Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 142 (2008) 675–697
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Randomized experiments from non-random selection in U.S. House elections ☆

David S. Lee*

Department of Economics and Woodrow Wilson School, Princeton University and NBER, Industrial Relations Section, Firestone Library, Princeton, NJ 08544, USA

Available online 21 May 2007

Abstract

This paper establishes the relatively weak conditions under which causal inferences from a regression–discontinuity (RD) analysis can be as credible as those from a randomized experiment, and hence under which the validity of the RD design can be tested by examining whether or not there is a discontinuity in any pre-determined (or "baseline") variables at the RD threshold. Specifically, consider a standard treatment evaluation problem in which treatment is assigned to an individual if and only if $V > v_0$, but where v_0 is a known threshold, and V is observable. V can depend on the individual's characteristics and choices, but there is also a random chance element: for each individual, there exists a well-defined probability distribution for V. The density function—allowed to differ arbitrarily across the population—is assumed to be continuous. It is formally established that treatment status here is as good as randomized in a local neighborhood of $V = v_0$. These ideas are illustrated in an analysis of U.S. House elections, where the inherent uncertainty in the final vote count is plausible, which would imply that the party that wins is essentially randomized among elections decided by a narrow margin. The evidence is consistent with this prediction, which is then used to generate "near-experimental" causal estimates of the electoral advantage to incumbency.

JEL classification: C21; C30; C90; D72

Keywords: Regression discontinuity; Randomized experiments; Average treatment effects; Elections; Incumbency

1. Introduction

There is a recent renewed interest in the identification issues involved in (Hahn et al., 2001), the estimation of (Porter, 2003), and the application of (Angrist and Lavy, 1998; van der Klaauw, 2002) Thistlethwaite and Campbell's (1960) regression–discontinuity design (RDD). RDDs involve a dichotomous treatment that is a deterministic function of an single, observed, continuous covariate (henceforth, "score"). Treatment is assigned to those individuals whose score crosses a known threshold. Hahn et al. (2001) formally establish

☆ An earlier draft of this paper, "The electoral advantage to incumbency and voters' valuation of politicians' experience: a regression discontinuity analysis of elections to the U.S. House", is available online as NBER working paper 8441.

*Tel.: +1 609 258 9548; fax: +1 609 258 4041.
E-mail address: davidlee@princeton.edu

0304-4076/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2007.05.004

minimal continuity assumptions for identifying treatment effects in the RDD: essentially, the average outcome for individuals marginally below the threshold must represent a valid counterfactual for the treated group just above the threshold. For the applied researcher, there are two limitations to invoking this assumption: (1) in many contexts, individuals have some influence over their score, in which case it is unclear whether or not such an assumption is plausible, and (2) it is a fundamentally untestable assumption.

This paper describes a very general treatment assignment selection model that (1) allows individuals to influence their own score in a very unrestrictive way, and (2) generates strong testable predictions that can be used to assess the validity of the RDD. In particular, it is shown below that causal inferences from RD designs can sometimes be as credible as those drawn from a randomized experiment.

Consider the following general mechanism for treatment assignment. Each individual is assigned a score V, which is influenced partially by (1) the individual's attributes and actions, and (2) by random chance. Suppose that conditional on the individual's choices and characteristics, the probability density of V is continuous. Treatment is given to the individual if and only if V is greater than a known threshold v_0. Note that there is unrestricted heterogeneity in the density function for V across individuals, so that each individual will in general have a different (and unobserved to the analyst) probability of treatment assignment.

Below it is formally established that this mechanism not only satisfies the minimal assumptions for RDDs outlined in Hahn et al. (2001); it additionally generates variation in treatment status that is as good as randomized by an experiment—in a neighborhood of $V = v_0$. Close to this threshold, *all variables* determined prior to assignment will be independent of treatment status. Thus—as in a randomized experiment—differences in post-assignment outcomes will not be confounded by omitted variables, whether observable or unobservable.

This alternative formulation of a valid RDD and the local independence result are useful for three different reasons. First, it illustrates that natural randomized experiments can be isolated even when treatment status is driven by non-random self-selection. For example, the vote share V obtained by a political candidate could be dependent on her political experience and campaigning effort, so that on average, those who receive the treatment of winning the election ($V > \frac{1}{2}$) are systematically more experienced and more ambitious. Even in this situation, provided that there is a random chance error component to V that has continuous pdf, treatment status in a neighborhood of $V = \frac{1}{2}$ is statistically randomized.

Second, in any given applied context, it is arguably easy to judge whether or not the key condition (continuous density of V for each individual) holds. This is because the condition is directly related to individuals' incentives and ability to sort around the threshold v_0. As discussed below, if individuals have exact control over their own value of V, the density for each individual is likely to be discontinuous. When this is the case, the RDD is likely to yield biased impact estimates.

Finally, and perhaps most importantly, the local independence result implies a strong empirical test of the internal validity of the RDD. In a neighborhood of v_0, treated and control groups should possess the same distribution of baseline characteristics. The applied researcher can therefore verify—as in a randomized controlled trial—whether or not the randomization "worked", by examining whether there are treatment-control differences in baseline covariates.[1] These specification tests are not based on additional assumptions; rather, they are auxiliary predictions—consequences of the assignment mechanism described above. The local random assignment result also gives a theoretical justification for expecting impact estimates to be insensitive to the inclusion of any combination of baseline covariates in the analysis.[2]

The result is applied to an analysis of the incumbency advantage in elections to the United States House of Representatives. It is plausible that the exact vote count in large elections, while influenced by political actors in a non-random way, is also partially determined by chance beyond any actor's control. Even on the day of an election, there is inherent uncertainty about the precise and final vote count. In light of this uncertainty, the local independence result predicts that the districts where a party's candidate just *barely* won an election—and

[1]Such specification checks have been used recently, for example, in Lee et al. (2004), Linden (2004), Martorell (2004), Clark (2004), Matsudaira (2004), DiNardo and Lee (2004).

[2]Hahn et al. (2001) do state that the "advantage of the method is that it bypasses many of the questions concerning model specification: both the question of which variables to include in the model for outcomes," but provide no justification for why the treatment effect estimates should be insensitive to the inclusion of baseline characteristics.