γ characterize the particular linear adjustment method. For example, if $\beta = 0$ and $\gamma = 0$, our estimate of τ is the simple difference of posttest means; while if $\beta = 1$ and $\gamma = 0$, we have gain scores. In all, thirteen such linear adjustment strategies have been listed in Table 1.

In addition to the linear adjustments, we also consider two forms of adjustment by subclassification (Cochran, 1968) which is a type of <u>standardization</u> (Wiley, 1971). The basic idea is to estimate the post-test means the treatment groups would have had if the pretest distribution of the treatment group were the same as that of the controls. To approximate this, class boundaries are found which divide the pre-test distribution for the control group into six approximately equal size classes (in terms of frequency). Subjects in all groups are assigned to a class on the basis of these pre-test boundaries. Post-test means are computed for each pre-test group. These six means are then proportionally weighted by the number in the cell to form an overall mean which is compared with the control group mean. The difference between the two is our subclassification estimate of the treatment effect.

The second subclassification technique is based on both the pre-test and age. We have elected to divide each variable into three classes, resulting in a nine-cell design.

Another strategy considered is a variant of the value-added approach which we have developed recently (Bryk and Weisberg, 1974). This method does not utilize the control group information per se. Rather than using the relationship between pre-test and post-test, the value-added technique utilizes the individual's pre-test status and age to project a post-test status in the absence of intervention. The actual post-test is then compared to the projected posttest. The average difference across subjects within the treatment group is an estimate of the treatment effect. In its application here, we use the ordinary

least squares regression of the pre-test on age as the basis for our projections.

Another approach to analysis which has received some attention is <u>matching</u> (c.f. Rubin, 197^{\downarrow}). Matching is, however, most effective as a component of the design rather than the analysis. <u>Post hoc</u> matching is possible, but its effectiveness generally depends critically on having a control group much larger than the treatment group, so that good matches can be found. These factors, we felt, would make a fair comparison of matching with the other analysis strategies mentioned above very difficult.

3.0 The Model

In this section we present a general mathematical model to represent the way in which quasi-experimental data arise in many educational settings involving growth. As in most models of dynamic processes, we shall employ the differential equation as our basic conceptual tool. As mentioned above, the model has three main components; representing the processes of natural growth, the effect of a treatment, and the assignment of individuals to treatment groups.

3.1 Natural Growth

We assume that each individual is characterized by a growth rate at any point in time which would apply in the absence of the experimental program. This growth rate may in general be a function of one or more parameters and of the individual's current status. Let $G_1(t)$ represent individual i's <u>natural</u> <u>growth function</u> in terms of some dimension of interest. Moreover, let $\pi_1 =$ $(\pi_{1_1}, \dots, \pi_{1_k})$ represent a vector of parameters characterizing individual in growth rate. Then G can be defined as that function which satisfies a differential equation of the form

$$\frac{\partial G_{i}}{\partial t} = g_{i}(\pi_{i}, G_{i}(t)) \qquad (3.1)$$

Of course there are many possible candidates for g_i . We will present four which seem particularly useful. The first two are selected because of their simplicity. The latter two suggest themselves because they can model a broad range of non-linear growth, and have seen some use in modeling certain developmental phenomena, such as human physical growth (Bock et al, 1973).

These are given by

$$\frac{\partial G_{i}}{\partial t} = \pi_{1i}$$
(3.2)

 $\frac{\partial G_{i}}{\partial t} = \pi_{1i} + \pi_{2i}t \qquad (3.3)$

$$\frac{\partial G_{i}}{\partial t} = \pi_{1i}(1 - G_{i}) \qquad (3.4)$$

$$\frac{\partial G_{i}}{\partial t} = \pi_{1i} G_{i}(1-G_{i})$$
 (3.5)

In analyzing the first three of these models, it is convenient to define t_{01} as the <u>onset time</u> for individual i. That is we assume

$$g(\pi_{i}, G_{i}(t)) = 0$$

for t < t_{oi1} (3.6)

and that g has one of the above forms for $t > t_{01}$. For the fourth case, it is convenient to define $t_{.s_1}$ as the time at which half of "full" growth is reached. With this notation we can express the natural growth curves as

$$G_{i}(t) = \pi_{1i}(t-t_{oi})$$
(linear)
$$G_{i}(t) = \pi_{i}(t-t_{oi})$$
(3.7)

$$\begin{array}{c} 1 & 1 & 0 & 1 \\ (quadratic) & (3.8) \\ G_{*}(t) = 1 - e^{-\pi} 1 & (t - t_{01}) \\ \end{array}$$

$$G_{i}(t) = 1$$

$$\frac{1}{1+e^{-\pi} i (t-t_{.5i})}$$
(logistic)
(3.10)

So far we have considered the growth for each individual. What about the population? To model the process of selecting a sample from a given population, we propose to consider the π_i 's as random variables with theoretically reasonable distributions. The actual choice of specific theoretical distributions for the growth parameters, will depend on the substantive meaning of the parameters themselves.

We shall assume that different parameters are independent. Particularly at the extremes of the distribution, this assumption is perhaps not strictly valid. We do not judge this to be a serious enough weakness, however, to warrant the additional complexity of joint growth parameter distributions.

To model some situations and because it is often used as a covariate in developmental research, it is useful to have a representation of each individual's age. To accomplish this, we do not generate t_{oi}, the onset time directly, but as the sum of time of birth and onset age. Let

 b_i = time of birth for individual i

 δ ,=onset age for individual i

Then

 $t_{o1} = b_1 + \delta_1$ (3.11)

and

 $a_i(t)=t-b_i=age at time t$ (3.12)

Finally, to make our model of growth more realistic, we allow the possibility of a random component to the growth. The systematic natural growth discussed so far may be considered as the smoothed trend underlying an observed time series. The random component reflects in part the instability characterizing our assessment instruments which is commonly termed measurement error. The random component also reflects, however, char-acteristics of individuals. While conceptually it may be useful to consider individual growth over a period of time as a smooth curve, --i.e., the systematic component--in reality, the individ-ual growth also consists of discontinuities, instability, and real variation. Thus, the random component is meant to include measurement error, but is not limited exclusively to it.

Mathematically we represent the random component by

 $R_{i}(t) = random component of growth$ where

 $E(R_{1}(t)|G) = 0$ (3.13)

Cov
$$(R_{1}(t_{1}), R_{1}(t_{2})) = 0$$
 (3.14)

The observed score for individual i at time t is then represented by

$$Y_{i}(t) = G_{i}(t) + R_{i}(t)$$
 (3.15)

3.2 Treatment Effect

So far we have developed a model to represent growth of a particular group on a particular dimension. Now suppose that during some specified period of time, a subset of the group is to be exposed to an experimental program. We put off until the next section the question of how this group is to be selected, and focus here on representing the treatment effect for this group. First let

t =start of the program 1 (time of pre-test)

t_end of program (time of post-test)

We assume that the program operates between times t and t , and measurements are obtained at these times.

We conceive of the program as effecting an alteration in the natural growth rate. We define a treatment effect rate,

$$\frac{\partial I_{1}}{\partial t} = 0 \text{ for } t < t \qquad (3.16)$$
$$=h(\varepsilon_{1},G_{1}(t)) \text{ for } t_{1} \leq t \leq t_{2}$$
$$=0 \qquad \text{ for } t > t_{2}$$

where ε is a vector or parameters characterizing individual i. The treatment effect rate can in principle be a function of one or more parameters as well as the current growth status. By proper choice of the h function, we can model a broad range of educational paradigms.

We can now define a treatment augmented growth rate,

$$\frac{\partial S_{i}}{\partial t} = \frac{\partial b_{i}}{\partial t} + \phi_{i} \frac{\partial I_{i}}{\partial t}$$
(3.17)

where ϕ_i =l if i is in the program group 0 if i is in the control group

and a treatment augmented growth curve,

$$S(t) = \int \frac{\partial S_1}{\partial t}$$
(3.18)

We define the observed score at time t by

$$Y_{i}(t)=S_{i}(t)+R_{i}(t)$$
 (3.19)

and for any individual the treatment effect at time ${\tt t}$ is given by

$$\tau_{1} = S_{1}(t_{2}) - G_{1}(t_{2}) \qquad (3.20)$$

For example, in the simple case of a linear growth system with a constant treatment effect rate,

$$\frac{\partial I}{\partial t} \mathbf{i} = \varepsilon$$
$$\frac{\partial S_{\mathbf{i}}}{\partial t} = \Pi_{\mathbf{i}} + \varepsilon$$

This yields upon integration,

$$S_{1}(t) = \Pi_{2}(t-t_{01}) + \varepsilon(t-t_{1})$$

for $t_{1} \le t \le t_{2}$

and a constant treatment effect across the subjects at t of

$$\tau = \varepsilon(t_2 - t_1) \tag{3.22}$$

3.3 Assignment Process

In this section we discuss the model to represent the assignment process. Let us assume that the total sample consists of n individuals, and that, of these, n are to be assigned to the program group and n to the control group $(n_c + n_p = n)$.

We want to be able to represent a general situation where each individual's probability of being assigned to the program group may be random or may to varying degrees depend on his growth parameter values. Moreover, rather than basing selection directly on values of the parameter, we prefer to view selection as sorting the total sample according to relative values. Thus, our algorithm will incorporate the idea that those n individuals who are highest in terms of some criterion (which may depend on their growth characteristics) will be assigned to the control group, and the n_p lowest to the program group.

One might argue that we also ought to consider assignment on the basis of sociological or demographic variables. Kenny (1975), using a path analysis model, takes just this approach. While clearly these variables may operate to determine who receives services -- for example, income qualifications for compensatory education programs--they are not operant variables in our causal models of human development. Rather, we view these gross observables as reflections of the normally unmeasured individual growth curve parameters. When these growth parameters are available, however, assignment to groups on this basis ex-tracts all of the useful information normally supplied by the gross proxy variables.

The assignment of units to the program and control groups will be determined through the use of an assignment variable, A_i . Let us define

$$A_{i} = \sum_{k=0}^{K} w_{k} \pi_{ki}^{*} + w_{K+1} d_{i}$$

where $\pi_{oi} = t_{oi}$ (3.23)

$$\pi_{ki}^{*} = \frac{\pi_{ki} - \mu_{\pi k}}{\sigma_{\pi k}} = \text{growth parameters in a}$$

standardized (unit)
 $\sigma_{\pi k}$ metric (3.24)

and

$$w_{K+1} = \frac{1}{k=0}^{K} |w_{k}|$$
 (3.26)

The assignment variable, A_1 , is a linear combination of two components. The first component is itself a linear combination of the growth parameters. These parameters are expressed now in standardized form, to adjust for potentially different metrics. The second component is a random normal deviate, generated independently for each subject. Any set of weights, [w where (k=1...K+1)], can be specified. This permits total flexibility to model situations where assignment is at random--w_{k+1}=1--to where assignment might be based totally on starting point--w_0=1. One question this permits us to explore is, "How much randomness is good enough?"

Once we have chosen a set of weights, w_k, and sampled for each subject an R₁, we can compute an A₁. If we rank order the subjects on the basis of their values on A₁, we can then assign the bottom $\frac{n_p}{n_T}$ X100 percent to the program group, and the remainder to the control group. In other than pure random assignment situations, this should be a reasonable replica of real-world selection processes.

Recalling the definition of ϕ_i from the previous section, we can write our assignment algorithm mathematically as

 $A_i \leq A(n_p) \implies \phi = 1$

 $A_1 > A_{(n_p)} \implies \phi = 0$

where $A_{(k)}$ represents the k-th order statistic out of the sample of size n.

4.0 Analysis of the Linear Growth System

4.1 The Linear Model We seek to explore in our work the implications of taking a developmental or growth perspective when examining quasiexperimental data. As described above, we assume an underlying growth system from which we have snapshots at times t₁ and t₂, the pre and post test respectively.

For the linear system,

$$\frac{\partial G}{\partial t} = \pi_i$$

the rate of growth for each subject i is some constant value π_i , but in general the values of the π_i 's will vary across subjects to form some distribution. Thus, we have

$$G_{i}(t) = \pi_{i}(t-t_{oi})$$
 where (4.1)
 π_{i} = growth rate for subject i
 t_{oi} = date of growth onset

In terms of analysis of this model, let us consider the simplest case where the treatment effect rate is a constant, ε , for all subjects

$$\frac{\partial I}{\partial t} = \varepsilon \qquad \begin{array}{c} t_1 \leq t \leq t_2 \\ o \text{ elsewhere} \end{array}$$

Thus,

$$\frac{\partial S_1}{\partial t} = \pi_1 + \phi_1 \varepsilon$$
 and, (4.2)

$$S_{i}(t) = \pi_{i}(t-t_{oi}) + \phi_{i}\varepsilon(t-t_{1})$$

$$t_{1} \leq t \leq t_{2}$$
(4.3)

At post test time t_2 , for subjects in the control group,

$$S_{ic}(t_2) = \pi_i(t_2-t_{oi})$$
 and

for subjects in the program group,

$$S_{ip}(t_2) = \alpha + \Pi_i(t_2 - t_{oi}), \text{ where } \alpha \in \varepsilon(t_2 - t_1)$$

The full model presented in section 3, also includes a random component $R_1(t)$. Since we are at present only considering the simple classical measurement model (errors have zero mean, constant variance, and are independent of systematic growth) this added complexity would provide little insight. Thus, we will focus in this section only on systematic growth $S_1(t)$.

4.2 Properties of the Linear System One question we are most interested

in exploring is the behavior of the basic statistics--means, variances, covariances --for these growth systems. This is an important issue since almost all of the analysis strategies presented in section 2 are simple functions of these basic statistics.

Let

$$E(\pi) = \mu_{\pi}$$
, $Var(\pi) = \sigma_{\tau}^{2}$ (4.4)
 $E(t_{o}) = \mu_{\tau}$, $Var(t_{o}) = \sigma_{\tau}^{2}$ (4.5)

Assuming that $t_{\rm O}$ and π are independent, it follows that

$$E[G(t)] = \mu_{\pi}(t-\mu_{\tau})$$
 (4.6)

To find the covariance between $G(t_1)$ and $G(t_2)$, we apply the definition of the covariance and equation (4.6) to obtain

Theorem: For the linear growth system with π and t_0 independent we have

$$E[G(t)] = \mu_{\pi}(t-\mu_{\tau})$$

$$Cov(G_{1}(t_{1})_{1}G(t_{2})) =$$

$$(t_{1}-\mu_{\tau})(t_{2}-\mu_{\tau})\sigma_{\pi}^{2}+\mu_{\pi}^{2}\sigma_{\tau}^{2}+\sigma_{\pi}^{2}\sigma_{\tau}^{2} \qquad (4.7)$$

Several useful corrolaries follow from this result. For example, if $t_1=t_2=t$, equation (4.7) reduces to

$$Var[G(t)] = (t-\mu_{\tau})^{2}\sigma_{\pi}^{2} + \mu_{\pi}^{2}\sigma_{\tau}^{2} + \sigma_{\pi}^{2}\sigma_{\tau}^{2}$$
(4.8)

Second, let us define p_{12} as the correlation between $G(t_1)$ and $G(t_2)$. From (4.9) and (4.10) we find

$$\rho_{12} = \frac{(t_1 - \mu_{\tau})(t_2 - \mu_{\tau})\sigma_{\pi}^2 + \mu_{\pi}^2 \sigma_{\tau}^2 + \sigma_{\pi}^2 \sigma_{\tau}^2}{\{[(t_1 - \mu_{\tau})^2 \sigma_{\pi}^2 + \mu_{\pi}^2 \sigma_{\tau}^2 + \sigma_{\pi}^2 \sigma_{\tau}^2]_{1/2}} [(t_2 - \mu_{\tau})^2 \sigma_{\pi}^2 + \mu_{\pi}^2 \sigma_{\tau}^2 + \sigma_{\pi}^2 \sigma_{\tau}^2]\}^2$$
(4.9)

From equation (4.9) we can study the behavior of the correlation coefficient as each of the parameters varies. Since many longitudinal studies have looked at the pattern of correlations between time points, equation (4.9) may be helpful in interpreting these results.

4.3 Application to Adjustment Strategies in the Analysis of Quasi-Experimental Data

Our primary interest is in examining growth systems where there is non-random assignment of subjects to groups. This results in a different distribution across the two groups for the growth parameters. The consequences for commonly employed analysis strategies are the focus of this investigation.

Let
$$\sum_{i=1}^{n} G_{i}(t) \sum_{i=1}^{n} \Pi_{i}(t-t_{oi})$$
$$G_{i}(t) = \frac{i-1}{n} = \frac{i=1}{n}$$
$$t_{o} = \frac{\Sigma t_{oi}}{n} \qquad \Pi_{i} = \frac{\Sigma \Pi_{i}}{n}$$

Then it is easy to show that

G.(t)=
$$\Pi$$
.(t-t₀.)+ $\frac{\sum_{i=1}^{n} (\pi_{i}-\pi_{i})(t-t_{0})}{n}$ (4.10)

Thus, the mean growth curve $G_{\cdot}(t)$ for a sample is simply the natural growth curve in the parameter means plus the sample

covariance. Moreover, if ${\rm I\!I}$ and ${\rm t_O}$ are assumed independent we have

 $E[G.(t)] = \Pi.(t-t_0.)$ (4.11)

Suppose now that for the program group

 $S.p(2) \cong \alpha + \Pi.p(t_{o.p})$

and for the control group,

$$S_{c}(2) \cong \Pi_{c}(t_{2}-t_{0,c})$$

Now since ${\rm II}$ and ${\rm t}_{\rm O}$ are independent, it follows that II. and ${\rm t}_{\rm O}.$ are independent. Moreover,

$$E(\Pi.) = \mu_{\pi}$$
$$E(t_{0.}) = \mu_{\tau}$$

Thus,

$$E[S._{p}(2)-S._{c}(2)] = \alpha + \mu_{\pi p}(t_{2}-\mu_{\tau p}) - \mu_{\pi c}(t_{2}-\mu_{t c}) \qquad (4.12)$$

and we have a bias which depends on $\mu_{\pi p}$, $\mu_{\pi c}$, $\mu_{\tau p}$, and $\mu_{\tau c}$ and will generally be non-zero. In fact, if non-random assignment results in <u>either</u> $\mu_{\pi p} \neq \mu_{\pi c}$ or $\mu_{\tau p} \neq \mu_{\tau c}$, bias will result.

Now let us consider the consequences of using a linear adjustment strategy in an attempt to obtain an unbiased estimate of α . The general form of a linear adjustment using the pre-test as a covariate is

$$\mathbf{\hat{a}} = S_{.p}(2) - S_{.c}(2) - \beta(S_{.p}(1) - S_{.c}(1))$$
(4.13)

And,

$$E(\mathbf{\hat{a}}) = \alpha + \mu_{\pi p}(t_{2} - \mu_{\tau p}) - \mu_{\pi c}(t_{2} - \mu_{\tau c})$$
$$-\beta \mu_{\pi p}(t_{1} - \mu_{\tau p}) + \beta \mu_{\pi c}(t_{1} - \mu_{\tau c})$$
(4.14)

Solving for β we find that linear adjustment will be unbiased if and only if $\beta = \beta^*$, where

$$\beta^{*} = \frac{\mu_{\pi p} (t_{2}^{-} - \mu_{\tau p}) - \mu_{\pi c} (t_{2}^{-} - \mu_{\tau c})}{\mu_{\tau p} (t_{1}^{-} - \mu_{\tau p}) - \mu_{\pi c} (t_{1}^{-} - \mu_{\tau c})}$$
(4.15)

This is a rather remarkable result. It represents the theoretically "correct" adjustment coefficient under our model.

One special case is of particular interest. Suppose $\mu_{\pi_D} = \mu_{\pi_C}$. Then the above expression reduces to $\beta^*=1$. That is, if assignment is effectively on the basis of t_o but not on I, then use of gain scores provides an unbiased estimate of α .

A rather straightforward question to ask at this point is what happens if we calculate an estimate of β in one of the usual ways, and substitute the result in (4.13)? Unfortunately, the answer is not so straightforward.

Under the usual linear model, it is assumed that

$$E(S_{1}(2)) = \alpha + \beta S_{1}(1)$$

conditional on any given value of $S_1(1)$. This leads to the result that any unbiased estimator of $\hat{\beta}$ can be substituted for β , and the linear adjustment will remain unbiased.

Under our model, it does not make sense to condition on S(1). Thus, if an estimator based on the data is used in 4.12), its distribution may be quite complex. This is one reason why we developed the simulation model described in the next section.

While the exact bias is complex, we can obtain some useful insights analytically. From equations (4.7) and (4.8) we can obtain an expression for the population regression coefficient of S(2) on S(1):

$$\beta = \frac{(t_1 - \mu_{\tau})(t_2 - \mu_{\tau})\sigma_{\pi}^2 + \mu_{\pi}^2 \sigma_{\tau}^2 + \sigma_{\pi}^2 \sigma_{\tau}^2}{(t_1 - \mu_{\tau})^2 + \mu_{\tau}^2 \sigma_{\tau}^2 + \sigma_{\pi}^2 \sigma_{\tau}^2}$$
(4.16)

The sample regression coefficient for the program group will be a consistent (though not necessarily unbiased) estimator of this expression for each group. Let us call the expression for the program group β_p and for the control group β_c . In general $\beta_p \neq \beta_c$. Thus, the data analyst may detect heterogeneity of regression and be unwilling to proceed with the analysis of covariance. If he does calculate an estimate on the basis of either regression, or some combination (e.g., the usual pooled within-group estimator), it is hard to say in general how badly he will miss the correct β given in (4.15).

There is one special, though rather trivial, case where $\beta_c = \beta_p = \beta^*$. This occurs when $\sigma_{\tau c}^2 = \sigma_{\tau p}^2 = 0$ and $\mu_{\tau p} = \mu_{\tau c} = \tau_{o}$. That is, all individuals in both groups "fan out" from the same starting point t_o . In that case $\beta_c = \beta_p = \beta^* = \frac{t - t_o}{t_1 - t_o}$. Thus, linear adjustments can be expected to perform well in situations approximated by this model.

On the other hand, suppose $\mu_{TP} \neq \mu_{TC}$ but that $\mu_{\pi P} = \mu_{TC}$. We showed above that gain scores ($\beta^{*}=1$) is appropriate. However, in this case $\beta_{C} \neq \beta_{P}$ and moreover both β_{C} and β_{P} are greater than 1. We can say in general that if our linear model is approximately correct, we would expect the data to reflect different regressions in the two groups, and that linear adjustment strategies will not be very effective in removing bias. Moreover, we have shown that while there are special situations where linear adjustment does well and special situations where gain scores are appropriate, there is no situation where both can be expected to work well.

5.0 Computer Simulation

5.1 The Simulation Program

In the previous section we developed some analytic results for the linear case. The component for the assignment model, however, is particularly difficult to manipulate analytically. Moreover, conclusions cannot easily be drawn for the more complicated growth and treatment effect models or for the non-linear adjustment strategies such as standardization. Thus, we resort to simulation.

We have developed a general FORTRAN IV computer program to generate and analyze data under a broad range of conditions as described in sections 2 and 3. Our main purpose is to assess the ability of various adjustment strategies to remove bias in quasi-experimental studies involving growth systems. For each set of conditions (a growth system, a treatment effect model, and an assignment procedure), we define the theoretical treatment effect by

$$\tau = E_p[S(t_1) - G(t_1)]$$

where the expectation is over the distribution of individuals assigned to the program group. For each analysis strategy there is an estimator $\hat{\tau}$. Thus, we can define the bias of any analysis strategy as

BIAS
$$(\hat{\tau}) = E(\hat{\tau}) - \tau$$

Two features built into the simulation program are worthy of comment. First, we place certain constraints on our simulations in order to generate "reasonable" data. The basic statistics for each generated data set are examined according to a set of reasonableness criteria: If a data set fails to pass any one of the criteria, it is rejected from the simulation run. If a particular simulation run (i.e. a growth system, an assignment procedure, and a treatment effect model) has more than 5% rejected data sets, the run is terminated. In this manner, we seek to guarantee that condidtions utilized in the simulation may be a reasonable reflection of data sets seen in common practice.

Second, a flexible simulation termination procedure is built into the program. A maximum number of runs (RMAX = 350) is specified, but we can terminate the simulation when sufficient accuracy is achieved.

5.2 Simulation Results

Table 2 presents the two treatment effects models for which simulation runs were carried out. Each of these was combined with each of eight different growth conditions (see table 3) and six assignment models (see table 4). For each interactive model and each of the 16 strategies, the bias is presented in table 5. Lack of space prevents us from presenting the constant treatment effect results. However, the following observations are based on the entire set of runs.

Several interesting patterns can be observed in the simulation results:

(1) As predicted in section 4, nonrandom assignment on the basis of t_0 but not on I, results in substantial bias for all techniques except gain scores, (strategy 2).

(2) The value-added technique (strategy 10) performs best in general over the range of conditions considered here. The Belson ANCOVA with reliability correction based on both age and pre-test (strategy 15) also does well. The need to know the pre-test reliability, however, is a limitation of this approach.

(3) When the treatment effect interacts with individual characteristics, strategies based primarily on the pre-post relationship in the control group only, (strategies 7, 8, 14, 15) perform consistently better than the corresponding strategies based on relationships in both groups (strategies 5, 6, 12, 13). This makes sense intuitively, since the interactive treatment can change the pre-post relationship in the program group but not the control group.

(4) When assignment of subjects to groups is based on either (but not both) of the growth parameters (assignment sets 1 and 2) or on both operating to create bias in the same direction (assignment set 6), all adjustment strategies reduce the initial bias.

(5) When the assignment is based on both I and t_0 in such a way that they tend to create biases in opposite directions (assignment sets 3, 4, 5), no adjustment strategy does well. In fact, almost all strategies considered do worse across all the growth conditions.

Bibliography

1. Bock, D.R., Wainer, H., Peterson, A., Thissen, D., Murray, J. and Roche, Alex. A Parameterization for Individual Human Growth Curves, <u>Human Biology</u>, 1973. Vol. 45.

2. Bryk, A.S. and Weisberg, H.I. A New Approach to Analyzing Quasi-Experimental Data: Value-Added Analysis. <u>Proceedings</u> of the American Statistical Association, <u>Social Statistics Section</u>, 1974.

3. Bryk, A.S. and Weisberg, H.I. Value-Added Analysis: A Dynamic Approach to the Analysis of Quasi-Experiments. Journal of Educational Statistics, 1976 (to appear).

4. Campbell, D.T. and Erlebacher, A. How Regression Artifacts in Quasi-Experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful. In J. Hellmuth (Ed.), <u>Compensatory Education:</u> <u>A National Debate</u>. Vol. III. <u>Disadvantaged</u> <u>Child</u>. New York: Brunner-Mazel, 1970.

5. Cochran, W.G. The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies. <u>Biometrics</u>, 24, 1968.

6. Cochran, W.G. and Rubin, D.B. Controlling Bias in Observational Studies: A Review. <u>Sankhya</u>, The Indian Journal of Statistics, Series A, Volume 35, Part 4, 1973.

7. Cronbach, L.J. and Furby, L. How We Should Measure "Change"--Or Should We? <u>Psychological Bulletin</u>. 1970, 74.

8. Gilbert, J.P., Light, R.J. and Mosteller, F. Analyzing Social Programs: An Empirical Base for Social Policy. In Arthur Lunsdale and Carl Bennett (Eds.). Experiment and Evaluation. Academic Press, 1975.

9. Kenny, D. A Quasi-Experimental Approach to Assessing Treatment Effects in the Nonequivalent Control Group Design. Psychological Bulletin, Vol. 76, 1975.

10. Lord, F.J. A Paradox in the Interpretation of Group Comparisons. <u>Psycho-</u> <u>logical Bulletin</u>, 1967.

11. Riecken, H.W. and Boruch, R.F. (Eds.) Social Experimentation: A Method for Planning and Evaluating Social Intervention. New York, Academic Press, 1974.

12. Rubin, D.B. The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies. <u>Biometrics</u>, Vol. 29, No. 1, 1973.

13. Smith, P. Estimating Linear Trends in Contaminated Quasi-Experiments of Pretest-Posttest Design: or: Gambling on Linear Models, 1972. (unpublished)

14. Tucker, L.R., Damarin, F. and Messick, S. A Base-free Measure of Change. Psychometrika, 1966, 31.

15. Werts, C.E. and Linn, R.L. A General Linear Model for Studying Growth. Psychological Bulletin, 1970, 73.

16. Wiley, D.E. Approximations to <u>ceteris</u> paribus: Data Adjustment in Educational Research. Paper presented at International Research Seminar on Learning and the Educational Process, Munich, Germany, 1971. <u>Table 1</u> The Lineup of Analysis Strategies

1) Difference of Post-test means

- 2) Raw gains
- 3) Standardized gains

4) Residual analysis (pre-test only)

- 5) ANCOVA-pooled within grp. regression(pre-test)
- 6) ANCOVA with reliability correction
- 7) Belson ANCOVA-control grp. regres. (pre-test)
- 8) Belson ANCOVA w. reliability correction
- 9) Subclassification (pre-test)
- 10) Value-added
- 11) Residual analysis (pre-test and age)
- 12) ANCOVA (pre-test and age)
- 13) ANCOVA w. reliability correction (pre-test & age
- 14) Belson ANCOVA (pre-test and age)

15) Belson ANCOVA w. reliability correction

16) Subclassification (pre-test and age)

Table 2:Treatment Effect ModelsTypeTreatment

constant ment for	incre-	$\frac{\partial I}{\partial t} = \epsilon$	
subjects			

interactive treatment effort: maximal $\frac{\partial I}{\partial t} = \epsilon_{i}[100-G_{i}(t)]$

								_
Table	3:	Grov	vth	Para	ameter	Conditions	for	-
	_	the	Lin	ear	System	n		
Q + 1-								

Growth Conditions	1	2	3	4	5	6	7	8
μ _d	9	9	9	9	9	9	9	9
$\sigma_{\rm d}^2$.6	• 3	.6	• 3	.6	• 3	.6	• 3
μ _π	2.5	2.5	2.5	2.5	5.0	5.0	5.0	5.0
σ_{π}^2	.4	.4	.8	.8	.4	.4	.8	.8

Distributions of d and π are assumed to be normal

Distribution of b is uniform (0,6) for all cases.

<u>Table 5</u> (next page) Simulation Results

LINEAR SYSTEM WITH AN INTERACTIVE TREATMENT EFFECT

	w = 0	0 w=	.20 w	= .8		$w_{t0} =20$	w = .20	$w_d = .6$	•
Anglucie	Distributio	ne for the	erowth na	rameters	1	Distribu	ntions for	the Growth Par	ameters
Strategy	1 2	3 4	5 6	7 8	1 1	21 32	4 5	6 7	8
1	-2.82 -3.02 -4	4.21 -4.04	-2.85 -2.8	5 -4.11 -4.00	51-5.32	-6.92	-6.84 -7.0	09 -7.14 -8.31	-8.57
2	-1.85 -1.93 -2	2.81 -2.66	-1.93 -1.8	9 -2.83 -2.70	-2.45	-3.52	-3.48 -2.	51 -2.42 -3.35	-3.49
3	-1.23 -1.20 -1	1.61 -1.44	-1.77 -1.7	2 -2.44 -2.20	550	52	45 -1.3	78 -1.55 -1.84	-1.82
4	-2.01 -2.14 -2	2.65 -2.50	-2.11 -2.0	8 -2.91 -2.78	3 -3.35	-3.67	-3.66 -3.5	52 -3.49 -3.96	-4.11
5	-1.75 -1.82 -	.232 -2.16	-1.97 -1.9	3 -2.77 -2.62	2 -2.18	-2.34	-2.31 -2.3	79 -2.68 -3.19	-3.26
6	-1.56 -1.60 -2	2.06 -1.87	-1.92 -1.8	8 -2.70 -2.54	-1.62	-1.72	-1.63 -2.	56 -2.44 -2.92	-2.98
7	-1.63 -1.70 -2	2.19 -2.01	-1.85 -1.8	1 -2.60 -2.44	-1.85	-2.01	-1.96 -2.2	26 -2.12 -2.58	-2.63
8	-1.42 -1.44 -1	1.91 -1.71	-1.80 -1.7	6 -2.52 -2.3	5 -1.24	-1.34	-1.23 -2.0	00 -1.86 -2.28	-2.32
9	-1.43 -1.49 -2	2.04 -1.92	-1.67 -1.6	7 -2.51 -2.34	4 -2.15	-2.61	-2.35 -2.0	62 -2.43 -3.20	-3.26
10	7290 -1	1.27 -1.12	588	8 -1.40 -1.3	.90	-1.57	-1.49 -1.2	21 - 1.31 - 1.42	-1.65
11	-2.04 -2.18 -2	2.46 -2.28	-2.11 -2.0	1 - 2.62 - 2.30	-3.45	-3.74	-3.80 -3.	50 -3.44 -3.76	-3.86
12	-1.52 -1.55 -1	1.82 -1.61	-1.78 -1.6	4 - 2.24 - 1.93	3 -2.13	-2.26	-2.31 -2.0	52 - 2.4/ - 2.82	-2.82
13	8968 -	1.1582	-1.51 -1.2	5 - 1.86 - 1.47	2 -1.22	-1.32	-1.30 -2.	14 - 1.88 - 2.24	-2.11
14	-1.40 -1.45 -1	1.03 - 1.42	-1.08 -1.3	4 -2.08 -1.70	-1.80	-2.01	-2.05 -2.	14 - 1.90 - 2.29	-2.31
15	(-30 - 32 -)	.0143	-1.37 -1.0	5 - 1.03 - 1.13		-1.00	92 -1.0	03 - 1.33 - 1.07	-1.07
136 mm	$\frac{2}{2}$	2.03 -2.30	-2.23 -2.1	5 -2.94 -2.0.	-2.0/	-3.29	-3.27 -3.4	41 -3.20 -3.81	-4.03
<u> </u>	' <u>W. A</u>	20 17 =	. 40 W.	= .4		W .=2	w = 0	w, = .8	
1	-6.13 -6.27 -9	9.62 -9.83	-4.37 -3.9	7 -7.42 -7.6	5 -1.42 -	1.51 -1.50	-1. 22 -2.1	85 - 2.96 - 2.93	-3.17
2	-5.38 -5.39 -7	7.68 -7.84	-5.37 -5.3	6 -7.61 -7.7	5 .05 -	10 - 12	030	051102	08
3	-4.94 -4.85 -6	6.22 -6.29	-5.53 -5.6	0 -7.66 -7.79	1.05	.91 1.08	1.23	45 .42 .91	.97
4	-5.46 -5.51 -7	7.46 -7.61	-5.46 -5.4	7 -7.68 -7.8	224 -	.3601	.12	566426	33
5	-5.30 -5.29 -7	7.09 -7.21	-5.36 -5.3	4 -7.61 -7.7	.25	.09 .37	.51	1522 .16	.14
6	-5.15 -5.10 -6	6.75 -6.82	-5.41 -5.4	1 -7.62 -7.73	.54	.39 .63	.800	0008 .32	.31
7	-5.24 -5.22 -7	7.02 -7.17	-5.45 -5.4	7 -7.63 -7.70	5 .47	.30 .58	.71 .3	22 . 17 .59	.58
8	-5.09 -5.02 -6	6.67 6.77	-5.51 -5.5	5 -7.64 -7.73	.80	.64 .86	1.03 .	38.33.78	.78
9	-5.07 -5.12 -6	5.99 -7.20	-5.32 -5.3	3 -7.50 -7.70	.51	.39 .81	.96 .3	24 .21 .51	.53
10	-2.55 -2.63 -3	3.49 -3.97	-2.78 -2.7	2 -3.63 -3.68	3 .19	.18 .38	.02 .0	0311 .29	20
11	-6.92 -7.32 -8	8.32 -8.69	-5.86 -6.0	8 -7.52 -7.63	368 -	.9571	80:	788966	93
12	-4.94 -4.82 -5	5.87 -5.77	-5.00 -4.7	9 -6.48 -6.22	2 .05 -	.2005	072	233311	29
13	-3.29 -2.39 -4	4.08 -3.32	-4.39 -3.7	6 -5.59 -4.8	.23 -	.10 .07	03	1022 .02	21
14	-4.92 -4.84 -5	5.63 -5.45	-5.23 -5.1	6 -6.67 -6.2	.25 -	.02 .11	.08 .	12 .02 .27	.06
15	-2.2481 - 2	2.7189	-4.43 -3.7	2 -5.43 -4.1	L .44	.09 .24	.14 .2	27 .13 .40	.10
	-6./5 -/.0/ -8	8.76 -9.09	-6.26 -6.5	$\frac{8 - 8.55 - 8.9}{2}$	<u> </u>	$\frac{18}{18}$	$\frac{.11}{}$	223214	0/
	wto	.40 w π	40 W	2		"t ₀ " .40	"π - •2	"d ".4	
	1 2	3 4	5 ¹ 6 ²	78	1	2 3	4 5	$1 6^2 7^3$	84
1	-4.55 -4.37 -8	8.11 -7.91		-3.81 -3.43	/05 -	.01 -1.64	-1.75		
2	-5.86 -5.79 -8	8.30 -8.17		-8.32 -8.1	5 -2.86 -	2.85397	-4.02		
3	-6.68 -6.72 -8	8.44 -8.38		-9.76 -9.83	3 -4.82 -	4.89 -5.97	-6.05		
4	-6.35 -6.37 -8	8.46 -8.36		-9.31 -9.32	2 -4.37 -	4.45 -5.49	-5.56		
5	-6.08 -6.06 -8	8.35 -8.26		-9.01 -8.9	9 -3.35 -	3.38 -4.92	-4.99		
6	-6.35 -6.38 -8	8.39 -8.32		-9.28 -9.28	3 - 3.93 -	4.02 -5.37	-5.47		
7	-6.17 -6.14 -8	8.35 -8.27		-9.35 -9.3	L -3.58 -	3.53 -5.07	-5.13		
8	-6.45 -6.47 -8	8.38 -8.33	-	-9.64 -9.63	2 - 4.20 -	4.20 -5.54	-5.64		
9	-5.88 -5.88 -8	8.19 -8.17		-9.12 -9.24	4 -3.21 -	3.14 -4.59	-4.65		
10	-2.64 -2.78 -3	3.78 -4.11		-4.22 -3.9	(-1.12 - 1.12)	1.37 - 1.83	-1.92		
11	-9.75 -10.61-1	1.5/-12.15		-9.72 -9.92	2 - 6.0	0.09 - 0.78	-7.25		
12	-6.21 -6.30 -	/./5 -/.58		-8.10 -/.//		2.00 - 3.44	-3.19		
13	-4.42 -3.50 -	5.65 - 4.49		-/.29 -0.30	-2.20 -	-1.58 - 2.70	-2.02		
14		7.00 - 7.03		-7 42 -5 9	2 - 2 + 18 - 2	-3.49	- 64		
15	-1.00 .37 -3	3.00 .44 1 52_12 07		-10 42 -10 9	8 - 47 - 47	$4 \ 97 \ -5 \ 59$	-6 10		
10	-0.90 -9.91 -1	.1. J2-12.07		-10.42 -10.9	q , -		0.10		
152 runs	$^{2}62$ runs.				¹ 16 ru	ns. $^{2}20$ rur	ns. ³ 55 ru	ns. ⁴ 23 runs.	
	Table 4: N	lon Rando	m Assignm	ent Condit	ions Co	nsidered	in this	Simulation	
	Weights		-	Process D	escript	ion of In	dividual	s Assigned	
₩t.	01 W _{π4}	Wd.		t	o the C	ontrol Gr	oup		
_	2 N	6		early eta	rtere				
<u> </u>	- v	.6		fast grow	ers				
U	4 1	.2		late star	ters, h	ut fast o	rowers		
	2 4	.4		late star	ters. b	ut fast g	rowers		
	4.2	.4		late star	ters. b	ut fast g	rowers		
	2.2	.6		early sta	rters a	nd fast g	rowers		

.