Journal of Econometrics 189 (2015) 132–147

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Regression discontinuity designs with unknown discontinuity points: Testing and estimation[☆]

Jack Porter[a], Ping Yu[b,*]

[a] Department of Economics, University of Wisconsin, 6448 Social Science Building, 1180 Observatory Drive, Madison, WI 53706-1393, USA

[b] School of Economics and Finance, The University of Hong Kong, Pokfulam Road, Hong Kong

ARTICLE INFO

Article history:
Received 19 March 2014
Received in revised form
23 November 2014
Accepted 11 June 2015
Available online 20 July 2015

JEL classification:
C12
C13
C14
C21

Keywords:
Regression discontinuity design
Unknown discontinuity point
Specification testing
Nonparametric structural change
Wild bootstrap
Degenerate U-statistic
Difference kernel estimator
Cross validation

ABSTRACT

The regression discontinuity design has become a common framework among applied economists for measuring treatment effects. A key restriction of the existing literature is the assumption that the discontinuity point is known, which does not always hold in practice. This paper extends the applicability of the regression discontinuity design by allowing for an unknown discontinuity point. First, we construct a unified test statistic to check whether there are selection or treatment effects. Our tests are shown to be consistent, and local powers are derived. Also, a bootstrap method is proposed to obtain critical values. Second, we estimate the treatment effect by first estimating the nuisance discontinuity point. It is shown that estimating the discontinuity point does not affect the efficiency of the treatment effect estimator. Simulation studies illustrate the usefulness of our procedures in finite samples.

1. Introduction

Since its invention by Thistlethwaite and Campbell (1960), the regression discontinuity design (RDD) has attracted much attention among econometricians; see Imbens and Lemieux (2008), van der Klaauw (2008) and Lee and Lemieux (2010) for excellent reviews on up-to-date theoretical developments and applications and Yu (2013) for a summary of treatment effects estimators in RDDs. In RDDs, an observable covariate is used to completely determine the treatment status, and is called the *forcing* (*running* or *assignment*) variable. When the value of the covariate for an individual is above a threshold or discontinuity point, the individual will be treated; otherwise, the individual will be put in the control group. Usually, the discontinuity point is set by the policy maker and is publicly known. However, such information is not always available in practice. Sometimes, the discontinuity point is only known to the policy maker but is unknown to the public (including econometricians) due to ethical reasons or privacy. In the classical application of RDDs in the effect of the scholarship offers on student enrollment decisions by van der Klaauw (2002), the forcing variable is an underlying index of various individual characteristics. To avoid manipulation by individuals or competition from other schools, the discontinuity point may not be disclosed. Another example with an unknown discontinuity point is Card et al. (2008) who analyze the tipping effect in the dynamic of segregation. Specifically, when the minority share in a neighborhood exceeds a "tipping point", all the whites leave. Such a tipping point depends on the strength of white distaste for minority neighbors and is generally unknown.

To date, all existing literature on RDDs assumes that the discontinuity point is known, especially to econometricians. This paper studies the testing and estimation problem when the discontinu-

[☆] Thanks go to the co-editor Jianqing Fan, one associate editor, two referees and the seminar participants at Auckland, Northwestern, and ESAM 2011 for helpful comments. The second author wants to thank FRDF at the University of Auckland for financial supports.

* Corresponding author.
E-mail addresses: jrporter@ssc.wisc.edu (J. Porter), pingyu@hku.hk (P. Yu).

http://dx.doi.org/10.1016/j.jeconom.2015.06.002
0304-4076/© 2015 Elsevier B.V. All rights reserved.

ity point is unknown. In testing, we try to check whether there are selection or treatment effects in our experiment. The first test is to check whether there is selection among individuals. This test is new in the literature. The second test is to check the presence of treatment effects. This test is very close to the nonparametric structural change test, and there have been at least three tests designed for this purpose in the statistical literature. Our test is inspired by the nonparametric specification testing literature started from Bierens (1982), and is novel in our context. We solve both testing problems by a unified test statistic, but adapt it to different problems by varying a smoothing parameter. The test statistic is constructed under the null, so is similar to the score test in spirit. In estimation, our main interest lies in the treatment effects evaluation, but a primary input, the discontinuity point, is unknown. So we estimate the discontinuity point first by an estimator called the *difference kernel estimator* (DKE). We show its superconsistency and find its asymptotic distribution, and then estimate the treatment effect as if the discontinuity point were known. It is surprising that estimation of the discontinuity point does not affect the efficiency of the treatment effect estimator asymptotically. The model we consider has many applications beyond RDDs; see Müller (1992), Wu and Chu (1993a), Wang (1995), Müller and Stadtmüller (1999), and the reference therein for applications in statistics.

This paper is organized as follows. In Section 2, we set up our framework and specify some regularity assumptions. Especially, we clarify what selection means and what it implies to observations. Section 3 presents our specification test statistic and develops its asymptotic distributions in different testing problems. Furthermore, a bootstrap method is suggested to obtain critical values which may have better finite sample performances. Alternative tests of nonparametric structural change in the statistical literature are also reviewed and compared with our test. Section 4 considers the estimation problem. We provide estimators of the discontinuity point and the treatment effect, and develop their asymptotic distributions. In Section 5, we extend the results in Sections 3 and 4 to other settings and solve an important practical issue, the bandwidth selection, in both specification testing and estimation. Section 6 includes some simulation results and Section 7 concludes. To save space, we put some intuitions and all technical proofs in the online supplementary materials (see Appendix A).

Throughout this paper, we concentrate on the sharp design and discuss the fuzzy design only briefly in Sections 5.1 and 5.2. Such an arrangement allows us to focus on the main idea of this paper. In the sharp design, we assume that only the response variable and the forcing variable are observable, while the treatment status is not. Otherwise, the testing and estimation problem will degenerate to the case with a known discontinuity point; see Section 2 of Yu (2012) and Section 2.2 of Yu and Zhao (2013) for a detailed discussion on this point. We further concentrate on the sharp design with at most one discontinuity point; generalization to finite and unknown number of discontinuity points is only discussed briefly in Section 5.3.

A word on notation: the letter C is used as a generic positive constant, which need not be the same in each occurrence. WLOG means "without loss of generality". DGP means "data generating process". LLS means the "local linear smoother" popularized by Fan (1992, 1993) and Fan and Gijbels (1996). The symbol $\approx$ means that the higher-order terms are omitted or a constant term is omitted (depending on the context). w.p.a.1 means "with probability approaching one". For a nonnegative real s, $[s]$ is its integer part. For any two random variables x and y, $f(x)$ means the density of x and $f(y|x)$ means the conditional density of y given x. The letter π_0 represents the true discontinuity point and π represents a generic discontinuity point in the parameter space $\Pi = [\underline{\pi}, \overline{\pi}]$, where $\underline{\pi}$ and $\overline{\pi}$ are constants and $\underline{\pi} < \pi_0 < \overline{\pi}$. For a function $g(x)$, $g(x)$ has a cusp at $x = \pi$ means that $g(x)$ is continuous at π but $g'(\pi+) \neq g'(\pi-)$, that is, the left and right derivatives at π are not the same. "Discontinuity" and "jump" are used exchangeably.

2. Framework and assumptions

We first put RDDs in the usual treatment framework and discuss a key "selection" assumption. Such a framework can be treated as the structural form of RDDs. We then impose some smoothness assumptions on the usual reduced-form formulation of RDDs. Such assumptions are necessary for the development of the testing and estimation procedures in this paper. Finally, we sketch the basic ideas of our specification testing and estimation.

2.1. Selection and treatment effects

Following Lee (2008), suppose the response $y = y(x, U)$, where x is the one-dimensional forcing variable which is observable, and U is the unobservable component such as students' ability in the scholarship example of van der Klaauw (2002). We assume that there can be any correlation between U and x.[1] Also, $y(x, U)$ satisfies the following smoothness assumptions.

Assumption Y. (a) If there is no treatment or there is treatment but are no treatment effects, $y(x, U) = \underline{y}(x, U)$ is continuous in (x, U) and is **continuously** differentiable in x for each U.

(b) If there are treatment effects, $y(x, U) = \underline{y}(x, U) + \alpha(U)1(x \geq \pi)$ with $\alpha(U)$ being continuous.

Under Assumption Y, when there are no treatment effects, the response y is a **smooth** function of x after controlling for all specific characters (except x) of an individual. In the special case where $\underline{y}$ takes the additively separable form, $\underline{y} = g(x) + U$, $g(\cdot)$ is assumed to be continuously differentiable. This is understandable because it is hard to imagine y changes dramatically when x changes from $x - \Delta$ to $x + \Delta$ for a small Δ given that human beings usually behave smoothly. Even if there are treatment effects, $y(x, U)$ only changes its size at $x = \pi$, and the slopes at the left and right sides of π remain the same.

Under Assumption Y, using notations of the conventional average treatment effects literature such as Heckman and Vytlacil (2007a,b), we can express the responses of the control and treated group as follows:

$$Y_0 = \mu_0(x, U_0), \qquad Y_1 = \mu_1(x, U_1) \quad \text{and} \quad D = 1(x \geq \pi),$$

where $\mu_0(x, U_0) = \underline{y}(x, U)$ with $U_0 = U$ and $\mu_1(x, U_1) = \underline{y}(x, U) + \alpha(U)$ with $U_1 = U$. The main difference of RDDs from the conventional average treatment effects framework is that the treatment status is determined by a single **observable** x, so the treatment status can be sharply observed (or there is a discontinuity in the propensity score at π, or the unconfoundedness condition is trivially satisfied). Such an advantage is not free. The usual overlap assumption is violated because given any x, we can observe either Y_0 or Y_1 but not both. We must rely on the continuity of $\mu_0(x, U_0)$ in the left neighborhood of π to predict its behavior in the right neighborhood of π, and similarly for $\mu_1(x, U_1)$. As a result, we only use the local information around π to identify the treatment effects, which makes the treatment effects estimator achieve only a

[1] From the Skorohod representation, y can be expressed as $Q(U|x)$, where $Q(\cdot|x)$ is the conditional quantile function of y given x, and $U|x \sim U(0, 1)$. Note here that U and x can have any correlation; see Section 2.1 of Yu (2014b) for more discussions on this point. Only in quantile regression, we assume U and x are independent.