AN INDIVIDUAL GROWTH MODEL PERSPECTIVE FOR EVALUATING EDUCATIONAL PROGRAMS

Judith F. Strenio, Harvard University & The Huron Institute
Anthony S. Bryk, Harvard University & The Huron Institute
Herbert I. Weisberg - The Huron Institute

1. Introduction

In evaluating educational programs it is often not possible to conduct a rigorous randomized experiment. Estimates of program effects must be based on uncontrolled observational studies or partially controlled quasi-experiments. These studies generally involve comparisons of treatment group performance with that of a nonequivalent control group. Because the groups being compared are not completely equivalent prior to the intervention, observed outcome differences may reflect these prior differences in addition to the treatment effect. That is, estimates of the effect based on a direct comparison of post-treatment measures will be biased.

Traditional analysis methods use an adjustment approach in attempting to reduce this bias. Pretreatment differences between a treatment and control group are modelled. Statistical techniques based on the model are used to compensate, or adjust for these initial differences when comparing outcome data for the two groups.

One of the major potential sources of bias in such studies derives from the fact that individuals grow at different rates in the absence of a treatment. Thus the effects of a program may be confounded with natural growth, or maturation. In a previous paper (Bryk and Weisberg, 1977) we have detailed some of the problems encountered by traditional statistical methods for analyzing quasi-experiments when individuals are growing.

In this paper we discuss an alternative analysis strategy based on a projection approach. Utilizing information in the data set on individual growth, the strategy involves explicitly projecting the growth the program group would have achieved without any intervention. Actual growth can then be compared with projected growth; the difference is termed the value-added by the program.

The value-added technique was originally presented (Bryk and Weisberg, 1976) in terms of a very restricted model: all individuals were assumed to have identical growth rates. In this paper we extend the model to consider variable individual growth rates.

Note that, unlike adjustment techniques, the value-added approach does not necessarily assume the availability of data on an untreated control group. It is essentially a single-group design. On the other hand, it does require a sufficient combination of theory and empirical data to estimate natural growth. In this paper we assume that subjects are tested twice: once prior to the program, at a pretest time that we denote by t_1, and once at the end, at a posttest time t_2. Our objective is to estimate the average increment at the posttest time which is attributable to program experience.

2. Model and Rationale

We assumed that each individual's growth is a linear function of age. Let us denote by $a_i(t)$ the age of individual i at time t. Individuals are assumed to vary in terms of growth rate π_i and onset age δ_i (the age when non-negligible growth begins). Moreover, they are assumed to be sampled from a population in which π and δ are independently distributed with means μ_π, μ_δ and variances σ^2_π, σ^2_δ.

This model represents the simplest situation that incorporates varying individual growth. While too simple to represent realistically many educational processes, linear growth may be a reasonable approximation over a short term even if long-term growth has a more complex form.

For the present, we will also assume that δ and π are distributed to children independently of their age at pretest. That is, the older children are not more likely to have started growth at a particular age than the younger, nor are they growing slower or faster. We examine this assumption, and some consequences of its violation, in a later section.

Finally, we assume that observed growth $Y_i(t)$ is made up of two components: systematic growth $G_i(t)$ and a random noise component $R_i(t)$ determined by the particular circumstances at time t. Our basic model can be represented as

$$(1)\quad G_i(t) = \begin{cases} \pi_i[a_i(t) - \delta_i] & \text{for } a_i(t) \geq \delta_i \\ 0 & a_i(t) < \delta_i \end{cases} \quad \text{and}$$

$$(2)\quad Y_i(t) = G_i(t) + R_i(t)$$

where for all t, t'

$$E\,[R_i(t)] = 0$$

$$\mathrm{Cov}[R_i(t),\, R_i(t')] = 0$$

We also assume that the pretest time t_1 is set so that all subjects have begun to grow by that time. Combining equations (1) and (2) we can write

$$(3)\quad Y_i(t_1) = \pi_i[a_i(t_1) - \delta_i] + R_i(t_1)$$

Let us for convenience define

$$(4)\quad \Delta = t_2 - t_1$$

Then if no treatment were introduced, we would have

$$(5)\quad Y_i(t_2) = \pi_i[a_i(t_2) - \delta_i] + R_i(t_2)$$
$$= G_i(t_1) + \pi_i\Delta + R_i(t_2)$$

In order to model a treatment effect, we assume that over the time interval t_1 to t_2 the treatment increases each subject's growth by an amount v_i (the value-added). The mean and variance of v_i are μ_v and σ^2_v and v is assumed to be uncorrelated with any other variable in the model. Since v_i is a random variable, this model in principle allows for individual effects. Finally, then, we can represent the measured growth that subject i in the program group achieves by time t_2 as

$$(6)\quad Y_i(t_2) = G_i(t_1) + \pi_i\Delta + v_i + R_i(t_2)$$

We take the estimation of μ_v as the object of our analysis.

Before proceeding further with the examination of this model, we present the rationale underlying the method. During the period between pretest and posttest, the average growth for the treatment group is $\overline{Y}(t_2) - \overline{Y}(t_1)$. The expected growth under the model is $\mu_\pi \Delta$. If we knew the value of μ_π, a natural estimator of the value-added would be $\overline{Y}(t_2) - \overline{Y}(t_1) - \mu_\pi \Delta$. So if we have an estimator $\hat{\mu}_\pi$ of μ_π we might use

$$(7) \quad V = \overline{Y}(t_2) - \overline{Y}(t_1) - \hat{\mu}_\pi \Delta$$

From equations (5) and (6) it is clear that any unbiased $\hat{\mu}_\pi$ will yield an unbiased estimator $\hat{\mu}_V$ of μ_V. In this paper we propose to use the ordinary least-squares regression coefficient of $Y(t_1)$ on age. This estimator is simple to compute nd intuitively appealing. In the next section we show that it is unbiased.

3. Examining the Value-Added Method: Properties of $\hat{\mu}_\pi$.

In this section we consider some properties of the least-squares regression coefficient we are proposing as an estimate of μ_π.

Lemma 1: Taking expectation over the distributions of π and δ,

$$E(\hat{\mu}_\pi) = \mu_\pi.$$

Proof: Our model is given by equation (3) with π_i, δ_i, and R_i mutually independent. We can rewrite this equation as

$$(8) \quad Y_i(t_1) = \mu_\pi a_i(t_1) - \mu_\pi \mu_\delta + \{(\pi_i - \mu_\pi) a_i(t_1) - (\pi_i \delta_i - \mu_\pi \mu_\delta) + R_i(t_1)\}$$

This equation is now in the form

$$(9) \quad Y_i = \alpha + \beta a_i + e_i$$

with

$$\alpha = -\mu_\pi \mu_\delta$$

$$\beta = \mu_\pi$$

$$e_i = \{(\pi_i - \mu_\pi) a_i(t_1) - (\pi_i \delta_i - \mu_\pi \mu_\delta) + R_i(t)\}$$

Under our assumptions it is straightforward to obtain

$$(10) \quad E(e_i | a_i) = 0.$$

Thus our model satisfies the usual conditions under which ordinary least squares yields unbiased estimates of α and β. Q.E.D.

We note, however, that the variance of the error term works out to be

$$(11) \quad \mathrm{Var}(\mathrm{error}_i) = a_i^2(t_1)\sigma_\pi^2 - 2a_i(t_1)\sigma_\pi^2 \mu_\delta + \sigma_\pi^2 \sigma_\delta^2 + \sigma_\pi^2 \mu_\delta^2 + \mu_\pi^2 \sigma_\delta^2 + \sigma_R^2.$$

Thus the error variance is a quadratic function of $a_i(t_1)$ and the OLS estimate, though unbiased, will be inefficient. In practice, we might wish to use a generalized least squares procedure to estimate μ_π. Implementing this idea involves some complex problems which we are currently investigating.

We next derive the variance of $\hat{\mu}_\pi$:

Lemma 2:

$$(12) \quad \mathrm{Var}(\hat{\mu}_\pi) = \frac{\sigma_\pi^2 \Sigma A_i^2 a_i^2(t_1) - 2K_2 \Sigma A_i^2 a_i(t_1)}{(\Sigma A_i^2)^2} + \frac{\sigma_R^2 + K_1}{\Sigma A_i^2}$$

Where $K_1 = \mathrm{Var}(\pi\delta) = \mu_\delta^2 \sigma_\pi^2 + \mu_\pi^2 \sigma_\delta^2 + \sigma_\pi^2 \sigma_\delta^2$

and $K_2 = \mathrm{Cov}(\pi, \pi\delta) = \sigma_\pi^2 \mu_\delta$.

Proof: Let $\overline{a}(t)$ be the average age of the program group at time t, and $A_i = a_i(t) - \overline{a}(t)$, noting that $\Sigma A_i = 0$. Then the usual least-squares estimate is given by:

$$(13) \quad \hat{\mu}_\pi = \frac{\Sigma A_i [Y_i(t_1) - \overline{Y}(t_1)]}{\Sigma A_i^2},$$

which simplifies to

$$(14) \quad \hat{\mu}_\pi = \frac{\Sigma A_i Y_i(t_1)}{\Sigma A_i^2} \quad \text{because } \Sigma A_i = 0. \text{ Now}$$

$$(15) \quad \mathrm{Var}(\hat{\mu}_\pi) = \frac{\Sigma A_i^2 \mathrm{Var}[Y_i(t_1)]}{(\Sigma A_i^2)^2}$$

There are no covariance terms, since $\mathrm{Cov}(Y_i, Y_j) = 0$.

Thus we require $\mathrm{Var}[Y_i(t_1)]$ (recalling $Y_i(t_1)$ from equation (3)):

$$(16) \quad \mathrm{Var}[Y_i(t_1)] = a_i^2(t_1)\sigma_\pi^2 + \mathrm{Var}(\pi_i \delta_i) + \sigma_R^2 - 2a_i(t_1)\,\mathrm{Cov}(\pi_i, \pi_i \delta_i)$$

Because we assume π_i and δ_i are independent, we find $\mathrm{Var}(\pi\delta) = K_1$, and $\mathrm{Cov}(\pi, \pi\delta) = K_2$ as given in the statement of the Theorem. Thus equation (11) is indeed the variance of $\hat{\mu}_\pi$. Q.E.D.

This gives the variance of $\hat{\mu}_\pi$ in terms of the parameters of the model. Note that the usual variance of $\hat{\mu}_\pi$ is simply $\frac{\sigma_R^2}{\Sigma A_i^2}$, one term of our variance.

4. Examining the Value-Added Technique: Properties of V.

We now consider some statistical properties of the value-added estimator itself. From equations (5) through (7) we have

$$(17) \quad V = \frac{\Delta}{n}\sum_{i=1}^{n} \pi_i + \frac{1}{n}\sum_{i=1}^{n} v_i + \frac{1}{n}\sum_{i=1}^{n} R_i(t_2) - \frac{1}{n}\sum_{i=1}^{n} R_i(t_1) - \hat{\mu}_\pi \Delta$$

Theorem:

(18) (a) $E(V) = \mu_V$

(b) $$\mathrm{Var}(V) = \frac{\sigma_V^2 + 2\sigma_R^2 - \Delta^2\sigma_\pi^2 + n\Delta^2\,\mathrm{Var}(\hat{\mu}_\pi)}{n}$$

Proof: (a) Apply expectation to both sides of equation (17);

(19) $$E(V) = \frac{\Delta}{n}(n\mu_\pi) + \frac{1}{n}(n\mu_V) + 0 - 0 - E(\hat{\mu}_\pi)\Delta$$

$$= \Delta\mu_\pi + \mu_V - \mu_\pi\Delta$$

$$= \mu_V.$$

(b) Take variances of both sides of equation (17);

(20) $$\mathrm{Var}(V) = \frac{\Delta^2\sigma_\pi^2}{n} + \frac{\sigma_V^2}{n} + \frac{2\sigma_R^2}{n} + \Delta^2\,\mathrm{Var}(\hat{\mu}_\pi)$$

$$- \frac{2\Delta^2}{n}\sum_{j=1}^{n}\mathrm{Cov}(\pi_i,\hat{\mu}_\pi).$$

We already have $\mathrm{Var}(\hat{\mu}_\pi)$ from Lemma 2.

We require $\sum_{i=1}^{n}\mathrm{Cov}(\pi_i,\hat{\mu}_\pi)$:

First, using equation (13),

(21) $$\mathrm{Cov}(\pi_i,\hat{\mu}_\pi) = \frac{\sum_{j=1}^{n} A_j\,\mathrm{Cov}[\pi_i, Y_j(t_1)]}{\sum_{j=1}^{n} A_j^2}.$$

We have

(22) $$\mathrm{Cov}[\pi_i, Y_j(t_1)] = \begin{cases} 0 & \text{if } i \neq j \\ -K_2 + \sigma_\pi^2 a_i(t_1) & \text{if } i = j. \end{cases}$$

So

(23) $$\sum_{i=1}^{n}\mathrm{Cov}(\pi_i,\hat{\mu}_\pi) = \frac{-K_2\sum_{i=1}^{n} A_i + \sigma_\pi^2\sum_{i=1}^{n} A_i a_i(t_1)}{\sum_{i=1}^{n} A_i^2}$$

$$= \sigma_\pi^2$$

Substituting this into equation (20) above and collecting terms, we get the expression in (b) of the Theorem. Q.E.D.

Comments On This Theorem:

1) V is unbiased, because $\hat{\mu}_\pi$ is. It may look as though we are using independent variables with error when we write the model--that is, we want $[a_i(t_1) - \delta_i]$ and we know only $a_i(t_1)$--but the proof of $\hat{\mu}_\pi$'s unbiasedness shows that age alone is valid for estimating μ_π, and we usually know age accurately.

2) If everyone had the same growth rate, and we knew what it was, $\mathrm{Var}(V)$ would be $\frac{2\sigma_R^2}{n} + \frac{\sigma_V^2}{n}$.

The $\frac{\Delta^2\sigma_R^2}{n}$ arises from the differences in growth rate, and the other terms from the estimation of μ_π from data.

5. Testing Significance of V

In practice, we generally wish not only to estimate the treatment effect, but also to test its significance and/or to state a confidence interval. To derive such tests and intervals requires derivation of the distribution of V under various assumptions about the distribution of π, δ, R and V. In the previous section we derived an expression for the variance of V. It is not obvious how to use it in developing the necessary statistical procedures.

While the development of procedures based on the distribution of V is worth pursuing, another general purpose approach may prove useful. The jackknife technique (described in Chapter 8 of Mosteller and Tukey, 1977) can provide both a test statistic and standard error for use in forming confidence intervals. To apply the jackknife to our situation is fairly straightforward. Let $\hat{\mu}_{\pi(\text{all})}$ be the least squares coefficient computed from the whole data set, and let $\hat{\mu}_{\pi(i)}$ be the coefficient computed with only observation i removed from the data. Then for each individual i a pseudo-value V_{*i} is calculated as

(24) $$V_{*i} = Y_i(t_2) - Y_i(t_1) - \hat{\mu}_{\pi*i}\Delta$$

where $\hat{\mu}_{\pi*i} = n\hat{\mu}_{\pi(\text{all})} - (n-1)\hat{\mu}_{\pi(i)}$

The V_{*i} are then treated as data points. Their mean $\bar{V}_*$ provides an unbiased estimate of μ_V, and the standard error allows calculation of a t-statistic with (n - 1) degrees of freedom for testing or interval estimation.

6. Illustrative Example

We take as an example a subset of the data collected to evaluate the Head Start Planned Variation program. We will consider the data on one curricular model for one outcome, the Preschool Inventory (described in Walker, Bane and Bryk, 1973). All children were pretested at ages between 50 and 63 months, with mean age 56.80 months.

The mean pretest score is 14.116 and the mean posttest score is 20.454, out of a possible 32. The mean time between tests is 7.40 months. The least-squares regression coefficient of pretest on age is 0.484. Thus the estimated value-added is given by

$$V = 20.454 - 14.116 - (0.484)(7.40) = 2.756.$$

To test this value for statistical significance, the jackknife procedure was carried out as described above. This resulted in a mean $\bar{V}_*$ of 2.764 which has a standard error of 1.192. The resulting t-value of 2.319 with 96 degrees of freedom is significant at the .05 level.

7. Independence of Age and Individual Growth Characteristics.

The value-added method as applied in this paper uses the cross-sectional relationship between score and age at a particular point in time, t_1, to estimate the mean growth rate for the program group. This approach assumes that individual growth characteristics (reflected by π_i and δ_i in our model) are independent of age. If there exists a systematic relationship between these characteristics and age, then the pretest/age relationship reflects not only individual growth but also the age gradient of δ_i and π_i.

Non-independence can occur in at least two different ways. First, in the population from which individuals are sampled, there may be historical trends causing children born at different times to differ. For example, during the period when Sesame Street was first being introduced, younger children exposed to the program may have had different characteristics from older children not exposed.

Second, even if this stable universe assumption (Kodlin and Thompson, 1958) is true for the population being studied, selection of the experimental sample may introduce an age by characteristic relationship. Criteria of selection may have operated so that younger children tend to have different characteristics from older ones. For example, the youngest children in a Head Start program may be there because they are unusually mature for their ages, possibly entering a bit below the age threshold. The oldest children may be particularly slow, possibly even old enough to enter kindergarten but not really ready.

To understand the effects of these phenomena, we develop a simple model. Let A_i represent the deviation of a subject from the group mean (as before),

$$A_i = a_i(t_1) - \bar{a}(t_1) \; .$$

Let us assume further that the expected values of π and δ are functions of A_i:

$$(25) \quad \begin{aligned} E(\pi_i | A_i) &= f(A_i) \\ E(\delta_i | A_i) &= g(A_i) \end{aligned}$$

To see how this would affect our value-added technique, we look first at $E[Y_i(t_1)|A_i]$, to see what the age versus pretest score graph will look like; that is, what the cross-sectional data will become. We have equation (3) for $Y_i(t_1)$. If we take expectations, substitute for $E(\pi_i|A_i)$ and $E(\delta_i|A_i)$ from equation (25) and rewrite $a_i(t_1)$ as $[A_i + \bar{a}(t_1)]$, we arrive at this result:

$$(26) \quad E[Y_i(t_1)|A_i] = f(A_i)\; [A_i + \bar{a}(t_1) - g(A_i)]$$

We can see that unless we choose some special f and g, or they have some special parameterization, $Y_i(t_1)$ will become a nonlinear function of age. Thus the age versus pretest score graph will show curvature, and we can test for age selection by testing the age by pretest score graph for nonlinearity.

8. Linear Individual Growth Assumption

Another possible problem is that individual growth may be non-linear. With extreme nonlinearity, the linear approximation will not be trust-worthy even in the short term. For example, on a particular test as soon as a subject has thoroughly mastered all items, Y_i flattens out at the perfect score (although the type of skill that had been measured may continue to improve).

If we wish to retain the idea that each subject has different parameters of the growth curve, then this problem becomes very complex. In Bryk (1977) an individual negative exponential growth curve is examined. This is a very appealing model for growth, which has been widely used in biological growth studies. Bryk derives the expected value of $\bar{Y}(t)$ and shows that it is not a negative exponential function of time. More generally, the average of non-linear growth curves, even when taken over subjects the same age, will not trace out the same shaped curve as the individuals are following. This will make model identification difficult when only cross-sectional data are available. But, we with the age-growth dependence problem, at least we can see that the age versus pretest plot will not be linear. So, again, a test for linearity can be used as an indicator of failure to meet the model's assumptions.

9. Directions for Further Research

The use of the ordinary least-squares regression coefficient to estimate μ_π was chosen for simplicity and intuitive appeal. We have shown that it leads to an unbiased estimate of μ_ν. In large samples, this estimator should be quite adequate. With smaller samples, however, it is not clear whether this approach yields estimates that are efficient enough for practical purposes. This question needs to be investigated. It may well be necessary to develop alternative estimation procedures with greater efficiency.

Secondly, the model we have assumed here is the simplest model which incorporated diffential growth rates across individuals. Investigation of more complex models and development of corresponding analysis strategies is needed. For example, models could reflect various kinds of dependence between π and δ, various forms of non-linear growth, and various kinds of age-selection effects.

Finally, a very important research area lies in the attempt to assess individual values of v_i. If we could do this, we would be able to estimate σ_v^2 and the distribution of v. We could also estimate interactions between the v_i and measured covariates. Particularly in this educational context, we are often interested in more than the simple average effect. Rather, we wish to discover which programs help which students, by how much.

In order to achieve this objective, the estimation of the individual v_i seems necessary. To accomplish this, however, more information will be needed. We have gone quite far with only two cross-sections, one as proxy for longitudinal data and the other to gauge progress. The next logical step is to gather more data on the same group, so that we really have, say, four or five data points on each subject. Through the combination of cross-sectional and longitudinal perspectives on the same data set we should be able to estimate more precisely both the mean effect μ_v and other aspects of the distribution of individual v_i's.

References

Bryk, A.S. An Investigation of the Effectiveness of Alternative Statistical Adjustment Strategies in the Analysis of Quasi-Experimental Growth Data. Thesis for Doctor of Education, Harvard University (1977).

Bryk, A.S. and Weisberg, H.I. Value-Added Analysis: A Dynamic Approach to the Estimation of Treatment Effects. Journal of Educational Statistics, Vol. 1, No. 2 (1976).

Bryk, A.S. and Weisberg, H.I. Use of the Nonequivalent Control Group Design When Subjects Are Growing. Psychological Bulletin, Vol. 85, No. 4 (1977).

Kodlin, D. and Thompson, D.J. An Appraisal of the Longitudinal Approach to Studies of Growth and Development. Monographs of the Society for Research in Child Development, Inc., Vol. 23, Serial No. 67, No.1 (1958).

Mosteller, F. and Tukey, J.W. Data Analysis and Regression: A Second Course in Statistics. Reading, MA: Addison-Wesley (1977).

Walker, D.K., Bane, M.J. and Bryk, A.S. The Quality of the Head Start Planned Variation Data. Cambridge, MA: Huron Institute (1973).