2

A Short History of Regression-Discontinuity

Why should we look at the history of a research design? Outside of several obvious considerations—the value of documentation, the need for a literature tradition—an historical perspective serves to underscore the fact that research designs are dynamic and evolving entities, not static mechanistic ones. As subsequent chapters will demonstrate, many of the issues in the regression-discontinuity framework remain unsolved or only partially concluded. Good understanding of current issues requires some sense of the contexts that generated them. The story of the regression-discontinuity design encompasses several interesting dramas including the technical disputes of methodologists and statisticians, the political and financial arguments of the evaluation research industry, and the social debate of Congress and the nation regarding the allocation and evaluation of social programs, especially in the educational arena.

A second reason for taking an historical view is that it enables us to learn from previous experience which practices and procedures seem to work in the field and which do not. As will be shown in later chapters, the regression-discontinuity design as presented by theoreticians and methodologists was often difficult to implement, at least within educational settings. An understanding of why this is so will provide the applied researcher who is interested in using the design with some prior warning about difficulties that might arise.

This chapter merely outlines the major historical events, indicates the relevant literature, and discusses some of the factors that influenced the selection or rejection of the regression-discontinuity design for the evaluation of compensatory education programs. Subsequent chapters will provide a more detailed sense of the arguments and controversies involved.

The history of the regression-discontinuity design can be categorized neatly into two major traditions. The first is termed here the "academic" tradition and describes the ongoing technical development of the design. The second is named the "compensatory education" tradition and refers to the use of the design to evaluate programs to that type. The design has almost never been used outside of compensatory education evaluation with the exception of some illustrative examples in the technical literature and a notable evaluation in criminal justice (Berk & Rauma, 1983). This two-fold classification is not meant to imply that the two traditions did not overlap or interact—rather, it is used because it enables us to make better sense of key issues, while at the same time it preserves the integrity of the historical story line.

Because of the important historical role of compensatory education evaluation for the regression-discontinuity design, many of the examples in this volume will be set in the context of compensatory education. This should not imply that the design is useful only in this context for, as will be seen, it has far more

general applicability. Rather, the discussion emphasizes the educational tradition because the design can only be understood well within the context in which it is applied and it has almost exclusively been applied to date in compensatory education settings.

THE ACADEMIC TRADITION

The regression-discontinuity design was first suggested in its present form in a paper by Thistlethwaite and Campbell (1960) entitled, "Regression-discontinuity analysis: An alternative to the ex post factor experiment." Initially, regression-discontinuity was not accorded the status of a research "design" but was rather viewed as an analysis. The example that they considered was a study of the effects of winning a scholarship on career plans. Students scoring above a given aptitude level were awarded a scholarship, while those below that score were not.[1] The central post measures reflected the student's desire to undertake graduate study or be a college teacher or researcher. The regression-discontinuity analysis was compared to a randomized experimental procedure. The logic behind this comparison is shown in Figure 2.1. The top graph, Figure 2.1a, shows a hypothetical randomized experimental situation. Students in the study are selected from a relatively small range of the pretest (indicated by two vertical lines). Within that range, students are randomly assigned to either receive a scholarship or not. The "X" in the figure represents the post mean for the no scholarship group. Because the two means differ one might conclude that the scholarship had an effect. Figure 2.1b shows how this logic might be extended. Here, information is also available for students with pretest scores above and below the interval of the random assignment. The high scoring students all receive a scholarship, the low scorers do not. In this case, one might conduct two analyses—one for the cases that were randomly assigned and one for the cases outside that interval. For the latter, regression lines could be fit to each group and projected into the interval. Figure 2.1b shows that the lines would be projected to their respective randomly assigned group means. The effect of the program would be the same for both analyses and is indicated by the vertical distance between the two posttest means. The traditional regression-discontinuity analysis is shown in Figure 2.1c. Here, random assignment is not used at all. The randomization interval is replaced by a single cutoff score. The program is the same as above but is reflected in the vertical distance (jump or discontinuity) in the regression lines at the cutoff. Thus, Thistlethwaite and Campbell (1960) saw regression-discontinuity analysis as a direct extension of or alternative to a randomized experiment. The program effect is estimated at the cutoff because that is where the two groups are most similar or comparable in pretest ability.

The regression-discontinuity design was a frequent topic in the writings of Donald T. Campbell. In 1963, it was included in a monograph on research design by Campbell and Stanley (1963) where it was still cast in the category of an analysis rather than a design. By 1969, however (Campbell, 1969), regression-discontinuity had become a "design":

> But if randomization is not politically feasible or morally justifiable in a given setting, there is a powerful quasi-experimental design available that allows the scarce good to be given to the most needy or the most deserving. This is the regression-discontinuity design. (p. 248)

This paper recognized a number of central issues for the design: the need for adherence to the cutoff criterion; the distinction between "sharp" and "fuzzy" regression-discontinuity; the use of uncorrelated pretest and posttest; the choice of where to estimate the program effect; and the possibility of using a cutoff on a composite measure. The design was recommended for such diverse contexts as the investigation of Job Corps Training or the study of the effects of military conscription.

In the early 1970s, Robert F. Boruch authored and coauthored a number of papers dealing with regression-discontinuity. In "Regression-Discontinuity Revisited," (Boruch, 1973) an un-

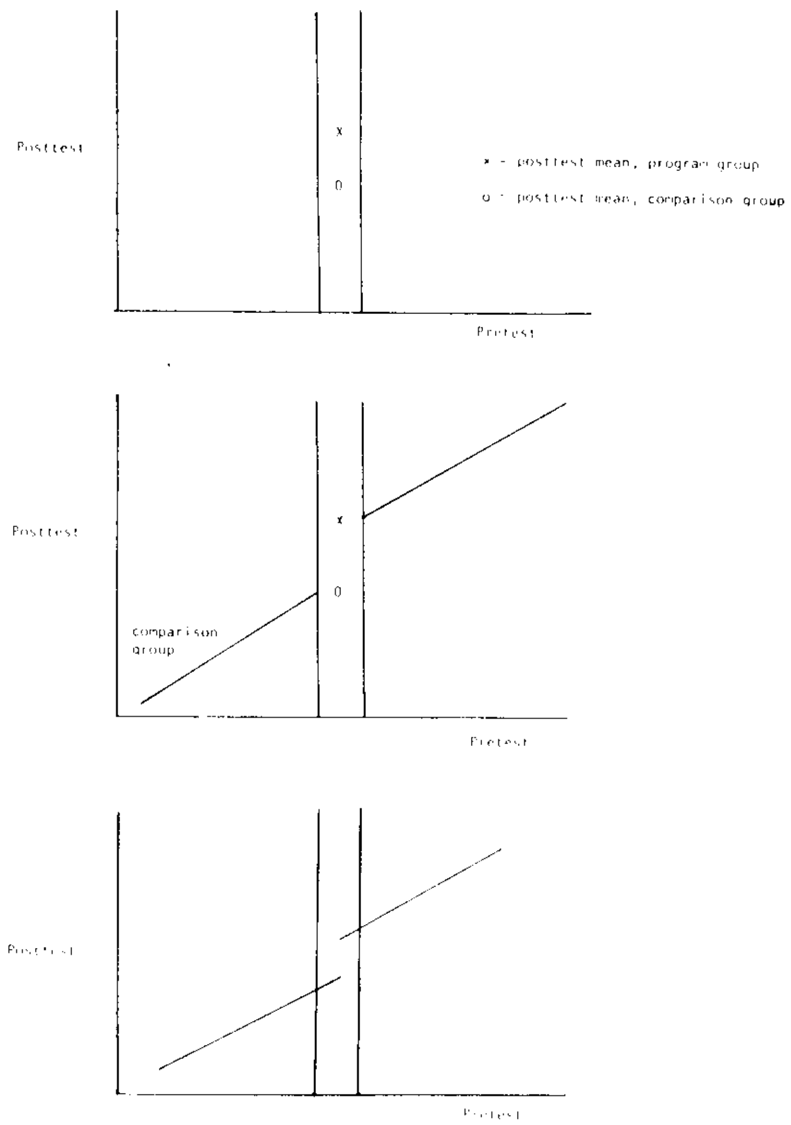

Figure 2.1 The Regression-Discontinuity Design as an Extension of a True Experiment

published paper, he discusses the statistical analysis of the design in some detail and distinguishes between cases where the preprogram measure can be considered fixed or random, and where the groups come from the same population or represent separate intact groups. In this paper, Boruch also points to the present recommended statistical analysis—the use of a single regression function with the group assignment included as a dummy-coded variable. Prior to this, the accepted analysis involved the calculation of separate group regression lines and construction of a t-test of differences between the intercepts at the cutoff point as discussed by Campbell. A number of applications are suggested for the design in the fields of education, health, mental health, criminal justice, and social welfare. In 1975, Boruch and DeGracie presented an evaluation of a compensatory education program using the regression-discontinuity design. The paper is particularly important for its recognition of the problem of model specification and the effects of curvilinearity on the statistical analysis.

By 1975, the regression-discontinuity design had still received almost no use outside of the illustrative analyses in the papers mentioned above. It had always seemed especially appropriate for evaluation in education where students are often assigned to program on the basis of need as measured on an achievement test. This fact and a desire on the part of Congress to see better evaluations of compensatory education provided the right mixture of ingredients for introduction of the design in that context. Because of its importance, the history of regression-discontinuity in compensatory education is discussed separately below.

More general discussions continued to emanate from the methodological and statistical communities. These ranged from the first dissertation on the design (Sween, 1971) to the statistical theorizing of Goldberger (1972), Rubin (1977), and Sacks and Ylvisaker (1976). The dissertation by Trochim (1980) brought together the academic and compensatory education traditions in a general discussion of regression-discontinuity supplemented with illustrative analyses.

By the late 1970s, the design was being included in both introductory (Kidder, 1981) and advanced (Cook & Campbell, 1979; Judd & Kenny, 1981) methodology texts. However, instances of the use of the design outside compensatory education remained negligible, a study in the criminal justice area by Berk and Rauma (1983) being a notable exception. The purpose of this study was to examine the effects of extending unemployment insurance benefits to ex-offenders from the California prison system. The assignment measure was the reported total number of hours worked for each person in prison jobs or vocational education programs. The legislation set the cutoff for program eligibility at $1500 or 652 hours at the minimum wage paid to prisoners. Eligible ex-offenders could apply for unemployment insurance like anyone else. The amount of payment depended on the hours worked in prison with benefits ranging from $30 to $70 per week for up to 26 weeks. The outcome measure was a reflection of recidivism, specifically, a parole revocation that meant, in effect, a return to prison. Because this outcome is dichotomous, the pre-post functional form was assumed to be logistic. The results indicated that program participants were about 13% less likely to return to prison than they would have been without the program.

Despite the absence of use of regression-discontinuity, the academic history of the design documents the steady increase in the knowledge of its mechanics and points to likely increased use in the future.

THE COMPENSATORY EDUCATION TRADITION

In 1965, the U.S. Congress passed Public Law 89–10, the Elementary and Secondary Education Act (ESEA), of which Title I is most relevant here.[2] It was the first major piece of legislation in the Great Society program of Lyndon Johnson and the largest single federal education grant of its scope. Title I was a singular domestic achievement that required a balancing of several divergent interest groups and not a little political straightarming (Eidenberg & Morey, 1969). The Statement of Purpose (Section 201) holds that Title I is to

> provide financial assistance . . . to local educational agencies serving areas of concentrations of children from educational programs by various means . . . which contribute particularly to meeting the special educational needs of educationally deprived children.

Each local school district or LEA (Local Education Agency) develops a proposal describing the programs for which it requests funds and submits this to the state department of education, or SEA, for approval. The programs are usually (but not necessarily) confined to basic skill areas such as mathematics, reading, and language arts. In 1979, over $5.5 billion was authorized by Congress for Title I, and nearly $3.4 billion was finally appropriated. Approximately 9 million low-income children were served, and between 5 and 6 million of these were of elementary school age. Nearly 87% of all school districts received some Title I funds (NCES, 1979) that accounted for between 3% and 4% of all national elementary and secondary education expenses (U.S. Office of Education, 1979). Only those aspects of Title I that are relevant to the application of the regression-discontinuity design are discussed here. The reader is referred to documents of the Office of Education and Wick (1978) for a more general description of Title I.

Title I was the first major social legislation that specifically required routine evaluation of its programs. Section 205(a)(5) reads

> effective procedures, including provision for appropriate objective measurements of education achievement will be adopted for evaluation at least annually, the effectiveness of the programs in meeting the special educational needs of educationally deprived children.

In the Amendments of 1974 (Public Law 93–380), Congress attempted to improve and standardize the evaluation procedures

at the federal, state, and local levels with the development of research models that

> specify objective criteria which shall be utilized in the evaluation of all programs and shall outline techniques (such as longitudinal studies of children involved in such programs) and methodology (such as the use of tests which yield comparable results) for providing data which are comparable on a statewide and nationwide basis. (Section 151(f))

In doing so, Congress moved away from an approach emphasizing demonstration-type studies by instead attempting to combine the evaluation and reporting functions. The research methods were to be clearly specified and results could be aggregated by district, state, and nation. In 1974, a contract was awarded by the Office of Education to the RMC Research Corporation for the development of such models. The system that they generated is characterized by a common measurement metric and by three alternative research designs that were presumed to yield comparable results (Tallmadge & Horst, 1976; Tallmadge & Wood, 1978). The metric, termed the Normal Curve Equivalent score (NCE), is simply a standard score with a mean of 50 and a standard deviation of 21.06. Estimates of program effect were reported on this scale to enable aggregation of gains at higher levels.

In order to understand the methodological and contextual issues of regression-discontinuity, it is essential to become familiar with the design choices which school districts were given. The three research designs layed out by RMC differed in the manner in which they generated the null case expectation for the program and can be described briefly as follows:

> Model A: The Norm Referenced Model. With this model only the program group students are pre and posttested. In an attempt to avoid regression artifacts it is required that program students be selected by some measure (or measures) other than the pretest. Thus, all students are given a "selection" test (usually, the annual spring achievement test) and only the selected program students are

pre and posttested. The average test scale score is calculated for the pretest and posttest and these can then be converted to percentiles or NCE scores. Because the average test scale scores are used and these are based on test scores obtained from norming samples the program effect consists of any gain over and above that which would be expected in the norming sample group. The comparison group consists of the national norming sample for the test and for that reason is labelled a "pseudo" comparison group here.

Model B: The Comparison Group Model. Both program and comparison group students are pre and posttested with this model. Ideally, students are randomly assigned to group, thus yielding a true experimental design. In practice, this is not often possible, and presumably "comparable" groups are used as program and comparison, yielding a non-equivalent group design. The program effect consists of any gain in the average of the program group over and above the gain in the average of the comparison group.

Model C: The Special Regression Design. This is the regression-discontinuity design as described here (the term "Model C" will be applied to refer to the Title I implementation of the design). Two separate estimates of program effect are computed. The "regression-discontinuity estimate" is the difference between the program and comparison group regression lines at the cutoff point. The "regression-projection estimate" is the difference between the lines at the program group pretest mean.

All three models can be used with either standardized or locally developed tests and are distinguished in the model names with a "1" for the former and a "2" for the latter. For example, Model A1 is the norm referenced model with a standardized test; Model A2 is the same model with a locally devised exam.

A system of technical assistance was set up to aid LEAs and SEAs in the design and execution of the evaluations. In each of the ten national Department of Health, Education and Welfare regions a Technical Assistance Center (TAC), upon request, was allowed to provide advice, training, and sometimes, assistance with statistical analysis. A good summary of the Title I system in general and the models in particular can be found in U.S. Office of Education (1976) and Tallmadge and Wood (1978). A useful bibliography of Title I evaluation issues was compiled by Strand (1979).

FREQUENCY AND LOCATION OF USE

In order to determine where the regression-discontinuity design was used, interviews were conducted primarily with personnel of the TAC Centers in each of the national Office of Education regions. From these interviews, a list of school districts that were mentioned in connection with the design was compiled.

In all, about 60 school districts (out of a total of about 15,000) were mentioned as possible users of the design in the 1979-1980 academic year. Of these, about 40 were confirmed verbally. The majority of users were located in East Coast regions, specifically, in the states of New Jersey, Virginia, and Florida. In the Midwest and Western states use was primarily restricted to a few large school districts and to "special case" districts such as the Trust Territories.[3]

Another source of information on the use of the regression-discontinuity design is a survey conducted by the National Center of Education Statistics (NCES, 1979). Of the districts that responded, 87% had Title I programs in the 1978-1979 school year and 63% had used some type of model. Of these, 86% used Model A, about 2% used Model B, and about 2% used Model C. The remaining 10% of the districts used an alternative or locally devised one. In general then, while the regression-discontinuity design was used, it was not used frequently, especially relative to Model A.

REASONS CITED FOR NOT USING MODEL C

It is important, when examining why the regression-discontinuity design is or is not used, to bear in mind the alternatives that a school district has. While two other models are available, the choice is most often between Model A and Model C (Echternacht, 1980). Consequently, many of the reasons listed as factors in the decision pertain to the perceived advantages of one model over the other.

In interviews and site visits, ten reasons were frequently mentioned as important in the decision not to use the regression-discontinuity design. These reasons often parallel the major factors cited in other papers such as those by Echternacht (1980) and McNeil and Findlay (1979). It is possible that for any given district some of the reasons that were cited were ill-considered or inappropriate. Nevertheless, these reasons are included because they serve to define, at least in part, how the design is perceived. The following discussion delineates the ten reasons.

1. Model C is not chosen because it is likely to yield "negative" program effect estimates. By far, the most frequently given reason for not using regression-discontinuity is the commonly held perception that it tends to yield "negative" gains for estimates of program effect. This reason was cited by representatives from almost every TAC, and by many persons at the state and local level. Because this issue is of central importance in assessing the appropriateness and feasibility of the design, the controversy is considered separately in Chapter 6.

2. Model C is not chosen because Model A is close to what most school districts had already been using for Title I evaluation. Prior to development of the three models, many districts evaluated Title I programs simply by testing program students before and after the program and determining whether their gain exceeded some goal or norm. In many cases, the only change that was needed in order to use Model A correctly was the addition of the separate selection measure. This was often easy to accomplish because many school districts give an annual achievement test to all students, usually in the spring. During any given evaluation cycle this test could be used as the selection test (which must be administered to all students) and, in the following year, could be used to obtain posttest scores for the program students as well as for the selection test for the next evaluation. All that a district

needs to add is a pretest of the program students, and this is usually accomplished through a smaller fall testing program.

3. Model C is not chosen because it requires adherence to the cutoff point criterion for assignment. Several administrators at the school district level claimed they want greater discretionary power in assignment both to allow for favoritism and for the possibility that a given test score may be inaccurate. Two administrators stated that they would like to have a "range of cutoffs" rather than a single cutoff point. In the simplest case, one could have two cutoffs. All students scoring below the lowest cutoff would automatically be assigned to the program group, all those scoring above the higher group would automatically be assigned to the comparison group, while all those scoring between the cutoffs would be assigned at the discretion of administrators, teachers, and so on. Although this strategy has not been used in Title I evaluation, it is a useful possibility, as will be discussed in Chapter 3.

4. Model C is not used because of premature release of research results which erroneously showed it yields biased estimates of effect. One frequently cited reason for not using the regression-discontinuity design provides an illustration of the interaction between social contexts and methodological decisions. Subsequent to the original contract for the development of research models awarded to the RMC Research Corporation, the Office of Education awarded a follow-up grant to the same corporation for purposes of investigating the models in more detail. To examine whether Model C might yield biased estimates, the RMC Research Corporation acquired two sets of data used to norm standardized tests. The first consisted largely of scores from the Comprehensive Test of Basic Skills (CTBS), while the second was comprised of the 1977 norming data for the California Achievement Test (CAT) issued by the McGraw-Hill company. For both of these data bases, it was assumed that the tested students did not receive compensatory education services. For Model C, cases were assigned to either program or comparison groups using an arbitrary cutoff point and an estimate of program effect was calculated. Presumably, since no program was administered, the analyses should have yielded an average program effect in the vicinity of zero.

These tests were carried out in two phases, first on the CTBS data base, and second on the CAT data. The initial results for the CTBS data indicated that Model C tended to yield a biased effect in the direction of a positive result. Subsequent to this portion of the study, and before looking at the CAT data, persons from the RMC Research Corporation, in seminars given throughout the country, appropriately suggested that there may have been problems with Model C that had not been previously anticipated. When the data from the CAT data base was examined using the same procedures, all three models yielded estimates of effect in the vicinity of zero NCE units. At this point, several researchers from RMC began to search for reasons for this apparent discrepancy (Wood, 1979). The reexamination of the CTBS data base turned up the existence of nonnormal or skewed distributions. After transforming the data to minimize this skew, the estimates for Model C were recomputed and, while slightly positive, approached the expected zero value. The final conclusion of the study was that Model C estimates are not biased, although they may be affected adversely by floor and ceiling effects and abnormalities in the marginal distributions (Stewart, 1980a, 1980b).

A distinction needs to be made between the actual results from this study and the effect of premature release of the results. A number of persons, especially at the TAC level, stated that to their knowledge RMC had shown that Model C did not work correctly and should be avoided. None of those who mentioned this seemed aware of the subsequent investigation and the change in the overall conclusion. While it is reasonable to expect that eventually this information was passed along, it is hard to gauge the effect that premature partial information had on the initial selection of research models by school districts. It is reasonable to argue that some districts used this information in deciding

against Model C, while others who wished to use Model A cited it to enhance the credibility of their decision.

Even granting that a consistent positive program effect may have been detected in this study, one must be careful to distinguish between the conclusion that Model C yields biased estimates or that implementation problems (e.g., testing problems, matching problems, etc.) degrade estimates of effect with this design. Each conclusion has different implications. If Model C yields biased estimates it would have been reasonable to recommend suspension of its use. If, however, with typical Title I implementations there are problems which act to degrade estimates, one can address these problems directly and attempt to improve the validity of the results.

5. Model C is not chosen because it is more difficult technically than Model A. Model C may appear to be more technically difficult to some degree because it is based on notions of regression lines and projections from regression lines, ideas that are not easily understandable to those who are unfamiliar with research and statistical analysis. In addition, several respondents mentioned that in presenting the three research models, the materials used and the explanations provided for Model C were the most difficult to understand. It is not at all clear whether this is a deficiency in the manner of presentation or a problem that stems from inherent differences in the technical difficulty of the models. Nevertheless, it is a commonly held perception that Model C is more difficult than Model A, that it requires a more technically trained staff, and that it is most appropriately used by districts that have ready access to computer facilities capable of computing the necessary regressions (Bridgeman, 1979). Although all of these considerations have some validity, the investigation of where Model C has been used turned up several districts that used the design despite the lack of sophisticated personnel and facilities. In most cases, these districts either pooled their resources with other small districts or hired outside consultants who were able to conduct the analyses.

6. Model C is not used because it requires testing of comparison students. Model C requires pre- and posttesting of comparison students. While it is not absolutely necessary that all non-Title I students be tested (e.g., a randomly selected subsample of non-Title I students could be used instead), the model does require more than simply the pre- and posttesting of program students. For those districts that use a district-wide annual standardized achievement test this requirement poses no problems. However, for other usually smaller districts, the increased cost of additional testing is a considerable burden.

7. Model C is not used because it requires at least 30 students in the program group. The original Users Guide for Title I Evaluation suggests that Model C should not be used unless there are at least 30 students in each group. The basis of this is that with larger samples one can achieve greater statistical precision and more accurate estimates of regression lines (McNeil, 1977). While this is certainly an important consideration, the requirement of at least 30 students was arbitrary. Estimation of a regression line is more critical for the comparison group where participants are typically more plentiful. A variation of the regression-discontinuity design that tests for differences between a program group posttest mean (rather than a regression line) and the projection from a comparison group line is discussed in Chapter 3.

8. Model C is not used because it requires that within-group correlations be at least .4. The original Users Guide suggests that Model C should not be used unless the pretest and posttest correlation is at least .4. While it may be desirable, from a conceptual point of view, to have a strong pretest-posttest correlation (McNeil, 1977), this is not technically necessary (see Echternacht, 1978). In fact, any measure can be used for the preprogram measure whether it is correlated with the posttest or not. Figure 2.2 presents an example of the regression-discontinuity design where the pretest and posttest are uncorrelated. In this case the regression line in each group is flat, that is, the slope is equal to zero. Here, one has effectively reduced the design to a test between the program and comparison posttest means. In this case, because there is no relationship between the pretest and posttest, one "approximates" random assignment to program. It is difficult to conceive of such a case occurring in Title I evaluation because of the desire to assign students to a program group on the basis of some measure related to achievement. Nevertheless, this example is included to illustrate that a low pretest-posttest correlation, should it occur, will not by itself result in biased estimates of program effect.

9. Model C is not chosen because it can cause individual schools to lose Title I teacher positions. Another problem that occurs in

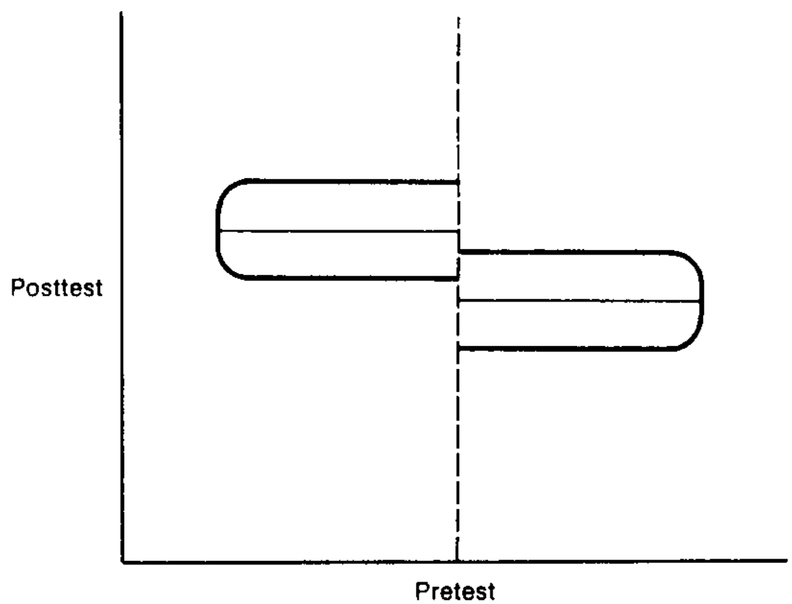

Figure 2.2 Regression-Discontinuity When the Pretest and Posttest Are Uncorrelated

relation to the sharp cutoff point concerns the allocation of Title I teachers to schools. From one year to the next, the number of children within any given school in a district who are eligible for Title I services might change considerably. If Title I teachers are assigned to schools on the basis of the number of children who will be served, there is a possibility that a given school will lose Title I teachers from one year to the next. Few schools are likely to accept unchallenged a reduction in the number of teachers they are allocated. Other types of allocation procedures could be developed as in one case, where a creative approach that capitalized on an already existing busing program equalized teacher allocation by means of moving students from one school to another.

10. Model C is not chosen because the SEA discourages its use. Another reason cited for not using regression-discontinuity is that in many states the SEA strongly recommends the use of another design, most often Model A. While the decision to use a particular research design technically is the option of the LEA, the SEA does have the right to review and approve program proposals. While no respondent stated that the SEA rejected their use of Model C, many mentioned that the existence of an explicit or implied state policy favoring Model A did have an effect on their decision.

REASONS CITED IN FAVOR OF USING MODEL C

Fewer reasons were listed as favoring the use of Model C. They are summarized as follows:

1. Model C is used because it is conceptually closest to the idea of Title I training. Title I programs are supposed to be given to those students who need them most. More than the other two models, Model C provides an explicit quantified measure of such need and a clear decision rule for the allocation of service. There are other ways to exploit the correspondence between the regression-discontinuity design and allocation procedures used

for Title I. For example, each school district must designate which schools are eligible for Title I programs on the basis of a quantified measure of poverty. If a cutoff on this measure is used to assign the schools, one can conceive of a regression-discontinuity analysis of the effects of Title I training at the school level. Variations of this type are discussed in Chapter 3.

2. Model C is used because it fits in well with annual district-wide testing programs. Those districts that have annual testing programs find that no additional testing is required for Model C. The test given in each year acts as the posttest for the current year and as the pretest and selection measure for the year to follow. In the typical implementation of Model A, all students in the school district are tested (for example, in the spring), those students selected for Title I service are pretested (for example, in the fall), and all students are posttested the following spring. Thus, in the annual cycle, the spring testing provides both the selection measure for the subsequent year and the posttest for the current year. However, it is necessary to include a separate pretest. Essentially, this issue involves a trade-off between the two models. Model A requires less testing in that only program students are given the pretest and posttest (although all are given the selection test). Model C involves less testing in that a separate selection measure is not needed. Depending on the type of district testing program that exists, a given district might find it less costly to use either Model A or Model C.

3. Model C is used because it is perceived as methodologically stronger than Model A. Another set of reasons cited in favor of using Model C is related to perceptions about the quality of the design. Those districts that employ researchers who have been specifically trained in research methodology are more likely to by
aware of the academic history of Model C and of its advanta ison
from a methodological viewpoint. In addition, because oca in a
history and the common perception within Title I ev Chapter
circles that Model C is the most technically difficult oCutoff. This
models, a degree of higher status might be accdn the cutoff

districts that use the model. One Title I evaluator from a large metropolitan school district said that he felt his district should [illegible] in attempting to implement more technically difficult res[illegible]ions. This is more likely to be the case for districts that have w[illegible]ined staffs and sufficient computer facilities.

SUMMARY OF USE ISSUES

The regression-discontinuity design is used under the name of Model C for Title I evaluation, although infrequently. While it is certainly appropriately used in this context, it tends to be judged relative to Model A. From the typical school district's perspective, Model A is easier to understand, easier to use, less costly, and likely to yield "favorable" results. If both models were of similar methodological quality, Model A would clearly be the better choice. In Chapter 6, the issue of the relative quality of these two designs is considered.

To summarize, we can see in the history of Title I the development of an evaluation system of national scope, specifying three presumably "equivalent" (in terms of program effect estimates) research designs, one of which is the regression-discontinuity design of interest here. These conditions make the context of Title I evaluation the richest source that is currently available for information on the application of the regression-discontinuity design.

NOTES

1. The situation is actually more complex. Students may be awarded a scholarship on the basis of several criteria (e.g., high school grades, test scores, recommendations), but only test score assignment was studied here.

2. Under the Reagan administration, the educational legislation has been considerably reorganized. For the most part, what is referred to here as "Title I" now falls under the name of "Chapter I." The Title I evaluation system is no longer formally required in the federal legislation, but nevertheless has been maintained with little or no change by many individual states. While the term "Title I" is used in this volume for historical consistency, the issues considered are applicable under present circumstances.

3. For instance, the design was used in Puerto Rico because there was no suitable achievement test in Spanish that had normative data (thus making Model A untenable).

3

Design Variations

The intent of this chapter is to expand upon the basic definition of the regression-discontinuity design as outlined in Chapter 1 by suggesting alternative variations and applications. It is important to keep in mind that the major distinguishing feature of the design is the assignment to conditions on the basis of a cutoff on some quantified measurement. Any analysis that is based at least in part on such a strategy will be considered here a variation of the design.

The regression-discontinuity design is far more versatile in principle than its present applications might suggest. This chapter is devoted to consideration of some of the major useful variants. Six major areas are considered: different strategies for handling assignment to condition; measurement variations; program-related variations; alternative postprogram measures; the