Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 142 (2008) 611–614
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Editorial

Special issue editors' introduction: The regression discontinuity design—Theory and applications

Regression discontinuity (RD) designs for evaluating causal effects of interventions where assignment to a treatment is determined at least partly by the value of an observed covariate lying on either side of a cutoff point were first introduced by Thistlewaite and Campbell (1960). With the exception of a few unpublished theoretical papers (Goldberger, 1972a, b), these methods did not attract much attention in the economics literature until recently. Starting in the late 1990s, there has been a growing number of studies in economics applying and extending RD methods, starting with Van der Klaauw (2002), Lee (2007), Angrist and Lavy (1999), and Black (1999). Around the same time, key theoretical and conceptual contributions, including the interpretation of estimates for fuzzy RD designs and allowing for general heterogeneity of treatment effects, were developed by Hahn et al. (2001).

The time appeared right for a conference focusing on these methods and most papers in this volume are the result of such a conference, organized by David Card and Thomas Lemieux at the Banff International Research Station (BIRS) in Banff, Canada, in the Spring of 2003. We first set the stage for the volume by reviewing some of the practical issues in implementation of RD methods. There is relatively little novel in this discussion, but it addresses some of the practical issues in implementing RD designs, including the new theoretical developments. Given the very recent nature of this renaissance of RD methods in economics, such reviews have so far been largely absent from the economic literature (exceptions include Van der Klaauw, 2008 and the discussion in Angrist and Krueger, 1999).

The volume then starts with an interesting history of the RD design in psychology, education, statistics, and economics by Thomas Cook. In his paper, Cook points out that although the RD method has been around for almost 50 years, applications have been sparse in psychology and education where it was first introduced by Thistlewaite and Campbell (1960), and only drew limited interest from statisticians. This is not unlike the situation that prevailed until recently in economics where the RD design was first independently rediscovered by Goldberger (1972a, b) as an aside to his work on selection based on observables (see also Cain, 1975), but then failed to attract further attention from econometricians and applied economists. Cook shows that the recent and growing interest for the RD design in economics is quite unique and that "this augurs well for new and rich life being breathed into RD design, especially if most of the design applications now under way turn out to be clear in their substantive findings and capable of generating new problems." Our hope is that the mix of methodological and applied papers contained in the volume will indeed help establish the RD design as an essential element in the "toolkit" of economist and other social scientists.

Most of the theoretical foundations of the RD design discuss the case where the probability of receiving the treatment is a discontinuous function of a continuous treatment-determining variable. This makes it possible, in principle, to draw causal inference by comparing outcomes for observations just above and just below the treatment threshold. In many applied settings, however, the treatment-determining covariate is discrete and the identification at the limit argument used in the continuous case does not hold. The paper by David Lee and David Card points out that, because of this limitation, researchers need to choose a functional form for the relationship between the treatment-determining variable and the outcomes of interest when the treatment-determining variable is discrete. Estimated regression models on both sides of the cutoff point have to be used to extrapolate what the difference between outcomes would be at that point.

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2007.05.008

An interesting feature of the discrete case is that it provides a natural way of addressing possible specification errors in the functional form of the regression function. Lee and Card show that these specification errors can be interpreted as introducing a group structure in the standard errors, so that conventional standard errors overstate the precision of the estimated program impacts. They show that standard clustering adjustments yield consistent standard errors when specification errors with or without treatment are similar.

David Lee's paper establishes that causal inferences from a RD analysis can be as credible as those from a randomized experiment provided that the treatment-determining covariate is partly determined by a random chance element. This means that even if individuals are able to partly manipulate the treatment-determining covariate, they do not have perfect control over it. When this condition holds, it follows that the conditional expectation of the outcome variable (under treatment and non-treatment) is continuous in the treatment-determining variable, which is precisely the condition under which the RD design is ‘‘as good’’ as a randomized experimental design (Hahn et al., 2001). Just as in a randomized experiment, the validity of the RD design can then be tested by examining whether or not there is a discontinuity in any pre-determined (or ‘‘baseline’’) variables at the cutoff point. 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.

When the treatment-determining covariate is partly determined by a random chance element, as in the case considered by Lee, it follows that the density of this variable is continuous at the point of discontinuity. In contrast, when individuals are able to perfectly manipulate the treatment-determining variable, its density is likely to be discontinuous. For example, if students with a GPA of 85 percent or more receive a generous scholarship and are able to perfectly ‘‘manipulate’’ their grades, we should expect to see an abnormal concentration of students with a GPA of exactly 85, but very few students with a GPA just below 85. This kind of manipulation of the treatment-determining variable will typically invalidate the RD design, but it can be detected by checking whether the density of the variable is discontinuous at the cutoff point. Building on this idea, the paper by Justin McCrary develops a test for manipulation related to continuity of the density of the treatment-determining variable density function at the cutoff point. The test is implemented by first estimating the density over a number of small bins. As in other applications of the RD design, McCrary then suggests to test for the presence of a discontinuity (in the estimated densities) using local linear regressions. The methodology is applied to popular elections to the House of Representatives, where manipulation is neither expected nor found, and to roll-call voting in the House, where manipulation is both expected and found. The density discontinuity test suggested by McCrary is a useful and unique (to the RD design) complement to standard tests that baseline variables are balanced around the discontinuity point, just as in a standard randomized experiment.

In many applications of the RD design, treatment is not purely assigned on the basis of a unique treatment-determining variable. This leads to the important distinction between a sharp RD and a ‘‘fuzzy’’ RD design discussed in detail in our guide to practice. Although the sharp RD design is attractive because it can be ‘‘as good’’ as a randomized experiment, it relies on the often unrealistic assumption that participation into a program is purely determined on the basis of a single treatment-determining variable crossing a specific threshold. Another well-known limitation of the sharp RD design is that it only allows for the identification of a local average treatment effect for individuals around the cutoff point in the presence of heterogeneous effects. To overcome some of the limitations of the sharp RD design, the paper by Erich Battistin and Enrico Rettore looks at a special case of the fuzzy RD design where some individuals who are not treated are eligible for the treatment, while others are not. Under this setup, Battistin and Rettore show that when individuals self-select into participation conditional on some eligibility criterion, a sharp RD design provides a natural framework to define a specification test for the non-experimental estimation of programme effects for all participants. They also show that, in this setup, the regularity conditions required for the identification of counterfactual outcomes for participants marginally eligible for a program are essentially the same as in a sharp RD design. These results are illustrated empirically in the case of the PROGRESA program in Mexico.

The rest of the volume presents a number of innovative applications of the RD design in a variety of settings. Wilbert van der Klaauw's paper provides an evaluation of the impact of Title I funding of compensatory education programs on school finance and student performance in New York City public schools during the 1993, 1997, and 2001 school years. One challenge in assessing the impact of these compensatory education programs is that they are typically directed at schools where students outcomes are below average to start with. This tends to generate a spurious negative correlation between Title I funding and students outcomes. The paper exploits a discontinuity in the rule that determines Title I eligibility, where schools with poverty counts above the district average are eligible for Title I funds, while most other schools are not. So although all schools near the average have comparable poverty counts, their Title I status differs, which provides the basis of an RD evaluation of the impact of Title I on school performance. The RD estimates indicate that the program was unsuccessful in improving student outcomes in high-poverty schools in New York City during this period, and may in fact have had adverse effects during the earlier years in the sample. Less evidence of a negative effect is found for the 2001 school year. Van der Klaauw concludes that these findings are related to the way in which the federal funds were spent.

Susan Chen and Wilbert van der Klaauw's paper evaluates the work disincentive effects of the Disability Insurance (DI) program during the 1990s using standard comparison group and RD methods. The earlier literature on the topic has relied on comparisons between individuals on DI and individuals in a non-experimental comparison group to estimate the disincentive effects of the program. Chen and van der Klaauw instead exploit the fact that the eligibility determination process is based in part on the age of individuals. They exploit this age discontinuity to estimate the impact of the program on labor supply for an important subset of DI applicants. Using merged survey-administrative data, they find that, during the 1990s, the labor force participation rate of DI beneficiaries would have been at most 20 percentage points higher had none received benefits. In addition, they find an even smaller labor supply response for the subset of "marginal" applicants whose disability determination is based on vocational factors.

Rafael Lalive's paper studies the impact of a program that extended the maximum duration of unemployment benefits from 30 to 209 weeks in Austria during the 1980s. Interestingly, the program was targeted to individuals age 50 or older, living in certain eligible regions of Austria. Lalive exploits the sharp discontinuities in treatment assignment at age 50 and at the border between eligible and control regions to identify the effect of extended benefits on unemployment duration. His RD results indicate that the duration of job search is increased by at least .09 weeks per additional week of benefits among men, whereas unemployment duration increases by at least .32 weeks per additional week of benefits among women. Lalive also discusses how the differences in responses to the program between men and women are consistent with women facing a lower minimum age for early retirement than men, so that they can use extended unemployment benefits as a "bridge" to early retirement.

Thomas Lemieux and Kevin Milligan also use a RD design to look at the adverse impact of income support programs on labor supply behavior. They study the case of the Canadian Province of Quebec where, prior to 1989, income assistance recipients under the age of 30 with no dependent children received much lower benefits than recipients age 30 or more. This sharp discontinuity in policy is used to estimate the effects of social assistance on various labor market outcomes using a RD approach. Lemieux and Milligan find strong evidence that more generous social assistance benefits reduce employment. The estimates exhibit little sensitivity to the degree of flexibility in the specification, and perform very well when controlling for unobserved heterogeneity using a first difference specification. In contrast, they show that commonly used difference-in-differences estimators may perform poorly with inappropriately chosen control groups.

Jordan Matsudaira's paper is motivated by the fact that nearly every school district in the United States' largest cities has implemented a mandatory summer school program for students failing to meet achievement standards on year-end tests. Previous work on the effectiveness of summer school has typically focused on voluntary programs, which is problematic from an evaluation point of view as there may be systematic differences between students who do and do not attend summer school. Using a large administrative data set, Matsudaira compares students who score just below a preset proficiency level on year-end exams, and are thus mandated to attend summer school, to students who score just above the cutoff. Students above and below the cutoff are shown to be nearly identical in terms of all other factors that might affect achievement, suggesting that the RD design is as good as a randomized experiment in this specific case. Overall, the paper finds an average effect of summer school attendance of about .12 standard deviations for both math and reading, which is at the low end of the range of prior estimates. These averages mask considerable heterogeneity, however, with effect size estimates ranging from (just below) zero to about one-quarter of a standard deviation. On the basis of these estimates, Matsudaira concludes that summer school may be a more cost-effective way of raising student achievement scores than class-size reductions.

References

Angrist, J., Lavy, V., 1999. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. Quarterly Journal of Economics 114, 533–575.

Angrist, J.D., Krueger, A.B., 1999. Empirical strategies in labor economics. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, vol. 3. North-Holland, Amsterdam, pp. 1277–1366.

Black, S., 1999. Do better schools matter? parental valuation of elementary education. Quarterly Journal of Economics 114, 577–599.

Cain, G.C., 1975. Regression and selection models to improve nonexperimental comparisons. In: Bennet, C.A., Lumsdaine, A.A. (Eds.), Evaluation and Experiment. Academic Press, New York, pp. 297–317.

Goldberger, A., 1972a. Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations. Madison, WI, Unpublished manuscript.

Goldberger, A., 1972b. Selection Bias in Evaluating Treatment Effects: The Case of Interaction. Madison, WI, Unpublished manuscript.

Hahn, J., Todd, P., van der Klaauw, W., 2001. Identification and estimation of treatment effects with a regression discontinuity design. Econometrica 69, 201–209.

Lee, D.S., 2007. Randomized experiments from non-random selection in U.S. house elections. Journal of Econometrics, this issue, doi:10.1016/j.econom.2007.05.004.

Thistlewaite, D., Campbell, D., 1960. Regression-discontinuity analysis: an alternative to the ex-post facto experiment. Journal of Educational Psychology 51, 309–317.

Van Der Klaauw, W., 2002. Estimating the effect of financial aid offers on college enrollment: a regression-discontinuity approach. International Economic Review 43, 1249–1287.

Van Der Klaauw, W., 2008. Regression-Discontinuity Analysis. The New Palgrave Dictionary of Economics, Palgrave MacMillan, forthcoming.

Guido Imbens
Department of Economics, Harvard University,
M-24 Littauer Center, Cambridge, MA 02138, USA and NBER
E-mail address: imbens@harvard.edu

Thomas Lemieux*
Department of Economics, University of British Columbia,
997-1873 East Mall, Vancouver, BC,
Canada V6T 1Z1 and NBER
E-mail address: tlemieux@interchange.ubc.ca

*Corresponding author. Tel.: +1 604 822 2092; fax: +1 604 822 5915.