DOCUMENT RESUME

ED 210 290 TM 810 931

AUTHOR Bryk, Anthony; Woods, Elinor
TITLE An Introduction to the Value-Added Model and its Use in Short Term Impact Assessment.
INSTITUTION Huron Inst., Cambridge, Mass.
SPONS AGENCY Department of Education, Washington, D.C.
PUB DATE Dec 80
NOTE 53p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Early Childhood Education; *Evaluation Methods; *Federal Programs; *Mathematical Models; Program Evaluation
IDENTIFIERS *Elementary Secondary Education Act Title I; Evaluation Problems; *Value Added

ABSTRACT

This resource book examines the value-added model approach when used in assessing early childhood Title I (ECT-I) programs. The evaluation design must be able to separate program effects from natural maturation. The basic idea behind the value-added model builds on the notion of natural maturation. The major strengths are that it does not require a comparison group or the use of a norm referenced test. The major weaknesses are its usefulness only for the assessment of skills or attributes which show a natural development with age over the duration of the program. Selection procedures may disguise the relationship between age and skill development among a particular group of program participants. Finally, it can require complex statistical calculations. This method attempts to derive a great deal of information from a situation with little data and little external control. The validity of results from the value-added model may be questioned in situations where one wishes to assess the short-term impact of an ECT-I program and the available sample size is relatively small. A possible solution is pooling data across multiple years of the program or across several sites that are implementing similar activities. (DWH)

The Huron Institute

ED210290

AN INTRODUCTION TO THE VALUE-ADDED MODEL
AND ITS USE IN SHORT TERM IMPACT ASSESSMENT

By Anthony Bryk and Elinor Woods

TM 810 531

THE HURON INSTITUTE 123 MOUNT AUBURN STREET CAMBRIDGE MASSACHUSETTS 02138

AN INTRODUCTION TO THE VALUE-ADDED MODEL AND ITS USE IN SHORT TERM IMPACT ASSESSMENT

December 1980

Anthony Bryk and Elinor Woods
with contributions from
Judith Strenio

The Huron Institute
123 Mt. Auburn Street
Cambridge, Massachusetts 02138

FOREWORD

This booklet has been prepared as part of a project sponsored by the United States Education Department (USED) on evaluation in early childhood Title I (ECT-I) programs. It is one of a series of resource books developed in response to concerns expressed by state and local personnel about early childhood Title I programs. The series describes an array of diverse evaluation activities and outlines how each of these might contribute to improving local programs. The series revolves around a set of questions:

- Who will use the evaluation results?
- What kinds of information are users likely to find most helpful?
- In what ways might this information aid in program improvement?
- Are the potential benefits substantial enough to justify the cost and effort of evaluation?

Together, the resource books address a range of issues relevant to the evaluation of early childhood programs for educationally disadvantaged children. The series comprises the following volumes:

- Evaluating Title I Early Childhood Programs: An Overview
- Assessment in Early Childhood Education
- Short-Term Impact Evaluation of Early Childhood Title I Programs
- An Introduction to the Value-Added Model and Its Use in Short-Term Impact Assessment
- Evaluation Approaches: A Focus on Improving Early Childhood Title I Programs
- Longitudinal Evaluation Systems for Early Childhood Title I Programs
- Evaluating Title I Parent Education Programs

The development of this series follows extensive field work on ECT-I programs (Yurchak & Bryk, 1979). In the course of that research, we

identified a number of concerns that SEA and LEA officials had about ECT-I programs, and the kinds of information that might be helpful in addressing them. Each resource book in the series thus deals with a specific concern or set of concerns.. The books and the evaluation approaches they describe do not, however, constitute a comprehensive evaluation system to be uniformly applied by all. Our feasibility analysis (Bryk, Apling, & Mathews, 1978) indicated that such a system could not efficiently respond to the specific issues of interest in any single district at any given time. Rather, LEA personnel might wish to draw upon one or more of the approaches we describe, tailoring their effort to fit the particular problem confronting them.

Finally, the resource books are not comprehensive technical manuals. Their purpose is to help local school personnel identify issues that might merit further examination and to guide the choice of suitable evaluation strategies to address those issues. Additional information and assistance in using the various evaluation strategies are available in the more technical publications cited at the end of each volume, and from the Technical Assistance Centers in the ten national regions.

TABLE OF CONTENTS

I. BACKGROUND AND RATIONALE

Young children are constantly growing and changing over time even in the absence of an educational program. This growth may occur along the very dimensions that are the objectives of an early childhood Title I (ECT-I) program. Thus, in trying to assess the effects of any ECT-I program, the evaluation design must be able to separate program effects from natural maturation. Each of the basic USED evaluation models can accomplish this, but in different ways. For example, with the norm-referenced model the adjustment for natural maturation is built into the norming tables. For any given percentile, as the child grows older the raw score is assumed to rise accordingly.*

The basic idea behind the value-added model builds directly on the notion of natural maturation and the fact that children of different ages will tend to display different test scores. Consider the sample of data shown in Figure 1. Notice that although there is considerable variation in the test scores on the Preschool Inventory measure, the scores tend to be higher for older children. We can summarize this relationship between test score and age by running a line through the scatter of points in such a way that the line is as close as possible, on average, to all of these points. Now there are many different lines that one could run through a scatter of points such as this. It is here that the statistical technique

---

* Unfortunately, as described elsewhere (see the resource book on Assessment in Early Childhood Education, Haney & Gelberg, 1980), there are few adequate norm-referenced tests for use with young children, and even where they exist, the norms may not be appropriate for use in assessing the short-term impact of an ECT-I program. In such situations, the value-added model described in this booklet and the other strategies discussed by Haney are possible alternatives.

or regression analysis is particularly helpful in determining a good choice of a single line. The resulting equation, called the regression line, for

predicted test score = -13.9 + .48 x child's age.

This equation is a statistical description of the relationship between test score and age. Given the age of any child, we can use this equation to predict a test score for that child. For example, if a child were 4 years and 7 months old (55 months) our equation would yield:

predicted test score = -13.9 + .48 (55)

predicted test score = -13.9 + 26.4 = 12.5.

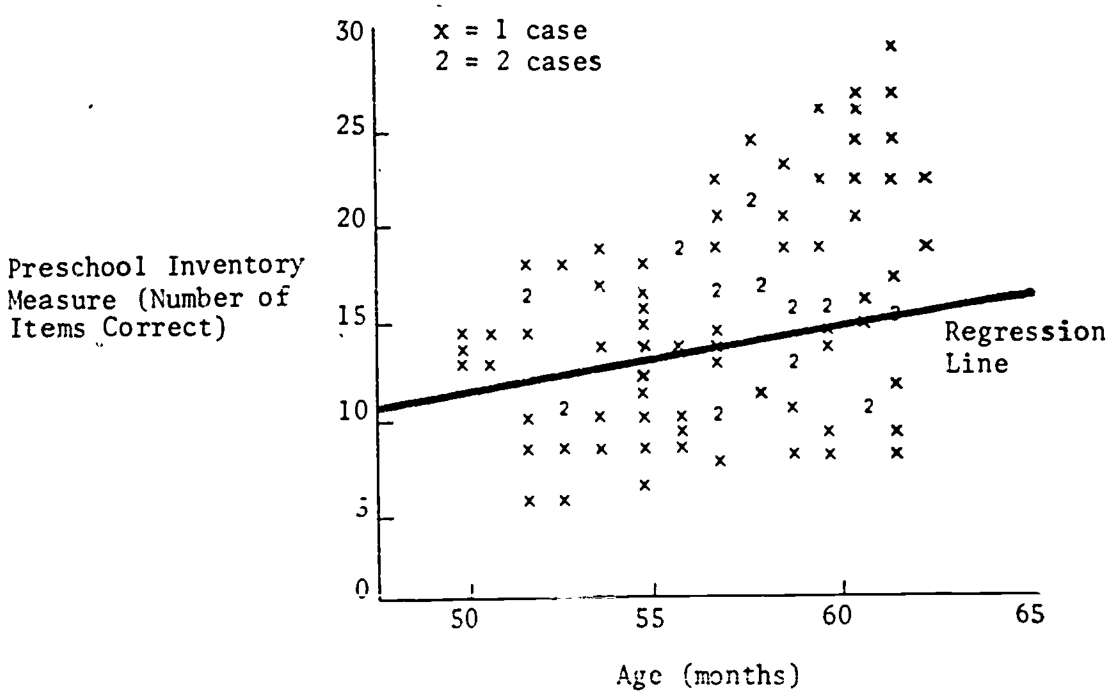

Figure 1. Scatterplot of PSI Pretest by Age (Head Start Planned Variation Evaluation).

Source: Anderson et al., 1980.

Note in Figure 1 that the test scores of children 55 months old tend to center around 12.5, although some scores are considerably different. Similarly, if a child were 5 years old (60 months) our equation would yield:

$$\text{predicted test score} = -13.9 + .48\ (60) = 14.9.$$

Thus, for older children the equation predicts a somewhat higher test score. So our regression (also called prediction) equation captures a fundamental feature of these data -- test scores tend to be higher on average for older children, although clearly not in every case.

The scatterplot and regression line in Figure 1 is a formal way of representing this maturation phenomenon. In particular, the feature of growth over time is captured in what is called the regression coefficient for age or the slope of the regression line. In the equation above, the slope is .48, and it represents the expected gain per month in test score points for any child. We can think of this as the natural growth rate for this group of children in the absence of any special program. For these data, we can expect children to gain, on average, about a half a test score point per month (to be precise, .48 points per month) as a result of natural maturation.

The value-added model builds directly on the relationship between test score and age and the information it contains about natural growth. Using the slope from the regression of pretest scores on age, we can project the growth that children in an ECT-I program would have attained in the absence of that program. This projection serves as the "no-treatment" expectation.

The observed or actual gain under the program is compared with the projected growth under natural maturation. The difference between the two provides an estimate of the effect of the ECT-I program, or what we refer to as the value added by the program.

One of the major differences between the value-added strategy and the other four models for examining the short-term impact of ECT-I programs* is the standard of comparison used for assessing the effects of the program. The norm-referenced model relies on test norms to generate this comparison value. The control-group model uses test data from a group of children who are similar to the Title I recipients but are not receiving Title I services. The special regression model is a variant on the control-group approach. In criterion-referenced approaches the standard of comparison is set by professional judgment. The value-added model, in contrast, does not require a control group, norms, or an externally set criterion. Rather, the relationship of age with pretest scores is used to estimate the natural growth that might have resulted if no special program had been offered during the interval between the pre- and posttest.

---

* See the resource book on Short-Term Impact Evaluation of Early Childhood Title I Programs, Haney (1980).

II. THE BASIC VALUE-ADDED MODEL

We introduce the basic value-added model through an example. The data for this illustration are drawn from a 1977-78 Title I kindergarten program in Iowa. As part of a larger evaluation effort, the staff administered the Boehm Test of Basic Skills as a pretest in the early fall and again as a posttest late in the spring. The actual duration between pre- and posttest was 7 months. By examining the relationship between the pretest scores and children's ages, we can develop an estimate of the natural growth rate in the absence of the program, and use this to estimate the expected gain under natural maturation.

Figure 2 displays the relationship between pretest score and age for these data. Again, the raw scores tend to be higher for older children.

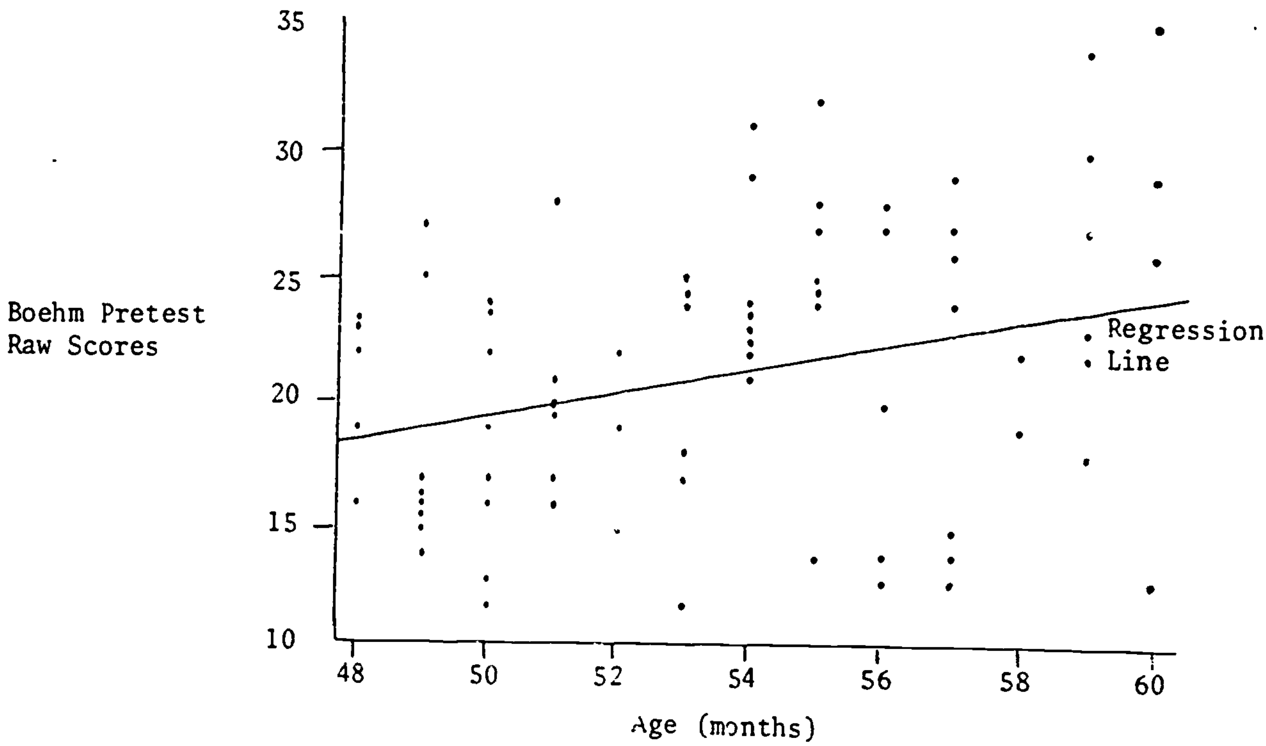

Figure 2. Scatterplot of Boehm Pretest by Age (Iowa ECT-I Data).

Through a simple regression of the Boehm pretest raw scores on age (measured in months at the time of the pretest), we estimate a natural growth rate (the slope of the regression line) of .47 points per month. In other words, the estimated average growth rate for these children in the absence of the Title I program is about half a raw score point per month on the Boehm.

Assuming that the children would continue to mature naturally at the same average rate even without an especially effective program, we can now estimate the expected gain over the course of the program interval due solely to natural maturation:

expected gain (due to maturation)	=	natural growth rate per month (slope of regression line)	x	program duration (measured in months).

For the Iowa Title I kindergarten program,

expected gain	=	.47 raw score points per month	x	7 months	=	3.29 points.

To estimate the program effect, we compare the expected gain with the actual observed gains of children in the program. For the Iowa program, the pretest mean was 21.27, and the posttest mean 30.47, yielding an observed gain (posttest mean minus pretest mean) of 9.20 points. The estimated short-term effect, or value added by the program, is simply

value added = observed gain - expected gain,

which for our illustration yields

$$\text{value added by ECT I program} = 9.20 - 3.29 = 5.91 \text{ points.}$$

If we compare the estimated value added by the program to the expected gain under maturation, we have a natural way to assess the educational significance of the program effect estimate:

$$\text{index of educational significance} = (\text{value added} \;/\; \text{expected gain}) \times 100 \,.$$

For the Iowa data this yields

$$\text{index of educational significance} = (5.91 \;/\; 3.29) \times 100 = 180\%.$$

In this case, the value-added model estimates that the program produced a 180% improvement in the average growth over what would have been expected in the absence of the program.

Thus, the basic value-added model only requires information on each child's pretest and posttest score, the age at pretest, and the duration between pre- and posttest points. Neither a control group nor information on background variables (e.g. demographic characteristics and home environment) is required. If the latter is available, however, more precise program effect estimates are possible, and there is also an opportunity to estimate

program effects for different subgroups of children (e.g. boys vs girls). The next section presents an illustration of this extension of the value-added model.

As is true of all of the short-term impact evaluation models for ECT-I programs, the validity of the program effect estimated by the value-added model depends upon some basic assumptions. For the value-added approach, the key concern is our ability to derive an estimate of the natural growth rate from the observed relationship between children's pretest scores and ages. The estimate derived from the regression of the pretest on age is valid only if we can assume that the average growth rate is stable across children of different ages within the program group (e.g., the older children cannot be the "slow developers"), and that the average growth rate will remain stable over the duration of the program. We refer to these conditions as the stable universe assumption. In a later section, we discuss some of the ways in which this assumption might be violated in ECT-I programs, and the consequences of such violations.

III. THE VALUE-ADDED MODEL WITH BACKGROUND VARIABLES

The basic value-added model estimates an average growth rate for all children that would be expected to occur in absence of the ECT I program. It is reasonable to assume, however, that the natural growth rate varies across children, and that children with different background characteristics (e.g. boys and girls) might have different average growth rates. Thus, it seems natural to extend the value-added model, when background information is available, to estimate different average growth rates for children with different background characteristics.

We introduce this extension of the model with a hypothetical example. Let's assume that in evaluating some ECT-I programs we had data on the child's sex in addition to the basic data on pretest, posttest, age, and program duration. The first step in the analysis would be to plot the pretest-age relationship separately for boys and girls, and to perform separate regressions of the pretest on age for each group. This is illustrated in Figure 3. The regression equation for girls might be:

$$\text{predicted test score} = -45.0 + 1.1 \times \text{age in months}$$

and for boys:

$$\text{predicted test score} = -36.0 + 0.9 \times \text{age in months} .$$

Thus, we now have different natural growth rate estimates (i.e. different regression slopes) for girls (1.1 points per month) and for boys (0.9 points per month). We can use these estimates to compute separate program effect

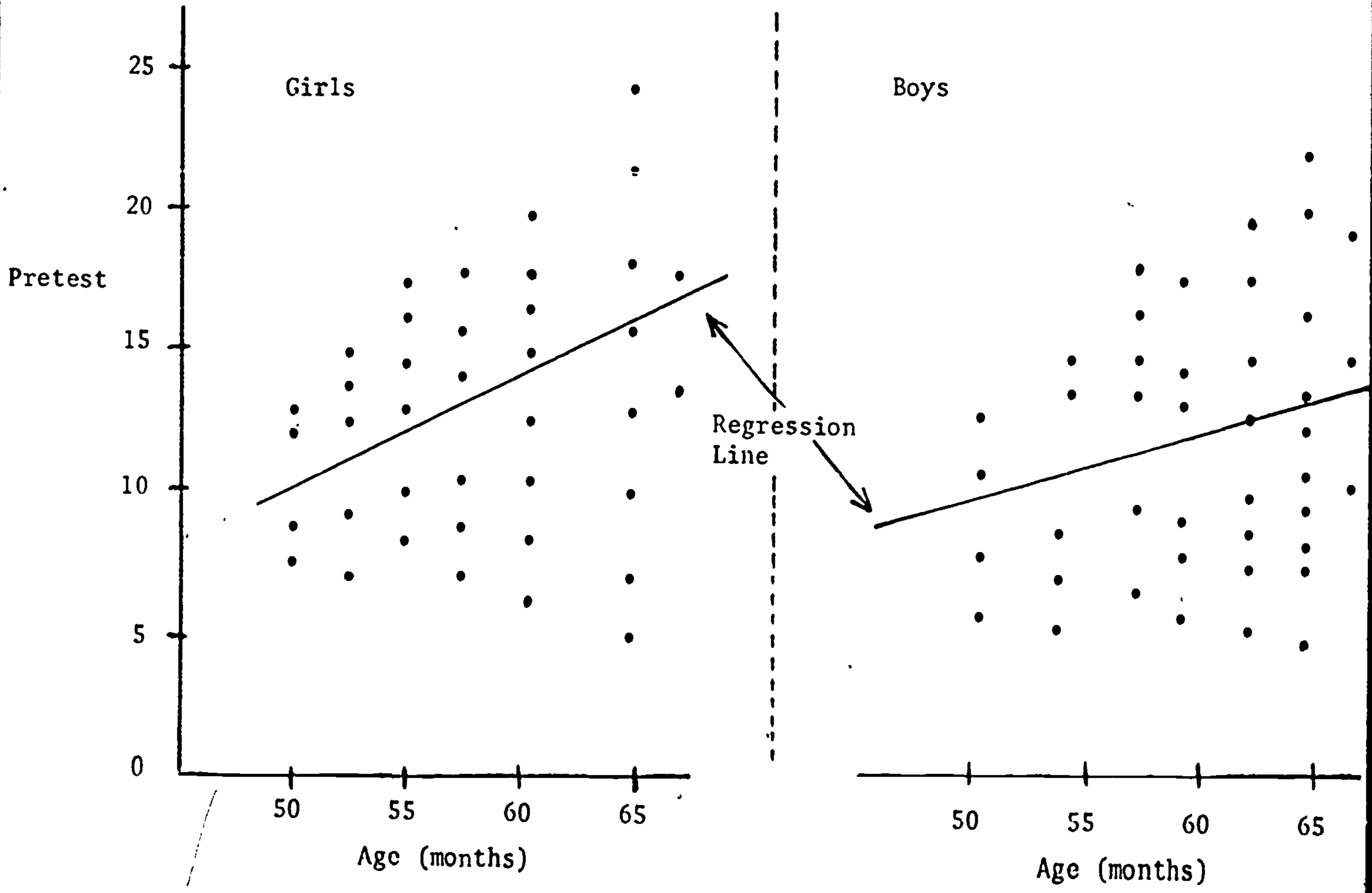

Figure 3. Scatterplots of Pretest by Age, Separately for Girls and Boys (Hypothetical Data).

estimates for boys and girls and an overall estimate. For girls,

expected gain = 1.1 points per month x 8 months (program duration) = 8.8 points,

and for boys,

expected gain = .9 points per month x 8 months = 7.2 points.

To estimate the program effect, we compare these expected gains with the actual observed gains. The mean test scores for boys and girls might be something like this:

	Girls	Boys
Posttest means	30.0	22.0
Pretest means	15.0	5.0
Observed gains	15.0	17.0
Expected gains	8.8	7.2
Value-added (observed gains minus expected gains)	6.2	9.8
Index of Educational Significance	70.5%	136.1%

Thus, the program seems to have had a somewhat larger effect for boys than for girls. In order to develop an overall estimate, we weight each of the separate estimates proportionally to that group's representation in the program. For example, if the program group consisted of 100 children--

30 girls and 70 boys--then the composite value-added estimate would be

$$\text{overall value-added estimate} = (30/100) \times 6.2 + (70/100) \times 9.8 = 8.7 \text{ points.}$$

Thus, the average effect of the program is 8.7 points.

While this example serves to illustrate the basic mechanism for incorporating background information into the value-added model, this approach becomes computationally unwieldly when we have multiple background variables, and when some of these are continuous--e.g., information on family income measured in dollars. While it is possible to take a continuous variable and create groups or categories (e.g. low, middle, and high income families), this is not a very efficient way to use the information. Rather, we can approach this as a multiple regression analysis problem. Appendix A presents a complete mathematical model, and a worked example of this approach. Here, we present only the simplest case using one background variable (again information on child's sex).

We can think of a growth rate for each child in the absence of an ECT-I program as consisting of a base, or average, growth rate plus adjustments to the base associated with particular background characteristics of the child. In the basic value-added model, the regression equation had only one independent variable: age in months. Now, we add to the equation other independent variables to represent the additional pieces of background information. In particular, we might want to know how sex group membership alters the relationship between pretest and age. We refer to this as the interaction of sex with age, and we represent it in the regression equation

as follows. We begin by designating some number to represent each sex group. The choice of number is totally arbitrary, but it is often convenient to use 1.0 for girls and -1.0 for boys. We can think of this as an indicator variable that identifies the sex of the child (values of 1 for girls, -1 for boys). Next, we multiply age by the sex indicator variable to create a sex-by-age interaction variable.* We then regress the pretest scores on age and on the age x sex interaction variable to determine a new prediction equation that contains more information on natural growth rate. Although the actual computational formulas are somewhat complicated, the results are fairly intuitive.

We illustrate the technique with data from a short-term impact evaluation of an ECT-I first-grade program in Rhode Island. Each child was pre- and posttested on the Peabody Picture Vocabulary test. Regressing the pretest scores on age and the age x sex interaction variable yielded:

predicted test score = 17.5 + .56 x age - .05 x age-by-sex interaction.

It can be shown easily that the regression slope for age (.56) represents the base growth rate, and that the regression slope for the interaction term (.05) represents the adjustment to the growth rate associated with sex group membership.

---

* This technique can be extended for background variables that have more than two categories. In this case, we would create a series of indicator variables to represent membership in the various possible groups. Alternatively, if the background variable is continuous, we do not create indicator variables, but directly multiply the background variable and age to create the interaction variable. More details on this technique can be found in a text on applied regression analysis such as J. Cohen and P. Cohen. Applied Multiple Regression/Correlational Analyses for the Behavioral Sciences. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.

For the Rhode Island data,

natural growth rate for girls = .56 - .05 (1.0) = .51 points per month

value indicating "girl"

natural growth rate for boys = .56 - .05 (-1.0) = .61 points per month.

value indicating "boy"*

The average program duration was 4.6 months for girls and 4.7 months for boys. Thus, for girls,

expected gain = .51 points per month x 4.6 months = 2.3 points,

and for boys,

expected gain = .61 points per month x 4.7 months = 2.9 points.

---

* While the choice of the values 1 and -1 is arbitrary, the size of the adjustment (i.e. the regression slope for the interaction variable) will depend on the particular values chosen. The separate growth rate estimates for girls and boys, however, will remain the same regardless of the choice of values for the indicator variable. For example, if we had chosen 2 for boys and -2 for girls the following prediction equation would have resulted:

predicted test score = 17.5 + .56 x age + .025 x age-by-sex interaction.

This equation, however, yields the same growth rates:

natural growth rate for girls = .56 + .025 (-2) = .51

value indicating "girl"

natural growth rate for boys = .56 + .025 (2) = .61.

value indicating "boy"

The observed test scores for the Rhode Island ECT-I kindergarten program were:

	Girls (n=21)	Boys (n=34)
	(total sample size = 55)	
Posttest means	53.8	59.4
Pretest means	50.5	57.5
Observed gains	3.3	1.9
Expected gains	2.3	2.8
Value added	1.0	-0.9
Index of Educational Significance	43%	-32%

Overall value added = (21/55) 1.0 + (34/55) (-.9) = -.17.

These results suggest a small positive program effect for girls, and a corresponding small negative effect for boys. On average, there appears to be no significant program effect.

IV. DISCUSSION OF MODEL ASSUMPTIONS

The idea underlying the value-added strategy is to use information about natural growth contained in the relationship between pretest and age (and background variables) to predict an expected gain in the absence of the ECT-I program. This expected gain represents the standard against which we compare the observed gain, and thereby assess the program's effectiveness. Thus, this approach requires neither a comparison group nor the use of a norm-referenced test in order to develop its standard of comparison. The validity of the approach, however, rests on an important assumption--the stable universe--which is key to our ability to estimate the expected growth in the absence of the program from the pretest and age relationship.

ASSUMPTION OF A STABLE UNIVERSE

The assumption of a stable universe is basic to any attempt to draw longitudinal inferences (e.g. expected growth for ECT-I children over a period of time) from cross-sectional data (e.g. test data such as pretest collected at one point in time). In particular, the value-added model assumes that the score at the pretest point represents the cumulative effect of natural growth up to that time point. While the influences of other factors (e.g. test-situational effects) are also represented in the observed test scores, the model requires that such influences be independent of the child's age.

Problems with this assumption can arise in several ways. First, there may be historical trends or accidents causing children born at different times to differ. For example, children born after a major outbreak of rubella might on average have somewhat lower growth rates. The model

assumes, however, that children born in say January of 1978 will grow at approximately the same average rate as children born in June 1978. For relatively homogeneous age groups (i.e., where the age range of children in the program group is less than 24 months), historical trends or accidents are unlikely to be a problem.

Second, schooling experiences prior to ECT-I programs can also be a source of concern. Since older children are more likely to have had such experiences, this pretest-age relationship may confound both natural growth influences and any effects of prior programs. Estimating expected gains on the basis of such data can be a serious misrepresentation, and as a result yield a very biased estimate of the program effect. Thus, application of the value-added model should generally be limited to the youngest groups (e.g., those in prekindergarten and kindergarten programs), who are most likely to have received little or no prior formal schooling. The more the ECT-I participants vary in previous school experiences, the more strain is placed on the stable universe assumption.

Third, the process of selecting children for the program may introduce a problem. For example, the oldest children in an ECT-I preschool might be delayed entrants into the school group, and the youngest somewhat more precocious than average. Such cases might show up as outliers in a pretest by age scatterplot (see Figure 4). If only a few cases like this occur in the program sample, they can be set aside and the value-added analysis applied to the remaining cases. For any case deleted as an outlier, there should be some corroborating evidence, such as teacher reports or parent interviews, to document that in fact this case is unusual. If the problem is more widespread, however, application of the value-added model is not appropriate.

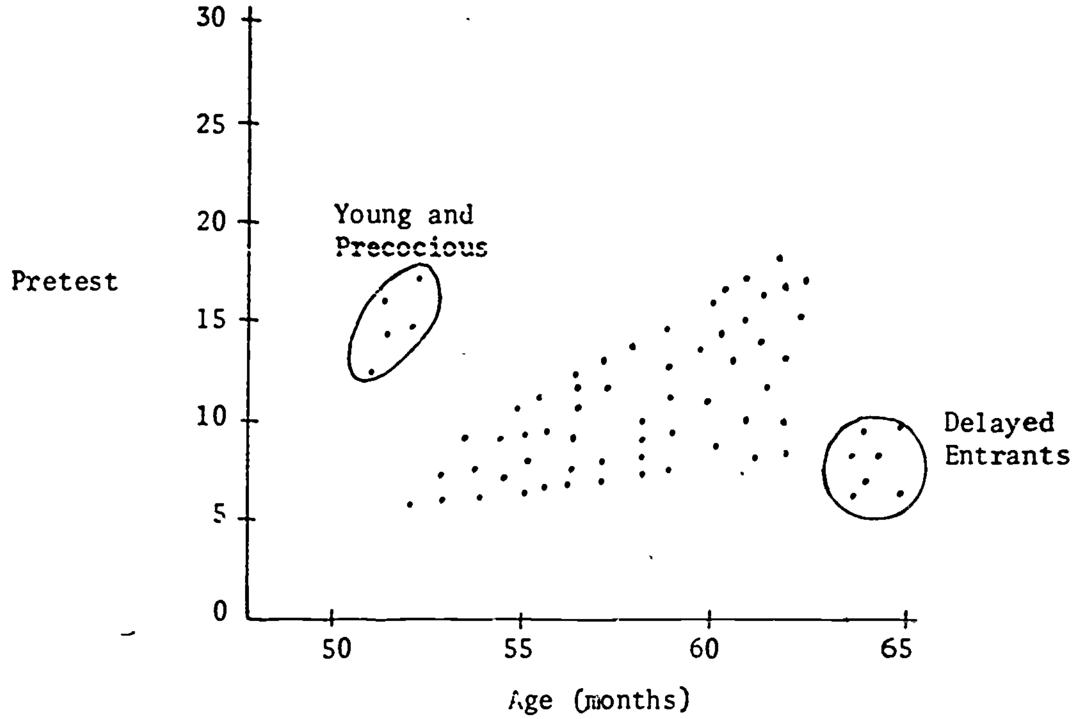

Figure 4. Differential Program Selection of Children by Age (Hypothetical Data).

LINEAR GROWTH ASSUMPTION

In using the simple regression of pretest on age to estimate the natural growth in the absence of an ECT-I program, we are assuming that the average natural growth trajectory is linear--that is, that the expected gain per month is independent of the age of the child (see Appendix A for more details). While this may be too simplistic a model to describe many educational processes in any detail, it should often be a reasonable approximation over a short period such as a six- to eight-month program. Over longer durations it is less likely to be adequate and use of the model should be approached with caution.

Now a violation of either the stable universe or the linear growth assumption will usually evidence itself in the pretest and age scatterplot. We have already seen in Figure 4 some data where differential selection by age had occurred. In general, if the linear growth and stable universe assumptions are valid, we should see pretest and age scatterplots similar to those in Figures 1 and 2. The line that we place through the scatter of points does a good job of summarizing the basic relationship between test score and age. Compare this with the situation represented in Figure 5. Here, the test score and age relationship is not well summarized by a line, but rather requires a curve to represent it. Thus, a partial test of the appropriateness of the value-added model is whether the pretest and age relationship appears linear. A further consideration is whether the pretest and age scatterplot appears as a tipped funnel (see Figures 1, 2, and 6). Under most situations involving linear individual growth, we expect test scores to be more variable for older children. This is reflected in the scatterplot by the widening of the funnel associated with older ages.

If a violation of either the linear growth or the stable universe assumption occurs, it will usually be reflected in the pretest and age scatterplots. Tipped funnels such as those found in Figures 1, 2, and 6 are very unlikely in such situations. While the existence of a tipped funnel scatterplot doesn't assure validity of the value-added model, it is a useful empirical tool for identifying possible problems. We have already indicated how differential selection effects may be identified and partially compensated for. More generally, if one suspects nonlinear individual growth and finds a nonlinear scatterplot, such as in Figure 5, it is often possible to transform

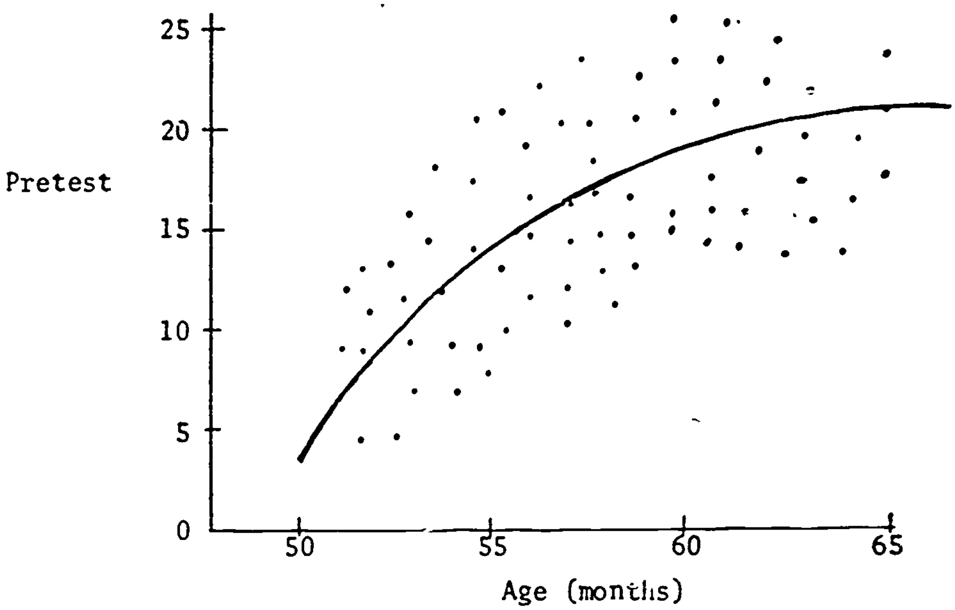

Figure 5. Nonlinear Relationship between Pretest Score and Age (Hypothetical Data).

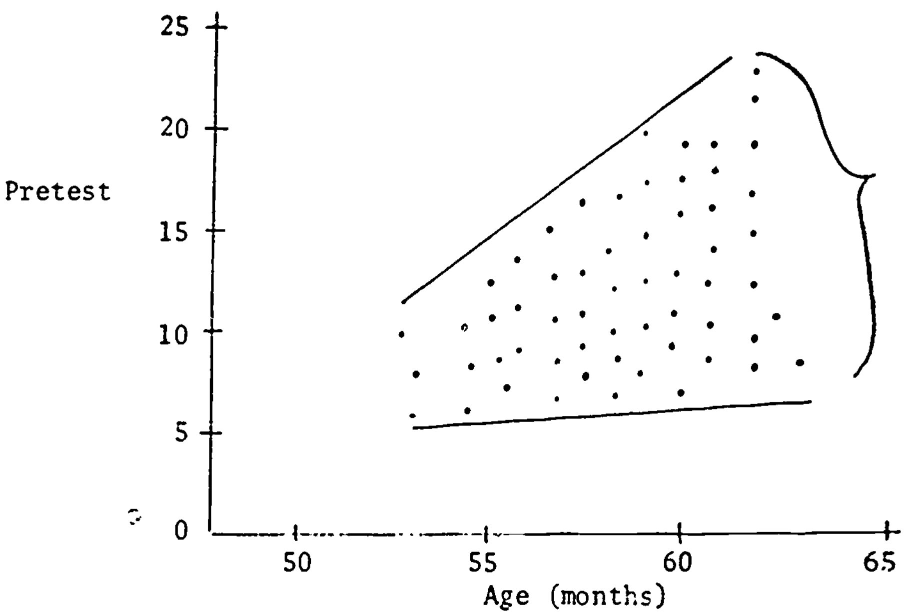

Figure 6. Pretest vs. Age: Tipped Funnel Scatterplot (Hypothetical Data).

the test score into an alternative metric (e.g. logarithmic or negative exponential transformation) in which the test-score/age relationship is again linear. The interested reader is referred to Bryk, Strenio, and Weisberg, 1980, for more details and an illustration.

One should, however, approach such transformation with caution. The existence of a nonlinear pretest and age scatterplot does not prove that individual growth is nonlinear. Violation of the stable universe assumption can also give rise to scatterplots which are nonlinear in appearance even though individual growth may be well represented by a straight line. Thus, if we observe a nonlinear scatterplot, we should first look carefully at the ECT-I evaluation design for possible violations of the stable universe assumption before going ahead with some nonlinear transformation.

ASSORTED OTHER ASSUMPTIONS

In varying degrees, each application of the value-added model requires some extrapolation beyond the available data. We can think of the value-added model as an approach to predicting an expected average test score for a group of children at some future time. The model implicitly assumes that the relation of age to test score apparent at pretest time will still be valid at posttest time when the children are significantly older. For example, if ECT-I participants are 45 to 57 months old at pretest time and the program lasts nine months, we must assume that the natural growth rates estimated at the pretest time point extrapolate into the age range of 54 to 64 months. Such extrapolation can sometimes be troublesome. Ideally, the age range within the sample should be considerably larger than the expected program duration. For example, if we were evaluating the effects of a six-month

preschool program, it would be desirable to have a spread of 12 to 18 months in children's ages. By contrast, in situations where the program duration exceeds the age range in the sample, the value-added approach is generally not recommended.

Finally, the model assumes that the outcomes to be evaluated show a natural increase with age across the duration of the program. For some variables this may not be the case. For example, an ECT-I program might use an ability test, e.g. the Slossen, as a program outcome. In the standard metric (mean equals 100, standard deviation of 15), such tests do not normally display any systematic relationship to age since the purpose of the standardization is to remove them. The value-added model can be applied in such cases by merely transforming the scores back into a mental age metric. In other cases, however, application of the value-added model may simply be inappropriate. By the first grade, for example, most children have developed the gross motor skills involved in skipping or running, so measures of gross motor skills would show little if any relationship with age in this range. Application of the value-added model in such a case might yield an expected gain of zero or even a negative amount. The validity of this as a standard for comparison vis-a-vis the observed gain is subject to considerable question.

V. PROBLEMS LIKELY TO BE ENCOUNTERED IN APPLICATION TO ECT-I PROGRAMS

In developing this resource book we had an opportunity to apply the value-added model to several existing ECT-I data sets. The examples in previous sections were drawn from these analyses. Since the value-added approach has not seen widespread use, these analyses were helpful in identifying new problems, particularly those that might commonly affect ECT-I applications.

First is the familiar problem of testing and instrumentation. Floor and ceiling effects on either pretest or posttest can cause special difficulties in the value-added approach, because it depends upon the assumed stable relationship between age and test scores for all children. A floor or ceiling effect on the pretest will obscure the natural growth relationship of test scores with age, and thus interfere with our ability to estimate expected gains in the absence of the program. Similarly, such effects on the posttest would make it difficult for the program to exhibit a positive effect no matter how worthwhile the effort. In short, the value-added model, like the other ECT-I short-term impact models, requires good instrumentation if the evaluation is to yield useful results.

Second, if selection for the ECT-I program is based on the pretest score or some other test score highly correlated with it, a routine application of the value-added model will not be valid. Figure 7 illustrates the scatterplot of pretest score and age for one of the ECT-I data sets that we analyzed. This LEA apparently used a cutoff score of 46 on the pretest in selecting children for the ECT-I kindergarten program. Figure 7 clearly shows how the cutoff rule eliminates any relationship that may normally

exist between age and pretest. As shown in Appendix B, the simple regression of pretest on age will underestimate the natural growth rate, and as a result overestimate the program effect, in such cases.

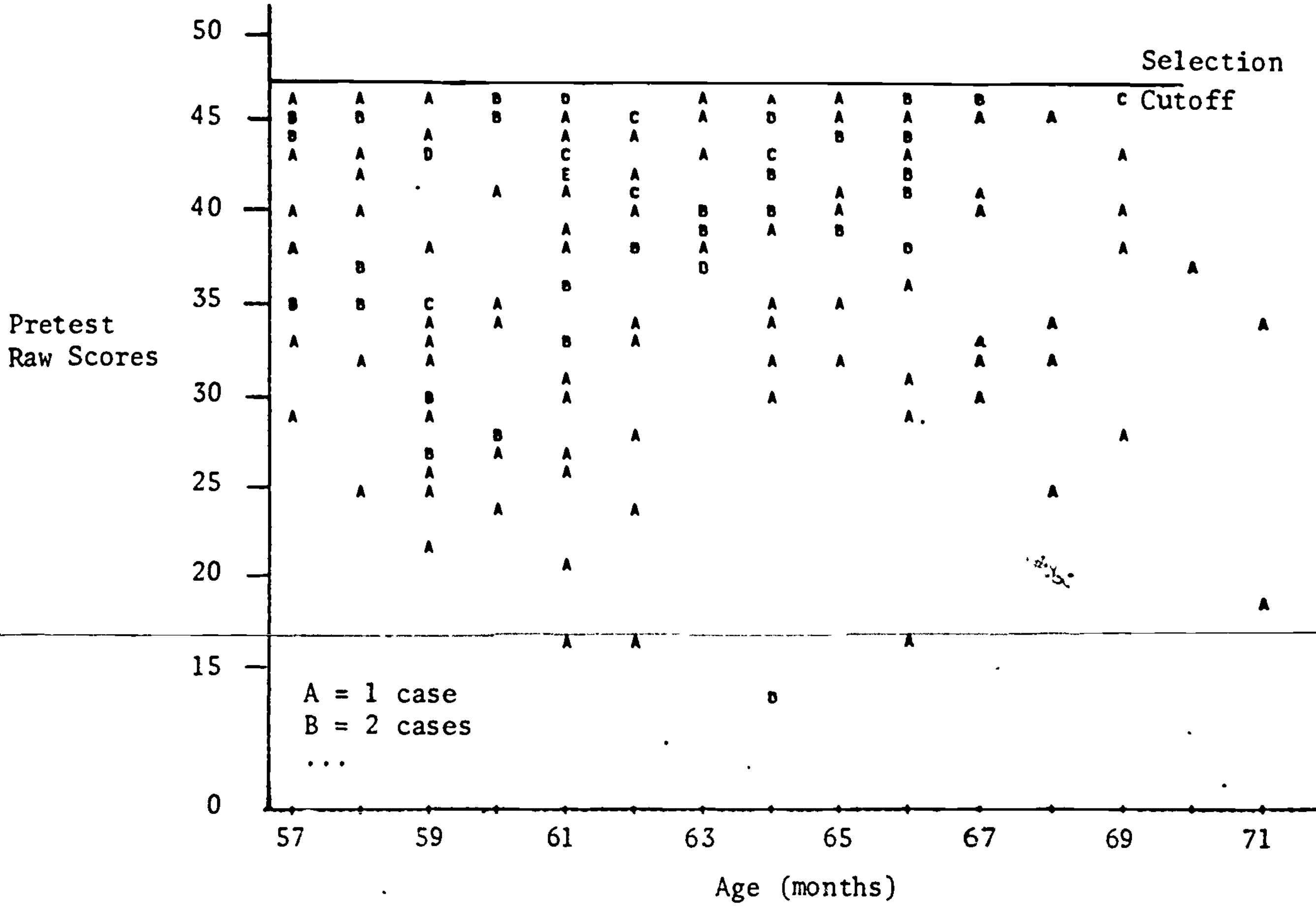

Figure 7. Scatterplot of Pretest Score and Age Under a Selection Cutoff Rule.

Now, if the program sample was drawn from a larger group through the use of a cutoff score rule, and if pretest and age data exist on this larger sample of children, then the basic natural growth can still be estimated by regressing the pretest scores on age for the entire group. The slope from

this regression should provide us with an unbiased estimate of the natural growth rate, which can be used to compute an expected gain and value added for the program sample. Alternatively, if data on the larger group do not exist, it is still possible to estimate the average natural growth rate, though the simple regression of pretest on age no longer suffices. Using more complex statistical methods, discussed in Appendix B, it is possible, although more difficult, to estimate the natural growth rate from the truncated data.

Third, since the value-added model hinges on this estimation of a slope or regression coefficient for the pretest and age relationship, its application with small sample sizes should be viewed with caution. With a small sample, the presence of just one or two outliers can significantly distort our estimate of the natural growth rate (i.e. the regression coefficient or slope). For sample sizes smaller than 30, even the basic value-added model (i.e. the simple regression of pretest on age) can be quite sensitive to sampling variations. As for application of the value-added model with background variables, this should be limited to fairly large sample sizes. If we wish to apply the model separately for distinct subgroups (e.g. boys and girls), then at least 30 subjects per group would be desirable. If we wish to apply the regression approach with several age-by-background interaction terms, then even larger sample sizes would be required.

Thus, in situations where we wish to assess the short-term impact of an ECT-I program and the available sample size is fairly small, the validity of results from the value-added model may be open to question. In these situations, pooling of data across multiple years of the program, or perhaps across several sites that are implementing similar activities, represents a possible

solution. An alternative formulation of the value-added model, employing an empirical Bayes approach to estimating the natural growth rate, appears particularly promising in dealing with data from multiple sites or years.*

* The algorithm for this method, however, is somewhat complicated, and requires some special computer programming. A discussion of the technique and illustrations can be found in J. Strenio, Empirical Bayes Estimation for a Hierarchical Linear Model (doctoral dissertation, Harvard University, 1981).

VI. CONCLUDING REMARKS

The value-added approach to measuring short-term impact of ECT-I programs has both strengths and weaknesses. Its major strengths are that it does not require a comparison group or the use of a norm-referenced test. Its major weaknesses are that (1) it is appropriate only for the assessment of skills or attributes which show a natural development with age over the duration of the program; (2) selection procedures may disguise the relationship between age and skill development among a particular group of program participants (thus either precluding application of the value-added approach or necessitating reliance on some additional data as a source of deriving appropriate age/skill-development projections); and (3) it can require some quite complex statistical calculations.

We should view each application of the value-added model with some reservations. We are attempting to derive a great deal of information (i.e. an estimate of the program effect) from a situation with little data (i.e. no comparison group or valid norm) and little external control (i.e. no random assignment). While it is possible to develop an estimate of the program effect, its validity should be carefully examined along lines suggested in the last two sections.

More generally, as the very first test, we should always ask the question "Does it make sense?" For a variety of reasons, mentioned above, it is quite possible that in an individual application the pretest and age relationship might appear negative, implying an average loss in test score points in the absence of a program. In most situations, expectations of negative gains due to natural maturation would be nonsensical, and the evaluator

should discard the analysis as clearly incorrect. Similarly, it is quite possible in an individual application that the expected gain under the value-added model, when added to the average pretest score, exceeds the maximum score for this test or instrument. If such a ceiling effect appears, application of the value-added model is again inappropriate.

Finally, only so much data analytic advice can be packaged in a fairly short and nontechnical resource booklet. It cannot substitute for technical expertise fully grounded in an understanding of the statistical model and its estimation procedures. Whenever possible, such professional assistance should be sought to help in the application of this approach.

APPENDIX A

TECHNICAL DISCUSSION:

(1) BASIC VALUE-ADDED MODEL
(2) EXTENDING THE MODEL TO INCORPORATE BACKGROUND VARIABLES

This appendix is a more technical introduction to the value-added approach. It summarizes Bryk, Strenio, and Weisberg (1980), and the interested reader is referred there for more details.

BASIC VALUE-ADDED MODEL

The value-added model focuses on the natural growth of subjects prior to an ECT I program, attempting to project explicitly a posttest status for the program group as if they had been subject to the control condition. The actual growth is then compared with projected growth, the difference representing the effects of the program.

The model assumes that over the duration between pre and posttest each individual's growth consists of two components: (1) systematic growth, which can be characteristized by a growth rate and a corresponding growth curve; and (2) an individual noise or random component, which is specific to a particular subject at a certain point in time. Thus we can represent the observed score for individual i at any time t as

$$Y_i(t) = G_i(t) + R_i(t), \tag{1}$$

where $G_i(t)$ represents systematic growth and $R_i(t)$ represents the random component.

The individual's systematic growth, $G_i(t)$, is represented as a function of age (or some other time metric). While in principle this function may take any form, it may often be adequate to assume that it is linear:

$$G_i(t) = \pi_i a_i(t) + \delta_i, \tag{2}$$

where π_i represents the slope, δ_i represents the Y intercept, and $a_i(t)$ is the age for subject i at time t. Individuals may vary in terms of a growth rate, π, and an intercept parameter, δ. The model assumes that π and δ are

distributed with means μ_π and μ_δ, variances σ_π^2 and σ_δ^2, and covariance $\sigma_{\pi\delta}$.

Note that this represents the simplest model for G(t), which incorporates varying individual growth. While too simple to fully describe many growth processes, linear individual growth may be a reasonable analytic approximation over a short term even if long-term growth has a more complex form.

As for the random component, the model assumes that

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

$$\mathrm{Var}[R_i(t)] = \sigma_R^2 \tag{3}$$

(i.e., fixed over subjects and time) and

$$\mathrm{Cov}[R_i(t), \pi] = \mathrm{Cov}[R_i(t), \delta] = 0.$$

Thus we represent the observed pretest ($t = t_1$) as

$$Y_i(t_1) = \pi_i a_i(t_1) + \delta_i + R_i(t_1). \tag{4}$$

For convenience, let us define

$$\Delta_i = a_i(t_2) - a_i(t_1), \tag{5}$$

where Δ_i represents the time duration between pre- and posttest for subject i. Note that we are assuming that t_1 and t_2 may differ across subjects, but are dropping the subscript i for notational convenience.

At the posttest ($t = t_2$), in the absence of a treatment, we would have

$$Y_i(t_2) = \pi_i a_i(t_2) + \delta_i + R_i(t_2)$$

$$= G_i(t_1) + \pi_i \Delta_i + R_i(t_2), \qquad (6)$$

where $\pi_i \Delta_i$ represents the expected growth between pre- and postmeasure due solely to natural maturation.

In the presence of a program, we assume that over the time interval t_1 to t_2 the treatment increases each subject's growth by an amount ν_i, called the value added by the program. Thus we can represent the measured growth subject i achieves by time t_2 under an intervention as

$$Y_i(t_2) = G_i(t_1) + \pi_i \Delta_i + \nu_i + R_i(t_2). \qquad (7)$$

Under this model, the treatment effect is fully described by the distribution of the ν_i. We assume that ν is a random variable with mean μ_ν, and variance σ_ν^2. Normally, we are interested in a summary measure of the treatment effect. This suggests that we estimate μ_ν, the average of the individual treatment effects.

During the period between the pre- and post-measure, the observed change in the treatment group is $\overline{Y}(t_2) - \overline{Y}(t_1)$. The expected growth under the model is $\mu_\pi \overline{\Delta}$. If we knew the value of μ_π, a natural estimator of μ_ν would be

$$\hat{\mu}_\nu = \overline{Y}(t_2) - \overline{Y}(t_1) - \mu_\pi \overline{\Delta}. \qquad (8)$$

Bryk et al. (1980) have shown, under the assumption that π and δ are independent of $a(t_1)$ and that π and Δ are independent, that the ordinary least squares regression of $Y(t_1)$ on $a(t_1)$ yields an unbiased estimate of μ_π, and as a result,

$$V = \bar{Y}(t_2) - \bar{Y}(t_1) - \hat{\beta}_\pi \bar{\Delta} \tag{9}$$

represents an unbiased estimate of μ_ν.

While Bryk et al. (1980) do not derive an estimate of the standard error of V, they suggest the use of the jackknife technique (described in Chapter 8 of Mosteller and Tukey, 1977) to provide both a test statistic and standard error of V.

EXTENDING THE BASIC VALUE-ADDED MODEL: INCORPORATING BACKGROUND VARIABLES

In the previous section, we assumed that the growth rate parameter, π_i, was a random variable that characterized each individual's determinants of growth on some outcome dimension of interest. We have implicitly assumed that π_i is unmeasurable. One obvious alternative is to consider models that incorporate additional background information besides age. We introduce in this section a model in which the individual growth rate π_i is represented as a function of measurable variables that stand in proxy for environmental and constitutional factors determining the individual's growth rate.

Model Specification

We assume that each individual's growth rate can be represented as a linear function of measurable background variables:

$$\pi_i = \theta_0 + \sum_{j=1}^{J} \theta_j X_{ij} + \varepsilon_i, \tag{10}$$

where

X_{ij} is the value on the jth background variable for subject i,

$\underset{\sim}{\theta}$ represents a vector of coefficients,

ε_i represents unmeasured determinants of individual growth rate, and

j=1 ... J variables; i=1 ... n subjects.

We assume further that

$$E(\varepsilon_i | \underset{\sim}{X}_i) = 0,$$

$$Var(\varepsilon_i | \underset{\sim}{X}_i) = \sigma^2_\varepsilon,$$

and $$Cov(\varepsilon_i, \underset{\sim}{X}_i) = \underset{\sim}{0}. \tag{11}$$

Equations 10 and 11 constitute our model for the individual growth rate parameter. We are assuming here that the differences in π_i among individuals can, at least partially, be expressed as a function of measurable variables. In particular, since participants will have different values for the vector of background variables, $\underset{\sim}{X}$, different individual values of π_i are predicted.

Substituting Equation 10 into our basic model, specified by Equation 4, we obtain

$$Y_i(t_1) = \mu_\delta + \theta_0 a_i(t_1) + \sum_{j=1}^{J} \theta_j X_{ij} a_i(t_1) + \varepsilon_i^*, \tag{12}$$

where

$$\varepsilon_i^* = R_i(t_1) + \varepsilon_i a_i(t_1) + (\delta_i - \mu_\delta). \tag{13}$$

Since this is a linear function of the parameters $\underset{\sim}{\theta}$, the simplest approach is to apply ordinary least squares (OLS) to the model in Equation 12. However, OLS will yield unbiased estimators only if

$$E(\varepsilon^* | a(t_1), \underset{\sim}{X}) = 0 \tag{14}$$

Bryk et al. (1980) have shown that if we assume for any given value of $\underset{\sim}{X}$ that $a(t_1)$ is uncorrelated with both π and δ, then

$$E(\varepsilon^* | a(t_1), \underset{\sim}{X}) = 0, \tag{15}$$

and OLS will therefore produce an unbiased estimate of $\underset{\sim}{\theta}$.

To understand this condition intuitively, let us consider a simpler (and less efficient) way that background information may be employed. If we simply stratified the data to create groups homogeneous in terms of their $\underset{\sim}{X}$ values, we could then carry out separate value-added analyses using the basic procedure derived earlier for each group and combine the results by averaging across groups. (A simple version of this analysis was illustrated earlier in the main body of the resource book.) However, for the analysis

on any group to produce an unbiased estimate of μ_V, the basic model assumption (i.e. that π and δ are independent of $a(t_1)$ and that π and δ are independent) must hold for that group. That is, it must hold conditionally on X.

To apply the least squares estimation procedure (following Equation 12) we regress the pretest, $Y(t_1)$ on age and the first order interactions of age and the background variables. From the $\hat{\underset{\sim}{\theta}}$ coefficients we can now estimate an individual growth rate, that is,

$$\hat{\pi}_i = \hat{\theta}_0 + \sum_{j=1}^{J} \hat{\theta}_j X_{ij} \tag{16}$$

and with these we can now calculate an average value-added, $V_{\hat{\theta}}$, where

$$V_{\hat{\theta}} = \overline{Y}_2 - \overline{Y}_1 - \frac{1}{n} \Sigma \hat{\pi}_i \Delta_i . \tag{17}$$

Under our assumptions we can easily demonstrate that

$$E(V_{\hat{\theta}}) = \mu_V . \tag{18}$$

So, we obtain an unbiased estimator of μ_V.

If the multiple correlation between π and $\underset{\sim}{X}$ is substantial, this estimator $V_{\hat{\theta}}$ should be more precise than the basic V. An exact expression for the variance, however, is rather complex, since it involves the covariances among the estimated θ_j's, which depend on the values of both $a(t_1)$ and $\underset{\sim}{X}$. Similarly, an expression for $Var(V_{\hat{\theta}})$ would depend on the unknown distributions for δ and $\underset{\sim}{\theta}$.

. A Day Care Study Application

As an illustration of how $V_{\hat{\theta}}$ can be applied, let us consider a small subset of the analyses performed as part of the National Day Care Study (NDCS). The Preschool Inventory (PSI) was administered to a large sample of day care participants. A score for each child was recorded in the early fall (pretest) and again in mid-spring (posttest). In addition, a set of background information on the child and family was collected.

One question of interest in the NDCS concerned the effects of different day care centers on the cognitive development (as reflected on the PSI) of children in these centers. The researchers wished to determine the average increment to a child's PSI score above that resulting from natural maturation. For technical reasons, irrelevant to this example, a transformation of the PSI score was actually used as the outcome measure.

Using the approach introduced above, we must first estimate θ coefficients for the individual growth rate model as specified by Equation 12. In the NDCS, four variables seemed most important on the basis of theoretical knowledge and preliminary analyses. These were:

X_1 = child's sex (1 = male; -1 = female),

X_2 = child's race (1 = black; -1 = white),

X_3 = mother's education (1 = more than 12 years; -1 = 12 years or less),

X_4 = public assistance (1 = receives; -1 = does not receive).

The regression of PSI pretest scores against age and the first-order interactions between age and these variables produced

$$\hat{Y}(t_1) = -.4086 + .2942a_1(t_1) - .0065a_1(t)X_1 - .0345a_1(t)X_2$$

$$+ .0123a_1(t)X_3 - .0167a_1(t)X_4.$$

So the estimated value π for any individual was given by

$$\hat{\pi} = .2942 - .0065X_1 - .0345X_2 + .0123X_3 - .0167X_4.$$

To interpret this equation, consider two illustrative individuals, one a "disadvantaged" child and the other a "middle-class" child. The individual participant data and resulting growth rate estimates are shown in Table A.1.

For Case I (disadvantaged child):

$$\hat{\pi}_I = .2942 - .0065(1) - .0345(1) + .0123(-1) - .0167(1)$$

$$\hat{\pi}_I = 0.2242$$

For Case II (middle class child):

$$\hat{\pi}_{II} = .2942 - .0065(1) - .0345(-1) + .0123(1) - .0167(-1)$$

$$\hat{\pi}_{II} = 0.3642$$

In general, these results are consistent with what we might expect. The estimated growth rate for the "disadvantaged child," $\hat{\pi}_I = .22$, is considerably lower than that for the more "advantaged child," $\hat{\pi}_{II} = .36$.

Moreover, because the background variables reflect group membership in this example, the $\hat{\theta}_j$ coefficients have a very simple interpretation.

Table A.1. Illustrative Data for Two Children in the National Day Care Study

Variables	Disadvantaged Child	Middle Class Child
PSI posttest (transformed metric)	.4000	.6000
PSI pretest (transformed metric)	.2000	.4000
Sex, X_1	1 (Male)	-1 (Female)
Race, X_2	1 (Black)	-1 (White
Mother's Education, X_3	-1 ($\leq$ 12 years)	1 (> 12 years)
Public Assistance, X_4	1 (Yes)	-1 (No)
Δ	.36 (Years)	.43 (Years)

If we think of $\hat{\theta}_0$ as a general growth rate for the total sample, then each $\hat{\theta}_j$ reflects an adjustment (either up or down) on the base rate associated with membership in some subgroup of the total sample. For example, for a male child $\hat{\pi}_i$ is reduced by .0065, while for a female child it is raised by the same amount. Being female, then, corresponds to an increase of .0065 over the average. We should note, however, that this straightforward interpretation represents only a description of the analytic model. No causal inferences are intended.

We are now in a position to calculate $V_{\hat{\theta}}$ for an individual day care center. We compute values of $\hat{\pi}_i$ and Δ_i for each participant. From Equation 17 an overall estimate of μ_V is obtained. In this example $V_{\hat{\theta}} = .172$.

Note that since the background variables are discrete, we could have estimated a separate μ_π for each subgroup of cases within the sample. In particular, we could have stratified the sample using the four independent variables, creating a 2^4 design, and then applied the basic value-added estimator, V, within each of the cells in the design. To be effective, this approach requires that the sample size within each cell be sufficiently large to generate a stable value for μ_π. The linear model approach illustrated above exploits the data in a more efficient way.

REFERENCES

Bryk, A.S., Strenio, J.F., & Weisberg, H.I. A method for estimating tests and effects when individuals are growing. Journal of Educational Statistics, Vol. 5, No. 1, 1980, 5-34.

Mosteller, F., & Tukey, J.W. Data analysis and regression. Reading, MA: Addison-Wesley, 1977.

APPENDIX B

EFFECT OF SELECTION ON THE PRETEST ON ESTIMATING A
NATURAL GROWTH RATE FROM THE
SIMPLE REGRESSION OF PRETEST ON AGE

This appendix discusses the problems involved in estimating a regression coefficient based on a data set in which the observations form a non-random, truncated sample. This problem arises in applications of the value-added method to situations where the pretest score was used as a selection criteria. Thus children who did well on the pretest did not enter the program and are not part of the data set. Since children usually do better with age, this selection process will mean that fewer older children and more younger children will enter the program. This differential selection procedure with regard to age creates a situation in which the least-squares regression of pretest on age will result in a biased estimate of the natural growth rate occurring in the absence of intervention.

Fortunately, the truncated dependent variable problem also occurs in econometrics. This appendix summarizes some of the findings of Hausman and Wise (1977), and applies them to the value-added situation.

THE BASIC PROBLEM

We begin by defining some variables and stating the value-added model:

$Y_i(t_1)$	= subject i's pretest score	$i=1,\ldots,n$
$a_i(t_1)$	= subject i's age at pretest	$i=1,\ldots,n$
β	= overall growth rate	
$R_i(t_1)$	= subject i's random component of growth	$i=1,\ldots,n$

Then the model is:

$$Y_i(t_1) = \beta a_i(t_1) + R_i(t_1)$$

where

$R_i(t_1) \sim n(0, \sigma_R^2)$ and is independent over subjects i, and independent from $a_i(t_1)$

However, since any child scoring better than some cut-off, L (or, more generally, L_i, where L_i might be related to age), is not included, we can also say:

if $Y_i \leq L_i$ we include individual i, or

if $Y_i > L_i$ we exclude individual i

For the most common case, where all L_i are equal, Figure B.1 illustrates the truncation problem. Here the solid line represents the underlying true relationship between age and test score in the population. If we screen children of say ages a_1, a_2, and a_3, we see that although the population, represented by dots, varies about this true line, we will accept into the program only those individuals with scores below the cut-off, L, whose dots are circled. Since regressing test score on age for just the program sample involves only the circled dots, the broken line represents the resulting estimate of the growth trajectory. Thus, we underestimate the natural growth rate. If we underestimate the natural growth rate, however, we will under-

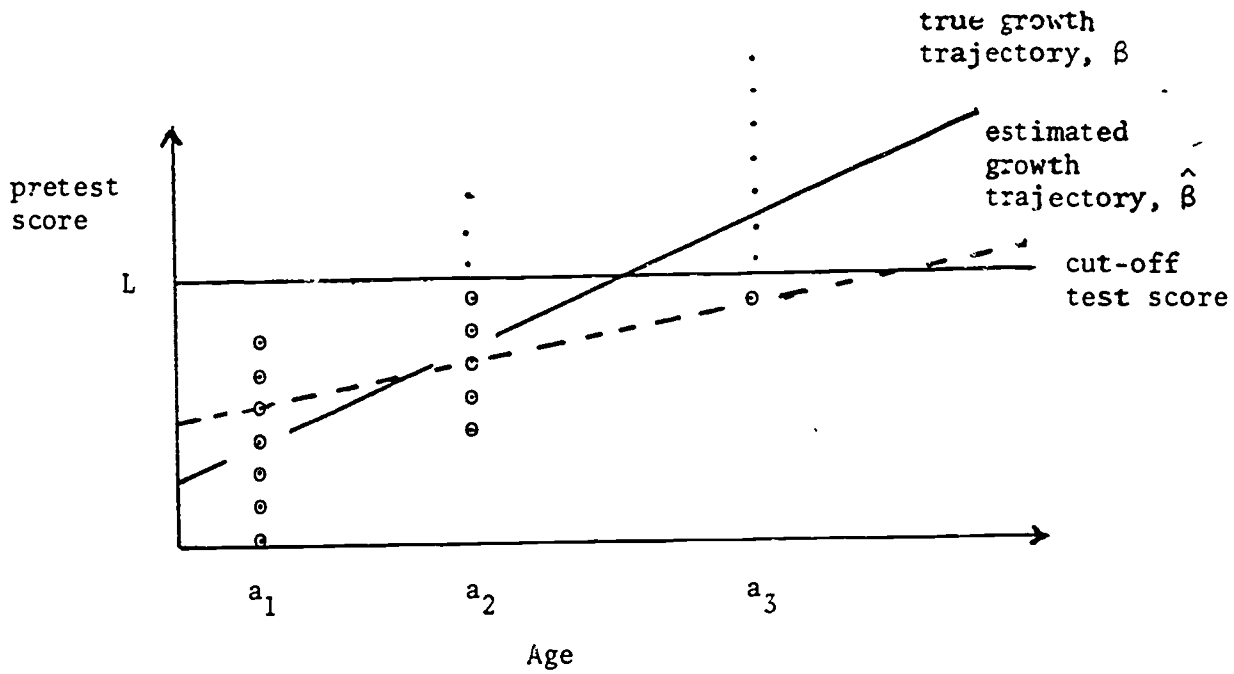

Figure B.1. Effects of Sample Truncation on Estimated Regression Line.

estimate the expected gain over the program period, and thus overestimate the value-added by the program.

From Figure B.1 we can see that the least-squares slope, $\hat{\beta}$, will always be flatter than the true β (for positive β, $\hat{\beta}$ will be smaller than β). For a given set of ages, the amount of the bias depends on three things: the value of L, the true growth rate β, and the variability of test scores around the growth curve, σ^2. As L increases, the bias decreases, as β decreases, the bias decreases; as σ^2 decreases, the bias decreases.

Hausman and Wise derive an exact formula for the bias. This formula depends on the assumption of normality of errors, R_i. First, they define

$$d_i = \frac{L_i - \beta a_i(t_1)}{\sigma}$$

This is the standardized distance from $E(Y_i | a_i)$ to the cut-off for age a_i. Now, for large samples, the bias of $\hat{\beta}$ is given in the following formula

$$\hat{\beta} = \text{true } \beta - \frac{\sigma}{\sum_{i=1}^{n} a_i^2(t_1)} \cdot \sum_{i=1}^{n} \frac{a_i(t_1)\phi(d_i)}{\Phi(d_i)}$$

where ϕ is the standard normal probability density function, and Φ is the standard normal distribution function.

From Figure B.1 we can also see that a lower cut-off would also bias the estimated slope downward, as would the operation of both upper and lower cut-offs together. Hausman and Wise do not consider this problem, but the bias equation would presumably be as above, with an additional third term adding in the probability lost below the lower cut-off as well.

As for the multivariate case, where more than one β are being estimated, the bias of individual parameter estimates cannot be determined a priori; but in general one should expect least squares estimates to be biased towards zero.

SOME ALTERNATIVE SOLUTIONS

In response to this problem, Hausman and Wise developed a maximum likelihood solution for estimating β from truncated Y data, and present an illustration of its application. The key assumption in the maximum likelihood estimation procedure is that the pretest score $Y_i(t_1)$ follow a conditional normal distribution with mean $\beta a_i(t_1)$. Glenday (1978) documents a general computer program which can be used for this situation.*

In terms of less technical (and also less efficient) ways to deal with the bias, there are several possibilities. If pretest and age data were collected on a larger population from which the program sample was eventually drawn, the simple regression of pretest on age for the larger population should yield an unbiased estimate of the natural growth rate to be used as the basis for the value added estimate. Another possibility is to use least squares, but to limit consideration to observations that fall within some restricted range. For example, use only values of age for which the expected value of Y given age is well below the truncation point. Still another possibility is to consider only observations having values of the dependent variable well below the truncation level. Again, this may lead to observations that are restricted to lie within a limited range.

* A copy of this demonstration is available at your regional technical assistance center, as well as information on how to procure a copy of the computer program.

REFERENCES

Anderson, S., et al. Statistical methods for comparative studies. New York: Wiley, 1980.

Glenday, G. The truncated data set program. Department of Industry, Trade and Commerce, November 1978.

Hausman, J.A., & Wise, D.A. Social experimentation, truncated distribution, and efficient estimation. Econometrics, Vol. 45, No. 4, 919-938.