3 Regression-discontinuity designs

In this chapter and the next, I consider two specific kinds of natural experiments: regression-discontinuity designs in this chapter, and instrumental-variables designs in the next. When appropriately invoked, both kinds of designs meet the key definitional criterion of natural experiments—namely, random or as-if random assignment. Yet they differ in other ways, and their use tends to raise characteristic issues of discovery, analysis, and interpretation that are specific to each kind of design. This makes it valuable to discuss these two kinds of natural experiments separately in various chapters, where they can each be given more detailed attention.

The regression-discontinuity design was proposed by Thistlethwaite and Campbell (1960) and discussed extensively by Trochim (1984). Yet, for somewhat unclear reasons—and despite the extensive efforts that Donald Campbell's research team at Northwestern University put into studying it—this research design was relatively underutilized for several decades after Campbell and his team originally proposed it. However, empirical applications have exploded in the social sciences over the past decade or so. In this chapter, I survey many recent applications, as a way to shed some light on the art of discovering productive opportunities for these modes of research. Issues of analysis and interpretation are mostly postponed for Parts II and III of this book, though some mention of them is also made here.

3.1 The basis of regression-discontinuity analysis

As part of a social or political process, individuals or other units are sometimes assigned to one or the other category of the independent variable (e.g., the treatment or control groups) according to whether they are above or below a given threshold value on some covariate or pre-test. For units very near the threshold, the process that determines treatment assignment may be as good as random, ensuring that these units will be similar with respect to potential

confounders. This in turn opens the possibility for a *regression-discontinuity design*, which may allow for a more compelling causal inference about the impact of the treatment on the dependent variable. The name of the design does not imply that regression analysis needs to be used to analyze data arising from such studies.

The contrast with the standard natural experiment is that as-if random assignment specifically involves the position of subjects in relation to the key threshold.[1] For example, in their study of the National Merit Scholarship program in the United States, Thistlethwaite and Campbell (1960) compared students who received public recognition of scholastic achievement with a similar group of students who did not, with the goal of inferring the impact of recognition on subsequent academic achievement. All students who scored above a given threshold on a Scholastic Aptitude Test received a Certificate of Merit—and thus had their names "published in a booklet distributed to colleges, universities, and other scholarship granting agencies and . . . received approximately two and one half times more newspaper coverage than commended students" (Thistlethwaite and Campbell 1960: 310). Students who scored below the threshold, however, merely received commendations, which confer less public recognition of scholastic achievement.

In general, students who score high on qualifying exams will be different from those who score low, in ways that matter for outcomes such as receipt of scholarships or later scholastic achievement. For instance, if high scorers have greater initial ability than low scorers, on average, then better *ex post* performance among the high scorers could be due to the effects of public recognition, differences in initial ability, or both. Comparisons between high scorers and low scorers may therefore be misleading for purposes of inferring the effect of receiving public recognition. Even comparing the change in outcomes from pre- to post-test could be misleading, for instance, if high scorers receive more coaching, invitations to apply to elite colleges, or differ in other ways besides the public recognition they receive in the form of Certificates of Merit.[2]

However, given the role of unpredictability and luck in exam performance, students just above and below the key threshold should be very similar, on average. In particular, they should not differ systematically with respect to the

[1] More technically, in a regression-discontinuity design, treatment assignment is determined by the value of a covariate, sometimes called a forcing variable, and there is a sharp discontinuity in the probability of receiving treatment at a particular threshold value of this covariate (Campbell and Stanley 1966: 61–64; Rubin 1977).

[2] Comparing the change on a dependent variable in the treatment and control groups, from pre-test to post-test, is *difference-in-differences* analysis.

ex post outcomes each student *would* experience if he or she were given a Certificate of Merit, or instead merely received a commendation. Thus, the outcomes that each student *would* experience in the presence and absence of public recognition should be independent of whether he or she actually received a Certificate.[3] This presumes that students cannot readily "sort" themselves on one side or another of the relevant threshold in ways that may be related to their potential outcomes; moreover, officials must not choose the threshold strategically to select particular candidates for certificates, who might differ from students in the control group in various ways. I discuss these conditions in more detail below. If such conditions hold, however, assignment to receive a Certificate of Merit can plausibly be considered as-if random in the neighborhood of the threshold. Then, comparisons near the threshold allow an estimate of the causal effects of certificates, at least for the group of students with scores near the threshold.[4]

Figure 3.1, which is similar to a figure in Thistlethwaite and Campbell (1960), shows why the "regression-discontinuity design" has this name. The figure plots average values of a hypothetical outcome variable—here, the percentage of students receiving college scholarships—against scores on the qualifying exam. The horizontal axis expresses exam scores in arbitrary units, because the specific threshold exam score varied across different US states; the vertical axis plots the average outcome for each student at each value of the exam score. Exam scores are scaled so that the key cutoff is at 11: students who scored 11 and above on the exam were awarded certificates, while those below the threshold merely received commendations. For ease of presentation, exam scores are rounded to the nearest mid-point between two integers (4.5, 5.5, and so on). The vertical dotted line above the score of 11 illustrates that this is the key value of the exam score determining assignment to certificates.

In Figure 3.1, we see three very different patterns. In Series A at the top of the figure, there is a clear jump in the value of the outcome variable between exam scores of 10 and 11. Thus, the group of students who scored 10 on the

[3] Chapter 5 more formally introduces the idea of *potential outcomes*. A potential outcome is the outcome a unit *would* experience, if it were assigned to treatment or control. Since each unit can only be assigned to at most one group, at least one of these potential outcomes is unobserved by the researcher.

[4] Oddly, Thistlethwaite and Campbell (1960) remove from their study group Certificate of Merit winners who also won National Merit Scholarships. Only Certificate of Merit winners were eligible for National Merit Scholarships, which are based on grades as well as exam scores. Thus, the control group includes students who *would have* won merit scholarships had they received Certificates of Merit and those who would not have, while the treatment group includes only the latter type. If type is related to potential outcomes, as seems plausible, this should lead to bias in the estimation of treatment effects (see Chapter 5).

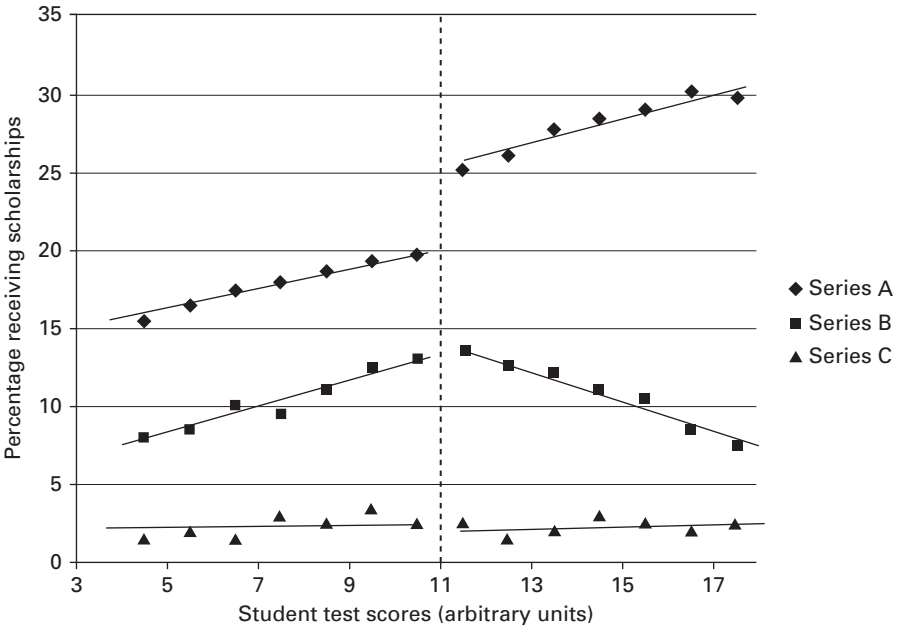

Figure 3.1 Examples of regression discontinuities

exam performed systematically worse on the outcome variable—i.e., they received proportionately fewer scholarships—than students who scored an 11 on the exam and thus received public recognition in the form of certificates. If we believe the assertion that potential outcomes are independent of exam scores for those who scored very near the threshold—and thus, assignment to certificates is as good as random—this is credible evidence for a causal effect of Certificates of Merit.

Here, the solid lines on each side of the vertical dashed line represent the regression lines relating this outcome variable to the exam score to the right and the left of the threshold.[5] Note that in Series A, the intercepts of these two regression lines—that is, the point at which they coincide with the dashed vertical line between 10 and 11—are clearly different. This discontinuity in the intercept of the regression line is in fact what gives "regression-discontinuity" designs their name. However, as we shall see in more detail in Part II, there is no need to use linear regression analysis to estimate causal effects using regression-discontinuity designs—and the use of regression can even be misleading.

[5] Note that the regression of the *average* outcome within each interval on the qualifying score produces the same fitted line as the regression of actual outcomes on the qualifying score, so long as the intervals are weighted by the numbers of students in each interval.

Series C at the bottom of Figure 3.1 presents an opposite case. Here, there is no detectable jump in outcomes between 10 and 11. Indeed, the regression lines on either side of the key threshold are nearly identical, as are their intercepts. While I discuss the role of sampling error in Chapters 5 and 6 and describe more systematic procedures for verifying the existence of causal effects, Series C does not seem to suggest any causal effect of certificates for students just above and just below the threshold.

Finally, Series B in the middle of Figure 3.1 seems to present a more mixed case, since the slopes of the regression lines on either side of the key threshold clearly differ: the regression line slopes up to the left of the threshold and slopes down to its right. Yet, as in Series C, the intercepts are nearly identical: there is no discernible difference in *ex post* outcomes between those who scored 10 and those who scored 11. In general, such a figure does not present compelling evidence for the effect of Certificates of Merit. For instance, recall that the outcome here is receipt of an academic scholarship. A pattern such as that in Series B might be produced if, for instance, scholarship committees tend to give scholarships to low-income students with relatively high abilities, and if income and ability are both correlated with test scores. Then, test scores might be positively associated with receipt of scholarships, up to a certain point, since low-scoring students may also tend to have low incomes. Yet beyond a certain income threshold, high-scoring students become less attractive candidates, since most of them may be too rich for scholarships. This could induce a negative relationship, on average, between test scores and receipt of scholarships among high scorers.

This last point underscores a central feature of regression-discontinuity designs: the assumption that assignment to the treatment (here, Certificates of Merit) is as-if random is typically only plausible for students just above and just below the key threshold—that is, in the neighborhood of the threshold. The problem with using the slope of regression lines to infer causal impacts in regression-discontinuity designs is that unobserved confounders, such as income, can influence the slopes of these regression lines—and even their intercepts. This suggests that the most reliable comparison group for students who scored 11 on the exam is students who scored a 10. The performance of students who scored, say, 17 or 4 may be less relevant, since many systematic differences—not just the luck of the draw on a particular exam—may distinguish these groups of students.[6]

[6] As Thistlethwaite and Campbell (1960: 311) put it, the effect of certificates may be "strictly demonstrated only for aptitude intervals adjacent to the cutting point, and inferences as to effects of the [Certificate of Merit] award upon persons of other ability levels would be made in hazard of unexplored interactions of award and ability level."

The example also suggests, in a cursory way, some of the potential dangers of using regression analysis to draw inferences about causal effects in regression-discontinuity designs: the slopes of the regression lines in Series B of Figure 3.1, and even their intercepts, are driven by data points that are relatively far from the key regression-discontinuity threshold—and as such, can be subject to confounding in a way that simple comparisons between students who scored 10 or 11 on the exam are not. I will return to these points in more detail in Chapters 5 and 6, when I discuss procedures for analyzing data from regression-discontinuity designs.

Finally, the example also illustrates another foundational point about regression-discontinuity designs: while limiting the analysis to students who scored 10 or 11 on the exam limits confounding—and thus bolsters the ability to draw valid inferences about the causal effect of recognition for this group of students—it may also limit the ability to say something about the effect of public recognition for students far from the cutting point. This point is typically discussed in terms of a trade-off between "internal validity" and "external validity" (Campbell and Stanley 1966). How sharp this trade-off is depends on the application: while in some settings the achievement of internal validity can come at the cost of external validity, in other settings it may not.

These central issues of analysis and interpretation are best saved for later sections of the book, however. The central goal of this chapter is to survey real social-scientific applications, as a way of answering the question: from where do regression-discontinuity designs arise?

3.2 Regression-discontinuity designs in the social sciences

Regression-discontinuity designs have become increasingly common in the social sciences over the past decade or two, particularly in economics and political science. Following Thistlethwaite and Campbell (1960), regression-discontinuity designs based on entrance exams have been used by many scholars; for instance, Matsudaira (2007) studies the effect of a remedial summer school program that is mandatory for students who score less than a cutoff level on a test (see also Jacob and Lefgren 2004).

However, scholars have greatly expanded the substantive domain of studies using this methodological approach, as Tables 3.1 and 3.2 attest. The first table lists some generic sources of regression-discontinuity designs used in recent

Table 3.1 Selected sources of regression-discontinuity designs

Source of RD design	Units in study group (at RD threshold)	Treatment variables	Outcome variables
Entrance exams	Students, others	Public recognition of scholastic achievement	Educational achievement
Population thresholds	Municipalities, citizens	Voting technologies	Effective turnout
		Federal funds	Voting behavior
		Cash transfers	Voting behavior
		Electoral rules	Voting behavior
		Politicians' salaries	Candidate entry
Size-based thresholds			
Voter numbers	Voters	Voting by mail	Voting behavior
School size	Students	Class size	Educational achievement
Firm size	Firms	Antibias laws	Productivity
Eligibility criteria			
Poverty rank	Municipalities	Antipoverty programs	Voting behavior
Criminality index	Prisoners	High-security incarceration	Recidivism
Age-based thresholds			
Voting age	Voters	Past voting	Turnout
Birth quarter	Students	Years of education	Earnings
Close elections	Candidates/parties	Incumbency	Candidates' performance
	Firms	Campaign donations	Public works contracts

Note: The table provides a non-exhaustive list of sources of regression-discontinuity designs. Specific studies are listed in Table 3.2. RD, regression discontinuity.

social-scientific research; the second provides a non-exhaustive yet large list of specific recent studies using this research design. For each study, the table lists the author(s), substantive focus, country in which the study took place, and the source of the regression discontinuity.[7] (Table 3.2 also lists whether a simple difference-of-means test is used to analyze the data, a topic for which I again postpone further discussion until later chapters.)

How are these designs discovered and leveraged in the service of diverse research agendas? As with standard natural experiments, discovering useful regression discontinuities is as much an art as a science. Yet, as the survey in this section will show, regression-discontinuity designs developed in one place or context have often served as inspiration for designs in other settings. One

[7] Most of these studies take place within a single country, a topic I will discuss elsewhere.

Table 3.2 Examples of regression-discontinuity designs

Authors	Substantive focus	Source of regression discontinuity	Country	Simple difference of means?
Angrist and Lavy (1999)	Effect of class size on educational achievement	Enrollment ceilings on class sizes	Israel	No
Boas and Hidalgo (2011)	Effect of incumbency on access to media	Near-winners and near-losers of close elections	Brazil	Yes
Boas, Hidalgo, and Richardson (2011)	Effect of campaign donations on access to government contracts	Near-winners and near-losers of close elections	Brazil	No
Brollo and Nannicini (2010)	Effect of partisan affiliation of incumbent mayor on federal transfers	Near-winners and near-losers of close elections	Brazil	No
Brollo et al. (2009)	Effect of federal transfers to municipalities on corruption and candidate quality	Population-based revenue-sharing formula	Brazil	No
Chamon, de Mello, and Firpo (2009)	Effects of second-round mayoral runoffs on political competition and fiscal outcomes	Population-based discontinuity in voting system	Brazil	No
Dunning (2010b), *Dunning and Nilekani* (2010)	Effects of caste-based quotas on ethnic identification and distributive politics	Rule rotating quotas based on caste population proportions[a]	India	Yes
Eggers and Hainmueller (2009)	Effects of holding legislative office on wealth accumulation	Near-winners and near-losers of close elections	UK	No[b]
Ferraz and Finan (2010)	Impact of monetary incentives on politician quality and performance	Salary caps for politicians based on municipal size	Brazil	No
Fujiwara (2011)	Effects of second-round runoff on first-round vote shares	Population-based discontinuity in voting system	Brazil	No[b]
Fujiwara (2009)	Effects of electronic voting technology on de facto enfranchisement and fiscal policy	Thresholds based on numbers of registered voters	Brazil	No[b]
Gerber, Kessler, and Meredith (2011)	Effects of campaign mail on voter turnout and vote choice	Discontinuity in rule used to select households to receive mail[c]	US	Yes
Golden and Picci (2011)	Incumbency advantage and distribution of pork	Within-party comparisons of near-winners and near-losers	Italy	No

Hidalgo (2010)	Effects of electronic voting technology on de facto enfranchisement and fiscal policy	Thresholds based on numbers of registered voters	Brazil	Yes
Kousser and Mullin (2007), *Meredith and Malhotra* (2011)	Effects of voting by mail on turnout and vote choice	Population-based thresholds used to select precincts for voting by mail	US	No
Lerman (2008)	Social and political effects of incarceration in high-security prisons	Criminality index used to assign prisoners to security levels	US	Yes[d]
Litschig and Morrison (2009)	Effects of federal transfers to municipalities on incumbent reelection probabilities	Discontinuities based on population-based revenue-sharing formula	Brazil	Yes
Manacorda, Miguel, and Vigorito (2011)	The effect of a cash-transfer program on support for the incumbent government	Discontinuity in program assignment based on a pre-treatment eligibility score	Uruguay	Yes
Meredith (2009)	The effect of past voting on subsequent turnout and partisanship	Voting-age restrictions	US	No[b]
Titiunik (2009)	Incumbency advantage in mayoral elections	Near-winners and near-losers of close elections	Brazil	Yes

[a] This RD design has an element of true randomization.

[b] Local linear regression with or without covariates, or polynomial regression without covariates, is used in these studies, and graphic difference-of-means comparisons are made.

[c] The rule is a function of income and other variables.

[d] Regression-discontinuity and instrumental-variables designs are both used.

goal of this survey is therefore to familiarize readers with these ideas, so they might be used in disparate contexts.

3.2.1 Population- and size-based thresholds

One of the most common regression-discontinuity designs, broadly conceived, is one in which researchers take advantage of the fact that a policy intervention is allocated to some geographic or political units but not to others according to some function of population size—and a rigid cutoff is adopted, so that units above or below the threshold differ in their exposure to the intervention. Like students with similar pre-test scores located on either side of a threshold score for public recognition, geographic units with similar population sizes may differ sharply in their exposure to the policy intervention. Error in the measurement of population size may introduce a further element of chance into location of units just above or just below the key threshold. If assignment to one side or the other of the threshold is as good as random, comparison of units on either side of the threshold can then be used to estimate the effect of exposure to the intervention. This presumes, of course, that census officials or other policy-makers do not manipulate the measurement of population size, or that politicians do not choose the key threshold to include or exclude particular units in the intervention group.

Hidalgo (2010), for instance, uses regression-discontinuity designs to study the impact of the introduction of electronic voting machines on de facto enfranchisement, patterns of partisan support, and fiscal policy in Brazil (see also Fujiwara 2009). In the 1998 Brazilian legislative elections, municipalities with more than 40,500 registered voters used electronic ballots, while municipalities with fewer than 40,500 voters continued to use traditional paper ballots. The introduction of electronic voting machines, which feature an ATM-like interface and display of candidates' names, party affiliations, and photographs for voters' confirmation, was thought to ease the process of navigating the complex voting system for illiterate and less well-educated voters, in particular. Brazil uses an open-list proportional representation system with high district magnitudes, meaning that many candidates are often elected from the same districts. In elections to the federal Chamber of Deputies, for instance, the "districts" are states; voters in large states such as São Paulo must elect as many as 70 deputies, choosing from among more than 200 candidates. This implies a bewildering number of individual candidates from which voters may choose, a process made all the more difficult for illiterate or less well-educated voters using traditional paper ballots because

of the need to write in candidates' names (or 5- or 6-digit identification numbers) for each of the offices being contested in a given election. The difficulty and inconvenience of navigating this process surely contributed to giving Brazil the highest rates of invalid or blank ballots in Latin America, reaching an average of about 33 percent of all ballots cast in the pre-reform period between 1980 and 1998 (as compared to the regional average of 8.5 percent).

The basis for the regression-discontinuity design is the use of the population threshold of 40,500 registered voters.[8] Municipalities "just above" and "just below" the threshold of 40,500 registered voters should on average be highly similar. Indeed, since the threshold was announced in May of 1998 and the number of registered voters was recorded in the municipal elections of 1996, municipalities should not have been able to manipulate their position in relation to the threshold. There is no apparent evidence that the particular threshold was chosen by municipalities to exclude or include particular municipalities in Brazil. (The threshold was applied uniformly throughout the country, with the exception of four states.) If location just above or just below the population cutoff is as good as randomly assigned, comparisons around the threshold can validly be used to estimate the impact of the voting technology.

Hidalgo (2010) and Fujiwara (2009) find that introduction of electronic voting increased the effective franchise in legislative elections by about 13–15 percentage points or about 33 percent—a massive effect that appears more pronounced in poorer municipalities with higher illiteracy rates. This is likely because electronic voting greatly simplified the process of voting, in a setting in which an open-list proportional representation electoral system with high district magnitude implied that voters would sometimes have to choose from among hundreds of candidates. In contrast to paper ballots, in which voters had to write in the name of their preferred candidate or party, voting machines use an ATM-like interface to display candidates' names, party affiliations, and photographs, which was thought to ease the voting process for illiterate voters, in particular.

Moreover, the introduction of voting machines led to substantial declines in the vote shares of incumbent "machine" parties in several Northeastern states. Thus, Hidalgo (2010) suggests that the introduction of voting machines contributed towards the strengthening of programmatic parties in Brazil.

[8] An earlier, more limited reform in 1996 used a population cutoff of 200,000 registered voters; see Chapter 10.

While the increasing importance of programmatic parties has been noted by several recent observers of Brazilian politics, the study by Hidalgo (2010) identifies a plausible cause of this tendency, and his use of a regression-discontinuity design suggests that it is highly unlikely that differences in de facto enfranchisement are driven by confounders, rather than by the introduction of the electronic voting technology.[9]

Regression-discontinuity research designs based on such population thresholds have also been used extensively to study the economic and political impacts of federal transfers in Latin America and other regions. T. Green (2005), for example, calculates the electoral returns of the Mexican conditional cash-transfer program, PROGRESA, using a regression-discontinuity design based on a municipal poverty ranking.[10] Litschig and Morrison (2009) study the effect of federal transfers on municipal incumbents' vote shares in Brazil, while Brollo et al. (2009) study the effect of such transfers on political corruption and on the qualities of political candidates. Both of these latter studies take advantage of the fact that the size of some federal transfers in Brazil depends on given population thresholds. Thus, these authors can construct regression-discontinuity designs in which municipalities just on either side of the relevant thresholds are compared.[11]

Scholars have also used population-based thresholds to study the political and economic impact of electoral rules. For instance, the Brazilian federal constitution states that municipalities with less than 200,000 registered voters must use a single-ballot plurality rule (a first-past-the-post system where the candidate with the most votes is elected) to elect their mayors, while municipalities with more than 200,000 voters must use the dual-ballot plurality rule (second-round "runoff"), a system where voters may vote twice. Fujiwara (2011) finds that in the neighborhood of this threshold, the change from single-ballot to second-round runoff systems increases voting for third-place finishers and decreases the difference between third-place and first- and second-place finishers—a finding consistent both with strategic voting and with the observations of Duverger (1954) and G. Cox (1997) that in elections for m seats, $m + 1$ candidates should command most of the votes. Chamon, de Mello, and Firpo (2009) extend this same idea, finding that the greater political competition induced by the discontinuous change in electoral rules in mayoral elections at the threshold of 200,000

[9] This study and the results are discussed more extensively elsewhere, for instance, in Chapters 5 and 10.
[10] This is the same program studied by De La O (forthcoming), as discussed in Chapter 2.
[11] See also the study in Uruguay by Manacorda, Miguel, and Vigorito (2009) discussed below.

voters induces greater investment and reduces current expenditures, particularly personnel expenditures.

Population-based thresholds have been used to study many other topics as well, among them legislative productivity and voter turnout. A constitutional amendment in Brazil sets salary caps on the wages of local legislators as a function of the population size of municipalities. Using this rule to construct a regression-discontinuity design, Ferraz and Finan (2010) find that higher wages increase legislative productivity and political entry but also increase reelection rates among incumbent politicians. Kousser and Mullin (2007) and Meredith and Malhotra (2011) take advantage of an administrative policy in California in which all voters in some precincts are assigned to vote by mail, based on an arbitrary threshold of the number of registered voters.[12] Analyzing the 2008 California presidential primary, and comparing precincts just above and below the relevant threshold, Meredith and Malhotra (2011) show that the practice of voting by mail both increases the probability of selecting withdrawn candidates and affects the relative performance of candidates remaining in the race.

This survey suggests that the use of population-based thresholds by politicians and policy-makers to allocate benefits or introduce other policy innovations has quite often provided the basis for productive regression-discontinuity designs. Obviously, opportunities will not always exist for the construction of this type of regression-discontinuity design in the context of a given substantive research agenda. Yet, the recent record suggests that the potential range of applications is very broad: many more possibilities for using this kind of design may also await researchers alert to their existence. Indeed, population-based thresholds and ceilings appear to be fairly widely used in the allocation of public benefits and other policies across a number of different countries, and scholars investigating a range of different causal hypotheses might be able to take advantage of such instances to construct regression-discontinuity designs.

Scholars have also used many other kinds of size-based thresholds, beyond those based on the populations of municipalities, to construct regression-discontinuity designs. A well-known example, which illustrates both the

[12] The election law states that if the number of registered voters in a precinct is no larger than 250 on a given date prior to the election, election officials may *choose* to establish voting by mail in that precinct. Because there are in fact some precincts above the 250-voter limit that also have vote-by-mail, and some below it that do not, this can be seen as an example of "fuzzy" regression-discontinuity analysis (discussed below and in Chapter 5). See Kousser and Mullin (2007) or Meredith and Malhotra (2011) for further details.

strengths and limitations of this design, is Angrist and Lavy (1999), who analyze the effects of class size on educational achievement. In general, comparing educational outcomes in schools with large classes to schools with small classes would be misleading, for purposes of inferring the impact of class size: students in schools with large classes differ in both potentially observable (income, ethnic background) and more difficult-to-measure (parental backgrounds, student motivation) ways.

However, Angrist and Lavy (1999) build on a requirement in contemporary Israel—known as Maimonides' Rule, after the twelfth-century Rabbinic scholar—that requires secondary schools to have no more than 40 students per classroom. Thus, in a school in which the total enrollment is near this threshold or its multiples—e.g., schools with just under 40, 80, or 120 students—the addition of a few students to the school through increases in overall enrollment can cause a sharp reduction in average class sizes—since more classes must be created to comply with the rule. The educational achievement of students in schools the enrollments of which are just under the threshold size of 40 (or 80 or 120) can then be compared to students in schools that are just over the threshold and who were thus reassigned to classrooms with a smaller number of students.

The key assertion is that, on average, schools just above each threshold size are just like schools below the threshold size—save for average class size. Thus, assignment to class sizes for students in these schools is as good as random, and the effect of class size can therefore be estimated in the neighborhood of the thresholds. A key feature of the design is that students probably do not self-select into smaller classrooms, since the application of Maimonides' Rule is triggered by increases in schoolwide grade enrollment. The design is interesting, and there is a plausible claim of as-if randomness in the neighborhood of the threshold—though this study also raises interesting issues of analysis and interpretation to which I turn in later chapters.

Beyond class size, scholars have used size-based thresholds to study effects for many other kinds of units. For example, Hahn, Todd, and Van Der Klaauw (1999) study the effect of an anti-discrimination law that only applies to firms with at least 15 employees. In such contexts, one must worry about confounding decisions related to the size of the firm; perhaps racist employers hold their size under 15 employees to avoid adverse consequences from the anti-discrimination law. Yet, such concerns notwithstanding, the potential range of applications of regression-discontinuity designs based on size-based thresholds also appears large.

3.2.2 Near-winners and near-losers of close elections

A different kind of regression-discontinuity design, which has also found growing use in recent years, takes advantage of the fact that in very close and fair elections, there is an element of luck and unpredictability in the outcome; thus, underlying attributes of near-winners may not differ greatly from near-losers. As Lee (2008) suggested in his study of the US Congress, near-winners and near-losers of close elections should be nearly identical, on average. Thus, comparisons of these groups can be used to estimate the effects of winning office. Indeed, due to the element of luck in very close elections, candidates and parties are plausibly assigned as-if at random to office in these elections. As with other natural experiments, whether the assumption of as-if random is valid in this kind of regression-discontinuity design should be evaluated on a case-by-case basis, using various quantitative and qualitative tools discussed in later chapters. In the context of the US House elections studied by Lee (2008), Caughey and Sekhon (2011) develop a compelling critique of the assertion of as-if randomness; in other contexts, this assertion may be more convincing.

For instance, Titiunik (2009) studies the incumbency advantage of political parties in Brazil's municipal mayor elections, comparing municipalities where a party barely lost the 2000 mayor elections to municipalities where it barely won. Contrary to findings in the US, she finds evidence of a strong *negative* effect of incumbency on both the vote share and the probability of winning in the following election; the significant estimated effect sizes range from around negative 4 percentage points of the vote share for the Liberal Front Party) to around negative 19 percentage points for the Party of the Brazilian Democratic Movement.

In a study of the relationship between political incumbency and media access, Boas and Hidalgo (2011) show that near-winners of city council elections are much more likely than near-losers to have their applications for community radio licenses approved by the federal government, a finding that reinforces previous research on the political control of the media in Brazil. Brollo and Nannicini (2010) use a related regression-discontinuity design to study the effect of partisan affiliation on federal transfers to municipalities in Brazil, comparing winners and losers of close elections and stratifying on whether the winner is a member of the president's coalition.

Beyond incumbency advantage for candidates, such designs have been used to study the effects for citizens and firms of supporting winning candidates. Boas, Hidalgo, and Richardson (2011), for example, study the effect of

campaign contributions on government contracts received by donors. These authors compare the returns to donations to near-winners and near-losers of campaigns, showing that public-works companies that rely on government contracts may receive a substantial monetary return on electoral investments. The effect size is striking: firms who specialize in public works projects can expect a substantial boost in government contracts—at least 8.5 times the value of their contributions—when a recipient of campaign donations from the ruling Workers' Party wins office. Golden and Picci (2011) adapt the near-winner and near-loser framework to the study of within-party effects of incumbency advantage in Italy's open-list proportional representation system, using a regression-discontinuity design to look at the relationship between the distribution of federal spending and the reelection prospects of incumbents. Finally, Eggers and Hainmueller (2009) use a regression-discontinuity approach based on near-winners and near-losers to estimate the returns to office in the British Parliament, finding that Conservative (but not Labour) MPs profited handsomely from office through outside employment they acquired due to their political positions.

Notice that regression-discontinuity designs in which analysts compare near-winners and near-losers of close elections differ in some respects from the classic design proposed by Thistlethwaite and Campbell (1960). For example, here there is not necessarily a single value of a pre-treatment covariate (or index) that determines treatment assignment. In plurality elections with just two candidates, the difference in the votes received by the two candidates can be thought of as the assignment variable—with a value of 0 being the key cutoff, positive values indicating assignment to treatment, and negative values indicating assignment to control. With more than two candidates, the difference in votes between the first and second vote-getters can be defined as the assignment covariate (also known as the "forcing variable"). Yet the adaptation of regression-discontinuity designs to other electoral systems, such as list systems with proportional representation, may imply that the margin between near-winners and near-losers is not zero at the assignment threshold, nor is it the same in every election. (For instance, the margin between the lowest vote-getter to enter the legislature on a party's list and the highest vote-getter not to enter the legislature need not be zero).[13] Thus, graphs such as Figure 2.1 showing outcomes as a function of the assignment covariate may not be as easy to draw. Yet the estimation of the effect of treatment assignment follows a similar logic as in other regression-discontinuity designs.

[13] See unpublished work by Luis Schiumerini as well as Olle Folke, and Golden and Picci (2011).

Whether electoral office is really assigned as-if at random in the neighborhood of the winning margin, even in very close elections, may be debatable in some contexts (Caughey and Sekhon 2011). The claim of as-if randomness, which motivates the analysis of the data from such natural experiments, should be carefully evaluated on a case-by-case basis. I will return later in the book to the type of a priori reasoning and empirical evidence that may be useful for such regression-discontinuity designs.[14]

3.2.3 Age as a regression discontinuity

Does past voting affect subsequent turnout or partisanship among voters? Many theories of political participation suggest that the act of voting itself has important consequences for political attitudes and behavior. Evaluating such causal claims empirically might seem rather hopeless at first glance, however, since people who choose to vote—for example, when they are first eligible to do so—probably differ in many difficult-to-measure ways from those who do not. For example, they may feel greater urge to heed civic duty, which may affect both voting and other political attitudes. Thus, initial differences in affect or behavior might then be responsible for subsequent differences in turnout rates or partisan identification.

Meredith (2009), however, takes advantage of the discontinuities imposed by voting-age restrictions to identify the effect of past participation on subsequent turnout and partisan identification. Thus, this paper compares individuals who turned eighteen just before US presidential elections—and were thus eligible to vote in those elections—with those who turned eighteen just after. Those twenty-two years of age with the opportunity to cast a ballot in a given presidential election are thus divided, as-if at random, into two groups: those who were eligible to vote in the previous election and those who were ineligible to do so, on the basis of birthdays that could be separated by as little as a few days or months. Because precise date of birth should not be related to confounding factors that affect turnout decisions or partisan identification, this regression-discontinuity design appears to provide a valid natural experiment for studying the effects of past eligibility on subsequent behavior.[15] The paper finds that past presidential election eligibility increases the probability that a voter will turn out in the next election by

[14] See Chapters 7 and 8.

[15] It may also provide an opportunity to evaluate the effects of past *participation*—rather than simply past eligibility to participate—on subsequent partisanship and voting behavior. See the discussion of instrumental-variables analysis in Chapters 4 and 5.

between 3 and 7 percentage points, depending on the election; moreover, these effects continue to persist for several election cycles after a voter first becomes eligible. Participation in past presidential elections also appears to affect partisan identification in subsequent elections.

Other kinds of age-based cutoffs have or could be used to construct regression-discontinuity designs. For example, individuals above or below a cutoff age often pay lower prices for access to libraries, museums, and other public resources (e.g., through senior citizen discounts or discounts for children under some age limit). Similarly, eligibility for medical services through Medicare (the US public health-care program) is restricted by age, which can be used to study the effects of access to these services, by comparing those who are just eligible on the basis of age to those who are just ineligible—but are otherwise indistinguishable from the treatment group (Card, Dobkin, and Maestas 2009). Again, the range of potential applications of age-based cutoffs to the study of the political and economic effects of access to various public services appears to be very large.

3.2.4 Indices

In many examples surveyed above, the key cutoff in each regression-discontinuity design was based on the value of a single pre-treatment covariate—such as a student's score on a qualifying exam or the population of a municipality. (These are called "pre-treatment" covariates because their value is determined before the intervention of interest takes place.) While this is perhaps the most common scenario for regression-discontinuity designs, it also sometimes occurs that policy-makers or other actors combine information from several variables into an index—and the value of the index then determines program assignment, with units scoring above a threshold score on the index being assigned to treatment. For example, Manacorda, Miguel, and Vigorito (2009) use a discontinuity in program assignment based on a pre-treatment eligibility score to study the effects of cash transfers in Uruguay on support for the incumbent Frente Amplio government; the eligibility score is constructed from a range of different variables, however. They find that program beneficiaries are much more likely than nonbeneficiaries to support the incumbent, by around 11 to 14 percentage points. Similarly, Lerman (2008) exploits an index used in the California prison system to assign convicts to higher- and lower-security prisons to study the effect of high-security incarceration, finding that assignment to high-security prisons has important consequences for social attitudes and behaviors.

3.3 Variations on regression-discontinuity designs

The basic regression-discontinuity design described in Section 3.2 has a number of extensions and modifications. Perhaps the most important is the distinction between regression-discontinuity designs in which treatment receipt is a deterministic function of placement above and below the threshold—as in the original Thistlethwaite and Campbell (1960) design, where every student scoring above the key threshold received a Certificate of Merit—and those in which placement relative to the threshold influences treatment receipt but does not determine it completely. Some regression-discontinuity designs also have an element of true randomization. I mention these topics only briefly below and return to them at greater length in subsequent chapters.

3.3.1 Sharp versus fuzzy regression discontinuities

I have presented the designs above as if the key regression-discontinuity threshold must determine program participation deterministically: e.g., a policy intervention is allocated to some geographic areas but not to others according to population size—and a rigid cutoff is adopted, so that units above the threshold all participate in the intervention while those below the threshold do not. While this description aptly characterizes so-called "sharp" regression-discontinuity designs (Campbell 1969; Trochim 1984), this is not necessary for successful regression-discontinuity analysis. In some studies, the key threshold merely influences program participation in a probabilistic, rather than deterministic, fashion.[16] That is, placement just above or just below the key threshold influences treatment receipt, but does not determine it completely.

Regression-discontinuity designs in which placement relative to the threshold influences treatment receipt, but does not determine it completely, are known as "fuzzy" regression-discontinuity designs (Campbell 1969; Trochim 1984). In a study by Gerber, Kessler, and Meredith (2011), for instance, a political organization in Kansas used a rule to select households to receive mailings critical of the political incumbent. Because of this rule, individuals living in census tracts in which a certain percentage of residents had

[16] In this case, as I discuss in Chapter 5, program assignment acts as an instrumental variable for program participation. Thus, there is an intimate connection between fuzzy regression-discontinuity and instrumental-variables designs.

household incomes above a given threshold were substantially more likely to be assigned to receive the mailings. Yet, the threshold did not deterministically influence treatment receipt.

Such "fuzzy" regression-discontinuity designs are closely linked to instrumental-variables designs, introduced in the next chapter: as in instrumental-variables designs, random assignment to treatment does not completely determine receipt of treatment. The analysis of fuzzy regression discontinuities thus raises issues that are similar to instrumental-variables analysis. I will therefore delay further discussion of this design for subsequent chapters.

3.3.2 Randomized regression-discontinuity designs

Some regression-discontinuity designs also incorporate elements of true randomization at the key regression-discontinuity threshold. For instance, thresholds may be relatively coarse, so that a broad set of (possibly quite heterogeneous) units are eligible for assignment to a policy on the basis of a score near the key threshold. However, actual assignment may be based on a true lottery.

Dunning and Nilekani (2010) provide one example from India.[17] In village council elections in the state of Karnataka, electoral laws require some council presidencies to be reserved for lower castes and tribes, i.e., in some councils the president must come from designated castes and tribes. The quotas are assigned by ranking councils within subdistricts in descending order, on the basis of the number of council *seats* that are reserved for lower castes and tribes.[18] Thus, quotas go first to the block of councils at the top of the list. (The total number of presidencies that must be reserved in a given election is determined by the proportion of the subdistrict population from lower castes and tribes.) In subsequent elections, quotas rotate down to the next block of councils on the list.[19] Thus, in any given election, councils with particular threshold numbers of seats reserved for lower castes and tribes are "eligible" for reservation of the council presidency. As in other regression-discontinuity

[17] See also Dunning (2010b).

[18] The number of council seats reserved for lower castes and tribes is in turn a proxy for the population proportion of lower castes and tribes, within each council constituency. What I call subdistricts here are administrative jurisdictions called "blocks."

[19] This system of rotation began in 1994, in the first elections after the passage of a national constitutional amendment requiring such electoral quotas for marginalized castes and tribes as well as women.

designs, this threshold determines treatment assignment, since councils below the threshold do not receive quotas while those above the threshold may receive quotas (though some do not, e.g., if they had quotas in previous elections; so this is also a fuzzy regression-discontinuity design).

The element of true randomization arises because, whenever the number of councils that are eligible for quotas (on the basis of having the threshold number of seats reserved for lower-castes or lower-tribe) exceeds the number of councils to which quotas must be assigned, councils are chosen at random for quotas from among the eligible group, using a lottery. In other words, at the regression-discontinuity threshold, there is then true randomization of treatment assignment. Comparisons among councils in the eligible group—including those in which the presidency was set aside for members of lower castes and tribes and those in which the presidency was not thus reserved—can then allow valid estimation of the effects of the electoral quotas.

3.3.3 Multiple thresholds

In some regression-discontinuity designs, the particular threshold that determines treatment assignment may vary across jurisdictional areas; or different thresholds may be used for different populations, or at different points in time. In Hidalgo's (2010) study of electronic voting in Brazil, for instance, the main threshold of interest is the one located at a municipal population of 40,500; in 1998, municipalities with populations greater than this threshold switched to electronic voting. However, an earlier reform in 1996 had introduced electronic voting in municipalities with populations greater than 200,000. Thus, the causal effects of electronic voting can be estimated at this threshold for the elections held in that year. In the studies by Dunning (2011) and Dunning and Nilekani (2010) discussed above, different Scheduled Caste population thresholds are used in each subdistrict in a given election, depending on the size of the subdistrict and the overall proportion of Scheduled Caste residents. Thus, the effect of caste-based quotas is estimated at different thresholds of the assignment covariate across different subdistricts.

As discussed later in the book, one of the potential advantages of regression-discontinuity designs with multiple thresholds is that causal effects at different thresholds may be explicitly compared. In general, one limitation of many regression-discontinuity designs is that the causal effect of treatment for units located near the threshold may differ from those located far from the threshold; and regression-discontinuity designs are properly used to estimate causal effects for the former population. However, with multiple thresholds

this question of external validity can, at least to some extent, be investigated empirically (see Chapter 10).

3.4 Conclusion

The recent growth of regression-discontinuity designs in the social sciences is impressive. In many contexts, the case for as-if random assignment at the threshold has strong a priori plausibility, making this a particularly useful kind of natural experiment for drawing causal inferences. Of course, as in other natural experiments, analysts are challenged to validate this assertion to the extent possible, and as-if random may be more compelling in some regression-discontinuity designs than in others.[20]

Discovering regression-discontinuity designs, as with other kinds of natural experiments, is an imprecise affair, subject to luck and inspiration. Yet, as the survey in this chapter suggests, scholars have adapted ideas developed in other contexts to study disparate substantive problems. For example, similar population-based thresholds have often provided the basis for different regression-discontinuity designs; indeed, sometimes the same threshold is used for different purposes or to study different outcomes. Comparisons of near-winners and near-losers of close elections have also been an increasingly prominent strategy for drawing inferences about the effects of a range of variables, from political incumbency to campaign donations.

Regression-discontinuity designs can have several limitations, however, and their use has inspired controversies related to both analysis and interpretation of effects. For example, there is a lively debate about strategies for analyzing data from regression-discontinuity designs: analysts have debated the virtues of simple difference-of-means tests in this context, and there are also important issues related to defining the size of the study group—essentially, how big the "window" should be around the key threshold. Such choices both depend on, and have key implications for, the claim that assignment to treatment is as good as random in a regression-discontinuity design. I return to these issues in Chapter 5.

As for interpretation, in regression-discontinuity designs the average causal effect is defined only for units near the threshold—that is, those units in the regression-discontinuity study group. Following Imbens and Angrist (1994), this parameter is sometimes called a "local average treatment effect" to

[20] See Chapter 8.