cemmap

centre for microdata methods and practice

Honest inference for discrete outcomes in regression discontinuity designs

Walter Beckert
Daniel Kaliski

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP07/24

Economic and Social Research Council

Honest Inference for Discrete Outcomes in Regression Discontinuity Designs

Walter Beckert and Daniel Kaliski*

March 3, 2024

Abstract

We investigate the consequences of discreteness in the assignment variable in regression-discontinuity designs for cases where the outcome variable is itself discrete. When the assignment variable is discrete, standard confidence intervals do not have the nominal level of coverage, but confidence intervals with the correct coverage rate can be constructed conditional on an assumed upper bound for the second derivative of the expectation of the outcome conditional on the assignment variable. We propose a novel method for estimating this bound on the second derivative in cases where the outcome variable is generated by a binary outcome threshold-crossing model. The method leverages prior results that show that key primitives that determine the second derivative are semiparametrically identified conditional on mild shape restrictions motivated by theory and a known distribution of unobserved heterogeneity. Applying our method to examine the effect of the Social Security claiming age eligibility threshold of 62 on claims in the United States, we find that the size of the spike in claims at the age of eligibility is sensitive to assumptions regarding the distribution of unobserved reservation utilities.

Keywords: Regression discontinuity design; Honest inference; Discrete choice

*Birkbeck, University of London. Corresponding author: d.kaliski@bbk.ac.uk. We are grateful to Bruno Ferman, Claudia Noack, Tomasz Olma, Christoph Rothe and seminar participants at the 2020 European Economic Association meetings, the 2021 Royal Economic Society meetings, and the 2022 International Association for Applied Econometrics meetings for helpful comments and advice.

1 Introduction

Regression discontinuity designs are a popular tool for causal inference. A subset of these designs exploits rules that assign units to treatment and control groups based on a discrete, ordered random variable, such as age in years. The object of interest in these cases is often a discrete outcome such as survival, or take-up of some intervention which is available to individuals just above the threshold for assignment to treatment, but not to their counterparts in the control group who have values of the assignment variable that are just below the same threshold value. The causal effect of the rule under investigation is then estimated as the difference in the average outcome between the treated units just above, and the control units just below, the threshold for treatment. Papers that include both of these types of discreteness include, but are not limited to: Card et al. (2008) and Card et al. (2009), who estimate the impact of Medicare on health care utilization and outcomes; Almond et al. (2010), who estimate the impact of birthweight-based cutoffs for medical interventions on the one-year mortality of infants; and Shigeoka (2014), who estimates the impact of changes in patient cost-sharing on health care utilization and outcomes among the elderly in Japan at age 70.

Recent methods for controlling the Type I error rate in these designs via "bias-aware" inference rely on researchers using *a priori* knowledge to specify bounds on features of the data-generating process (Armstrong and Kolesár, 2018, Kolesár and Rothe, 2018, Imbens and Wager, 2019, Noack and Rothe, 2021). This leaves room for guidance based on an explicit model of the data-generating process and any bounds that it implies.

In this paper, we exploit the fact that if the outcome of interest Y is generated by a binary outcome threshold-crossing model $Y = 1[g(R) \leq \eta]$, where R is a running variable used to assign units to treatment and η is unobserved heterogeneity, then under the assumptions that (i) the distribution of η is known, and $g(\cdot)$ is (ii) monotone increasing and (iii) concave in the assignment variable, we can semiparametrically identify the curvature of the conditional expectation of the binary outcome Y given the assignment variable R

from the data. This allows researchers to use information from within the same data set to generate "Honest" confidence intervals, which have at least the nominal Type I error rate, for a regression-discontinuity design with a discrete running variable. This provides a link between the literature on the non- and semiparametric identification of discrete choice models of binary response (Matzkin, 1991, 1992, Heckman et al., 2010) and the literature on inference in regression discontinuity designs with discrete running variables (Lee and Card, 2008, Kolesár and Rothe, 2018, Imbens and Wager, 2019, Noack and Rothe, 2021).

We first provide a direct link between a process that generates binary outcomes and the smoothness tuning parameter introduced to the literature by Kolesár and Rothe (2018) that must be chosen by the researcher. Part of this linkage is facilitated by proving a new result linking the scale of random variables with log-concave PDFs to the first derivative of their PDFs. In our Appendix, an empirical example demonstrates how researchers can examine the sensitivity of regression-discontinuity results with a discrete outcome to different assumed distributions for unobserved heterogeneity across individuals.

The confidence intervals that we construct and their properties originate in Kolesár and Rothe (2018), who introduce Bounded Second Derivative (BSD) confidence intervals for regression-discontinuity designs with discrete running variables. They show these confidence intervals have the correct size conditional on researchers assuming some value for K, the uniform bound on the second derivative of a function $m(\cdot)$ that is equal to the conditional expectation function at the discrete points of support of the running variable with probability one.[1] Armstrong and Kolesár (2018) moreover show that any "data-driven" method for constructing confidence intervals necessarily implicitly assumes some value for K even if one isn't explicitly chosen. In discrete outcome models, the distribution of unobserved heterogeneity and the functional forms present in the process that generates the outcomes are jointly necessary and sufficient to determine the second derivative of $m(\cdot)$. Hence the results in this paper, which show that choosing K in the context of threshold-crossing

[1] It is a separate question as to whether a continuous function $m(\cdot)$ that is only observed at non-randomly spaced discrete points $\{R_{min}, ...R_{max}\}$ can be identified. For the purpose of this paper, we also adopt this framework, and leave this question to future research.

binary outcome models is equivalent to choosing some set of distributional assumptions for the unobserved η and functional-form assumptions on $g(\cdot)$, which can be interpreted as the observable systematic component of a utility or production function. While economic theory is typically silent on the distribution of unobservables, it does provide some guidance on the shape of $g(\cdot)$ if it can be interpreted as a utility function or production function. A general review of the econometrics of shape restrictions motivated by economic theory can be found in Chetverikov et al. (2018).

We apply our method to estimating the spike in the take-up of Social Security benefits in the United States at age 62. We find that several choices of the tuning parameter required for constructing the confidence intervals in Kolesár and Rothe (2018) can be motivated under different identifying assumptions, producing wide variation in the set of possible confidence intervals for the discontinuous increase in claims. Proceeding in this way allows researchers to analyse the sensitivity of their results to economically interpretable assumptions.

The advantages of our approach are the following: first, the shape restrictions we use can be motivated from theory and/or institutional detail; second, these shape restrictions are in principle testable (though we do not recommend carrying out formal tests before implementing them, since this introduces pre-testing bias); third, our approach allows the researcher to examine the robustness of their conclusions to different distributional assumptions; fourth, the resulting confidence intervals are less conservative than the ones constructed using the R package RDHonest's default method (described in Armstrong and Kolesár (2020)), as evidenced by the simulation results. These tighter confidence intervals produced by our approach are a natural consequence of the use of stronger assumptions. They are more conservative, by contrast, than the method for bounding the second derivative recommended by Imbens and Wager (2019), but our simulations suggest that their method can produce unreliable inferences when the function that generates the discrete outcome is not concave. An open question which we leave for future research is whether a subset of the optimality class of confidence intervals identified by Armstrong and Kolesár

(2018) is optimal under our additional assumptions.

Our method illustrates another benefit of combining methods typically used in structural estimation with quasi-experimental evidence. The literature has already noted the benefits of using quasi-experiments to identify the parameters of structural models (Hastings, 2004, Petrin and Train, 2010, Kleven and Waseem, 2013, Andrews et al., 2020) and of using models of discrete choice to assess the external validity of and generalize from quasi-experiments (Todd and Wolpin, 2006, Kaboski and Townsend, 2011, Manoli and Weber, 2016, Todd and Wolpin, 2023). To our knowledge, ours is the first paper to propose that a semiparametric model of discrete choice be used to aid statistical *inference* in a quasi-experimental setting rather than external validity. The explicit reliance of the construction of Honest confidence intervals on a tuning parameter that depends on the underlying data-generating process is the feature of the method we consider that makes this possible. Armstrong and Kolesár (2018) show that "data-driven" methods for constructing valid inference in regression-discontinuity designs implicitly restrict the value of this tuning parameter (the maximum of the second derivative of the underlying conditional expectation function), so assuming this parameter exists is less restrictive than it might otherwise appear.

A recent strand of the literature takes a different approach to ours and applies randomization inference to the regression-discontinuity design (Cattaneo et al., 2015, 2017). That approach assumes that there is a neighbourhood around the cutoff in which the running variable is as good as randomly assigned, and hence the unobservables are locally constant in that neighbourhood, which, like the assumptions invoked in this paper, is significantly stronger than the usual regression discontinuity assumption of continuity of the unobservables in the running variable at the cutoff. Whether that approach is preferred to our own depends on the empirical context; the uniformity property of the confidence intervals constructed using our assumptions is suitable for when there are many observations at each discrete point of support, whereas the randomization approach is better suited to when few observations are available near the cutoff. Bias-aware inference can also deal with cases where the randomization mechanism is implausible, such as regression-discontinuity designs that use age as a discrete running variable (since individuals do not appear at ages above or below the cutoff due to any plausible randomization procedure, but due to the passage of time and survival to subsequent ages). Our paper also invokes a version of local exogeneity in order to justify estimation of the primitives of the bound on the second derivative, and shows that additional assumptions often used in the literature on discrete choice have implications that can be used to estimate that upper bound. Two advantages of this approach are that it allows the estimated effect and the bound on the second derivative to both be expressed in terms of these underlying primitives, and that it allows for sensitivity analyses based on different distributional assumptions for the unobserved heterogeneity term.

The rest of this paper is organized as follows. Section 2 introduces our setup and notation. Section 3 reviews current recommendations for choosing the bound on the second derivative *a priori*. Section 4 presents our results for some well-known distributions and the implicit bounds on the second derivative of the conditional expectation function necessary to yield tests with the nominal size under different assumptions regarding the data-generating process, as well as conditions under which this second derivative can be nonparametrically and semi-parametrically identified. It also contains a general result which proves the existence of a general relationship between log-concave random variables' maximal first derivative and their variance. Section 5 describes our proposed estimation procedure. Section 6 presents simulation results that examine the performance of Honest confidence intervals using our semiparametric method and the closest alternative methods when different combinations of our key assumptions of monotonicity and concavity are either true or false of the underlying DGP. Section 7 concludes.

2 Setup and Notation

2.1 Sharp Regression Discontinuity Designs

Let $Y_i(j) \in \{0,1\}$ be the i^{th} unit's potential binary outcome in a random sample of n units, with $j=1$ for the treated outcome $Y_i(1)$ and $j=0$ for the untreated outcome $Y_i(0)$. Let $R_i \in \mathcal{R} := \{R_k : k \in \mathbb{Z}, R_k < R_{k+1}\}$ be the value of the discrete ordered running variable for that unit, and treatment status $D_i \in \{0,1\}$ be assigned if $R_i \geq R^0$ for some R^0. In most applications, R^0 is normalized to $R^0 = 0$, implying that $D_i = 1[R_i \geq 0]$. The treatment effect of interest is

$$\tau = E[Y_i(1) - Y_i(0) | R_i = 0], \tag{1}$$

the average effect of treatment on the treated (ATT) at the point $R_i = R^0 = 0$, the cutoff value, which is not identified from the data alone as we do not observe untreated outcomes for treated units (equivalently in this setup, $R_i \geq 0$). Throughout this paper, we will assume that the following identifying assumption holds:

Assumption A1. $\lim_{r \to 0^-} E[Y_i | R_i = r] = E[Y_i(0) | R_i = 0]$

This assumption is equivalent to assuming that unobservable differences between treated and untreated states trend smoothly at $R = 0$. It can be shown that, under the above assumption (Lee and Lemieux, 2010),

$$\tau = E[Y_i(1) - Y_i(0) | R_i = 0] = \lim_{r \to 0^+} E[Y_i | R_i = r] - \lim_{r \to 0^-} E[Y_i | R_i = r], \tag{2}$$

which follows from the treated state tending to 0 from above and Assumption A1. In a regression equation, if we write the potential outcomes conditional on R as $Y_i(0,R) = f(R) + u_i$ and $Y_i(1,R) = \beta + g(R) + u_i$, where u_i is unobserved by the econometrician, we can estimate

$$Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i) = f(R)(1 - D_i) + \beta D_i + g(R)D_i + u_i, \tag{3}$$

in which, under Assumption A1,

$$\beta = \lim_{r \to 0^+} E[Y_i|R_i = r] - \lim_{r \to 0^-} E[Y_i|R_i = r] = E[Y_i(1) - Y_i(0)|R_i = 0], \tag{4}$$

so that any consistent estimator of the average difference in levels between the treated and untreated states at $R = 0$, β, will also be a consistent estimator of $\tau = E[Y_i(1) - Y_i(0)|R_i = 0]$.[2] If instead assignment to treatment status is not certain, but probabilistic, so that some units with $R_i < 0$ are treated or some units with $R_i \geq 0$ are not treated, then the design is a "fuzzy" design and the causal effect estimated is interpretable as a local average treatment effect (LATE) as in the case of instrumental variables (Hahn et al., 2001). While we do not deal with the fuzzy design in this paper for the sake of clarity, the extension of our method to the fuzzy case is straightforward given the method for constructing Honest confidence sets in the fuzzy design proposed by Noack and Rothe (2021).

2.2 Honest Confidence Intervals

Throughout this paper, the assumed goal is to construct confidence intervals that have the following property: given functions $m(\cdot)$ that belong to the set $\mathcal{M}$, a significance level α, coverage probability P_m (which depends on $m(\cdot)$), and the object of interest defined above τ,

$$\liminf_{N \to \infty} \inf_{m \in \mathcal{M}} P_m(\tau \in C^{1-\alpha}) \geq 1 - \alpha, \tag{5}$$

which, following Li (1989), Kolesár and Rothe (2018) and Noack and Rothe (2021), is what we will mean by the statement "$C^{1-\alpha}$ is honest with respect to $\mathcal{M}$". The above definition makes clear that Honesty of the confidence intervals we focus on in this paper relates to their asymptotic uniform coverage probability. Honesty is desirable as a property of confidence

[2]For comprehensive overviews of the regression discontinuity literature, see Lee and Lemieux (2010), Cattaneo et al. (2019) or Cattaneo and Titiunik (2022).

intervals since every Honest confidence interval has the nominal rate of pointwise asymptotic coverage $1-\alpha$ if $m(\cdot) \in \mathcal{M}$. Kolesár and Rothe (2018) show that there are at least two different ways to specify the class $\mathcal{M}$ to construct honest CIs with respect to $\mathcal{M}$: Kolesár and Rothe (2018) define it as either the set of functions $m(\cdot)$ such that $m(\cdot) = E[Y|R]$ at the discrete points of support of R with probability one and there is an upper bound on the bias of the local linear estimator $\hat{\tau}^{LL}$ of τ, or the set of functions such that $m(\cdot) = E[Y|R]$ at the discrete points of support of R and is twice continuously differentiable with a global upper bound K on its second derivative $m''(\cdot)$. In the rest of this paper, we focus on the latter type of assumption, which produces "Bounded Second Derivative" (BSD) CIs.

2.3 Bounded Second Derivative (BSD) Confidence Intervals

Throughout the remainder of this paper, we will make the following assumption:

Assumption A2. *There exists a function $m(\cdot) : \mathbb{R} \to \mathbb{R}$ that is twice continuously differentiable either side of $R = 0$, and potentially discontinuous at $R = 0$, such that at each point of support $R \in \mathbb{Z}$, $m(R) = E[Y|R]$ with probability one.*

This assumption is necessary since the conditional expectation function $E[Y|R]$ is only defined for the discrete points of support for the running variable R. Since $m(\cdot)$ is equal to $E[Y|R]$ at these discrete points of support, it is relevant to any bias that arises from misspecification error in the estimation of τ. If $m(\cdot)$ is linear, then a local linear estimator of the difference between the treated and untreated groups at the cutoff $R = 0$ does not suffer from misspecification error. Since it is twice continuously differentiable, its second derivative exists (unlike that of $E[Y|R]$) and can be bounded. We are then in a position to define the class of functions $m(\cdot)$ that will allow us, conditional on a bound on the second derivative of these functions, to construct honest confidence intervals. Without additional assumptions, since $m(\cdot)$ is only observed at the discrete points of support of the running variable R, we cannot identify $m(\cdot)$ or its derivatives.[3] In Section 4, we show that if the

[3]This important caveat applies to the entire existing literature on CIs in RDDs with discrete running

outcome Y is a binary variable generated by a threshold-crossing model, then identification results already exist for identifying an upper bound for the second derivative of $m(\cdot)$, conditional on assuming a known parametric distribution for the unobserved component of the utility function, η, when the outcome Y is generated by a threshold-crossing model $Y = 1[g(R) \leq \eta]$.[4]

Assumption A3. *(Kolesár and Rothe, 2018) There exists some $K > 0$ and some $q \geq h > 0$ such that $m(\cdot)$ belongs to the set $\mathcal{M}(K) = \{m : |m'(a) - m'(b)| \leq K|a - b| \text{for all } a, b \in [-q, 0) \text{and all } a, b \in [0, q]\}$.*

We can note that this condition applies equally to the left- and right-derivatives of $m(.)$ from either side of the normalized cutoff value $R = 0$. Kolesár and Rothe (2018) prove that if the function $m(\cdot)$ defined above is in the class $\mathcal{M}(K)$ for the correct assumed value of K or larger, then confidence intervals constructed on that basis (for which they provide a method of doing so) are honest with respect to $\mathcal{M}(K)$. Note also that $m(\cdot)$ belongs to the Hölder class within a bounded range of the support of the running variable, so that the second derivative need not be bounded on the entire real line. This will be useful later on when we restrict $m(\cdot)$ to be a composition of a CDF and a monotone, concave function, since the set of such functions with bounded second derivatives on a subset of the real line is larger than the set of functions that are monotone, concave, and have bounded second derivatives for all real inputs.[5]

---

variables. Our analysis proceeds within the technical framework offered in this literature and leaves filling this caveat for future research.

[4]Chernozhukov et al. (2023) provide finite-sample guarantees for debiased machine-learning methods that assume that the conditional expectation function of interest is linear and mean square continuous in the parameter space. These conditions appear to be stronger than the ones that we require for the population function $m(\cdot)$ that generates the conditional expectation function at the points of support of R.

[5]Kolesár and Rothe (2018) also consider, in addition to Bounded Second Derivative (BSD) confidence intervals, Bounded Misspecification Error (BME) confidence intervals that pre-specify a maximum deviation of the estimated CEF from the true CEF. We do not examine this alternative in this paper as few econometric models imply a bound on the difference between the true and estimated CEFs, while commonly used discrete choice models do imply an upper bound for the second derivative of the CEF provided that the second derivative of the index function is bounded (cf. Section 4, below).

2.4 Local Linear Regression

A common method for estimating τ via a regression-discontinuity design is to approximate the conditional expectation function either side of the cutoff value $R = 0$ via a local linear regression (Fan and Gijbels, 2018). Suppose, given data on the outcome Y_i and running variable R_i, we use a bandwidth $h > 0$ to choose observations "close" to the cutoff, so that $R_i \in [-h, h]$ for all i included in the sample, and let N_h denote the sample size for the observations included when the bandwidth is h. Then the local linear estimator $\hat{\tau}^{LL}$ is just the coefficient on the indicator $1[R_i \geq 0]$ in an OLS regression using the N_h observations that lie in the interval $-h \leq R_i \leq h$ that includes $1[R_i \geq 0]$, R_i and $1[R_i \geq 0] \times R_i$ as regressors, viz.:

$$Y_i(1)D_i + Y_i(0)(1 - D_i) = Y_i = \beta_0 + \beta_1 R_i + \beta_2 1[R_i \geq 0] + \beta_3 R_i \times 1[R_i \geq 0] + u_i, K\left(\frac{R_i}{h}\right) < 1,$$

so that $\hat{\beta}_2 = \hat{\tau}^{LL}$. If $Y_i \in \{0, 1\}$, then local linear estimation is equivalent to estimating a linear probability model over the range of R chosen by the bandwidth and kernel function, so that $E[u|R] = Pr(Y = 1|R) - (\beta_0 + \beta_1 R_i + \beta_2 1[R_i \geq 0] + \beta_3 R_i \times 1[R_i \geq 0])$. Throughout the rest of this paper we use the Uniform kernel $K(\cdot) = |\cdot|$, so choosing a bandwidth h amounts to choosing a range of observations with $R_i \in [-h, h]$.

Kolesár and Rothe (2018) show that the local linear estimator $\hat{\tau}^{LL}$ is biased for τ, with the bias bounded from above by (with $w(R_i)$ defined in Kolesár and Rothe (2018) as $w(R_i) = \frac{e_1' \hat{W} M_i}{N_h}$, $M_i = m(R_i) = (1[R_i \geq 0], 1[R_i \geq 0] \times R_i, 1, R_i)'$, $\hat{W} = \frac{1}{N_h} \sum_{i=1}^{N_h} M_i M_i'$, and $e_1 = (1, 0, ..., 0)'$),

$$\left|\hat{\tau}^{LL} - \tau\right| \leq -K \frac{\sum_{i=1}^{n} w(R_i) R_i^2 sgn(R_i)}{2}, \tag{6}$$

which holds with probability 1, regardless of the sample size. Since all other components of the above expression are data, specifying some value for the tuning parameter $K > 0$ is sufficient to construct honest confidence intervals with respect to the class $\mathcal{M}(K)$. Bounding the finite-sample bias *a priori* and constructing Honest confidence intervals based on that bound produces valid inference when the running variable is discrete, while Eicker-Huber-White standard errors do not, for the following reason: (6) holds regardless of whether R is discrete or continuous, since it is derived conditional on particular realizations of the running variable $\{R_1, ..., R_{N_h}\}$ for a given bandwidth h. By contrast, EHW standard errors rely on an asymptotic justification that the bias in the local polynomial estimator converges in probability to zero as the number of observations goes to infinity; this argument does not hold in the case of a discrete running variable, as there are no observations in between the points $R^0 = 0$ and R^{-1}, the last point of support to the left of the cutoff, regardless of the sample size. By avoiding these asymptotic arguments, confidence intervals constructed using the bound on the bias of the estimator in (6) provide at least the nominal rate of coverage regardless of the bandwidth or sample size, conditional on K being chosen so as to be larger than the second derivative of the unknown function $m(\cdot)$.[6]

In Section 4, we show that the terms that need to be bounded *a priori* by the researcher can be written in terms of population parameters that are interpretable and can in principle be identified from the data under some common assumptions made in the literature on discrete choice models. Before doing so, we review prior recommendations for choosing the uniform bound on the second derivative of $m(\cdot)$.

[6]An alternative approach to discreteness in the running variable is discussed in Dong (2015), where the realizations of the running variable are rounded versions of the true, continuous running variable. Letting $\triangle_R$ represent the "size" of the gaps between the realizations of the running variable, Dong (2015) recommends a procedure for estimating $\triangle_R$ under the assumption that the rounding errors are Uniformly distributed on $[0, 1]$, thereby obtaining bias-aware confidence intervals by directly estimating the bias that is due to measurement error in the running variable. The Honest confidence intervals explored here are valid more generally regardless of the source of the discreteness in the running variable (see main text and Kolesár and Rothe (2018)). A related situation arises when the running variable is continuous but there is nonrandom "heaping" of observations at discrete points of support in the running variable (Barreca et al., 2016). We view this as complementary to the setting we consider, as nonrandom selection into particular points of support of a discrete running variable comes with its own inferential problems (as laid out in Barreca et al. (2016)), but we do not pursue the interaction of these two approaches in this paper.

3 Existing Proposals for Uniform Bounds on the Second Derivative

This section reviews the current guidance on selecting the tuning parameters that are necessary for bias-aware inference. These tuning parameters often take the form of a "smoothness condition" such as a bound on the second derivative of an unknown function, which is the focus of this paper. They may also directly set an upper limit for the extent of the bias of the local linear estimator, as is done in Kolesár and Rothe (2018) via "Bounded Misspecification Error" (BME) Honest confidence intervals. In both cases, it is not possible without further assumptions to choose these upper bounds based on the data, as the data only speak to the minimum variability of the outcome of interest, not the maximum (as it could always be the case that the latter is not attained in the sample).

Kolesár and Rothe (2018), who introduce Bounded Second Derivative (BSD) confidence intervals, themselves make the following two recommendations. First, they write, "One can show by simple algebra that if [$m(\cdot) \in \mathcal{M}(K)$], then it differs by at most $K \times (x^{**} - x^{*})^2/8$ from a straight line between [$m(x^{**})$ and $m(x^{*})$]." An advantage of this heuristic is that there are many data-generating processes that can be consistent with it. A disadvantage is that it can be relatively difficult to infer which features of data-generating processes are being ruled out by this restriction, and hence whether those are reasonable restrictions in the context at hand.

Kolesár and Rothe (2018) go on to apply this heuristic in an empirical setting. In their empirical application (a reexamination of Oreopoulos (2006)), the outcome of interest is the natural logarithm of annual earnings (in constant 1998 pounds sterling). To motivate different choices of the uniform bound K, they write: "Given that a typical increase in log earnings per extra year in age is about 0.02, we consider $K = 0.04$ and $K = 0.02$ to be reasonable choices, with the other values corresponding to a very optimistic and a very conservative view, respectively". One problem with this criterion is that it does not distinguish between changes in log earnings per year over a given worker's life cycle, and differences in log earnings between cohorts born a year apart. It is the latter that is relevant for the treatment effect of interest in Oreopoulos (2006), since the comparison being made in that study is between different cohorts that are observationally similar except in the number of years of compulsory schooling they face. A second potential problem is that choosing $K = 0.04$ as a reasonable bound on the second derivative when the first derivative is on average 0.02 places implicit shape restrictions on the data generating process, as noted above, which can be nontrivial to reverse-engineer from the heuristic they recommend. As a result, it is unclear what restrictions there may be on e.g. capital-labor complementarity, the evolution of the labor share of output, or technical progress that make this a reasonable choice for a given observed trend in annual earnings.

Noack and Rothe (2021) add their own suggestions for choosing smoothness parameters and review a number of other suggestions in the literature. One approach they recommend is visual inspection of the data combined with estimating functions of different curvature to examine the plausibility of those functions as data-generating processes. The key assumptions we use in our method of monotonicity and weak concavity of an index function (defined below) share the advantage of this approach in that their plausibility can be assessed by inspection. Two rules of thumb are mentioned and assessed in their Monte Carlo simulations: the first, due to Armstrong and Kolesár (2020), is based on estimating the maximal second derivative of a fourth-order polynominal fitted to the data. Noack and Rothe (2021) find that the Armstrong-Kolesar approach leads to extremely conservative confidence intervals in the case of a binary outcome variable, such as a first-stage treatment status indicator (a finding echoed in the simulation study in this paper). The other, due to Imbens and Wager (2019), is to multiply the second derivative estimated of a quadratic fit to the data by a "suitably large" constant, such as 2. This is found to be less conservative in the settings they consider. Neither of these rules of thumb make use of subject matter knowledge or economic theory to derive their heuristic estimates of the uniform bound on the second derivative. In our Monte Carlo simulations and empirical application, we compare the confidence intervals that result from using these rules of thumb with our own method.

4 Binary Outcome Threshold-Crossing Models with Bounded Second Derivatives

We propose an approach that allows researchers to use the sample data to bound the second derivative of the function $m(\cdot)$ in absolute value. This approach relies on assumptions that are common in the literature on discrete choice modeling and economic theory.

Discrete-choice models have an important advantage when the goal is to identify a uniform bound on the second derivative of $m(\cdot)$: it can be semiparametrically identified from the data even when the running variable R is not conditionally independent of the unobservables (Matzkin, 1991). This is not true for generic functions of R with bounded second derivatives. In regression-discontinuity designs, researchers typically leave the relationship between the running variable R and unobservables u unrestricted except for the identifying assumption that $E[u|R]$ is continuous at the cutoff point $R = R_0$. In discrete-outcome settings, the assumption that the outcome is generated by an additively separable threshold-crossing model allows us to identify the derivatives of $m(\cdot)$ without assuming that R is exogenous, as long as we are prepared to adopt a distributional assumption for the unobserved latent "reservation utility" (a type of assumption that is common in the discrete-choice literature) and shape restrictions on utility functions that can be motivated by economic theory or institutional detail.

Assumption A4. *The binary outcome potential outcome $Y_i(j) \in \{0,1\}$ for $j \in \{0,1\}$ is generated by the threshold-crossing rule $Y_i(j) = 1[\eta_{ij} \leq g_j(R_i)]$, where $g_j(\cdot)$ is twice differentiable and η_{ij} is unobservable and iid across individuals.*

The above assumption implies that, if we let F_{η_j} denote the CDF of η_j, $Pr(Y_i(j) = 1|R_i) = F_{\eta_j}(g_j(R_i))$.

Assumption A5. *Let η_1 be the iid unobservable reservation utility in the treated state*

$(D=1)$ and η_0 iid unobservable reservation utility in the untreated state $(D=0)$. Let F_{η_1}, F_{η_0} denote the CDFs of these random variables. Then $F_{\eta_1} = F_{\eta_0} = F_\eta$ within the neighborhood of $R=0$, $[-q,q]$ in which $m(R)$ has its second derivative bounded from above in absolute value by K. This CDF F_η has a density f_η, which has a first derivative f'_η defined everywhere in $[-q,q]$.

In the rest of this paper, while we do not make the same assumption for the index functions (i.e. we let $g_1(\cdot) \neq g_0(\cdot)$), we suppress the potential outcomes notation for compactness and write $Y_i = 1[\eta_i \leq g(R_i)]$. An upper bound on $|g(\cdot)|$, for example, can be understood in what follows to mean an upper bound on $\max_{j\in\{0,1\}} |g_j(\cdot)|$. Note that the foregoing assumption does not imply that $Y_i(1) = Y_i(0)$ or $E[Y_i(1)] = E[Y_i(0)]$, since g_1 and g_0 can still differ. We do not observe g_0 when $R \geq 0$, or g_1 when $R < 0$, so a common distribution of the unobserved reservation utilities in the treated and untreated states does not imply that we can identify $E[Y_i(1) - Y_i(0)|D_i = 1]$, the effect of treatment on the treated, from comparing treated with untreated units, since the difference between treated and untreated units in the data is $E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] = F_\eta(g_1(D_i = 1)) - F_\eta(g_0(D_i = 1))$, and the term $F_\eta(g_0(D_i = 1))$ is not observed in the data.

Assumption A6. *η is independent of R.*

The plausibility of this assumption, on its own, will vary from context to context. If it is violated, then a control function approach can be employed, with a suitable choice of instrumental variable Z, with the control function denoted $\xi(Z)$ so that η is independent of R conditional on $\xi(Z)$.

Assumption A7. *There exists a function $m(R_i) \equiv F_\eta(g(R_i))$, with F_η and $g(\cdot)$ defined as above, which is twice differentiable.*

Hence, by $g(\cdot) \equiv D_i g_1(\cdot) + (1 - D_i) g_0(\cdot)$ being twice differentiable (except at $R = 0$, where it may be discontinuous) and the composition of functions being twice differentiable, $F_\eta(g)'$ and $F_\eta(g)''$ exist. This assumption refines the earlier assumption that there exists some such function equal to $E[Y_i|R_i]$ at the points of support of R with probability one which is twice differentiable; here the more specific function given is clearly equal to $E[Y_i|R_i]$ at integer values of R given the binary threshold-crossing process that defines Y. Similarly, we can write the average effect of treatment on the treated at the point $R = 0$ as $E[Y_i(1) - Y_i(0)|D_i = 1] = F_\eta(g_1(0)) - F_\eta(g_0(0))$, where $F_\eta(g_0(0))$ is not observed (since $g_0(\cdot)$ is only observed for $R < 0$). The local linear regression-discontinuity estimator estimates $\hat{\tau}^{LL} = F_\eta(g_1(0)) - \hat{F}_\eta(g_0(0))$, where $\hat{F}_\eta(g_0(0))$ is predicted by the local linear fit. Since F_η is in general a nonlinear function in most discrete-outcome settings, a linear approximation to it is unlikely to be exact; two jointly sufficient conditions for this (see below) occur in the special case where η has the Uniform distribution and $g(\cdot)$ is itself linear.

We can now state the requirement from Assumptions (A2) and (A3) that the second derivative of $m(\cdot)$, which is equal to $E[Y|R]$ at all of the discrete points of support of R, be bounded in absolute value in terms of the functions F_η and $g(\cdot)$:

$$\left| f_\eta(g(R))g''(R) + f'_\eta(g(R))[g'(R)]^2 \right| < K, \text{some } K > 0, \tag{7}$$

where f_η and f'_η denote the pdf of η and its first derivative. Since the distribution of η is assumed to be known, whether the uniform bound can be identified from the data, if it exists, depends on the identifiability of $g(\cdot)$. Note that $K = 0$ only in the special cases where either $f_\eta(g(R))g''(R) = -f'_\eta(g(R))[g'(R)]^2$ or $\eta \sim Uniform(\omega_1, \omega_2)$ and $g(\cdot)$ is linear.

Common choices for the distribution of η in the discrete choice literature, conditional on location and scale normalizations, have analytic bounds on their probability distribution function f_η and its first derivative f'_η.[7] Suppose that η has the Logistic distribution. Its CDF is then

$$F_\eta(\eta) = \frac{1}{1 + \exp(-\frac{\eta - \mu}{s_\eta})}, \tag{8}$$

[7] An alternative approach to the one we pursue here is to use Probit and Logit estimation in conjunction with a regression-discontinuity design, derived by Xu (2017). The result in Armstrong and Kolesár (2018) that necessitates an assumption regarding bounds on the second derivative of the conditional expectation function (CEF) implies that the confidence intervals derived in Xu (2017) are conservative relative to the local linear estimator if the assumption that the second derivative of the CEF is no more than K is true.

from which it is clear that, given that the PDF and its first derivative appear in the expression for $m''(R)$, the scale of the distribution plays a key role in the second derivative of $m(\cdot)$. In fact this generalizes to all log-concave distributions, a family which includes the Logistic and Normal distributions as well as other commonly used distributions such as the F-distribution. The following Proposition formalizes this result.

Proposition 1. *If X and Z are log-concave random variables, with $Z = \alpha X$, and $|\alpha| > 1$, then the maximum of the first derivative of $f_Z(.)$ is smaller than the maximum of the first derivative of $f_X(.)$.*

Proof. See Appendix. ■ □

The location parameter μ and scale parameter s cannot be identified from the data separately from the index function $g(\cdot)$ (Train, 2009), so in practice the location-scale normalization $\mu = 0$, $s = 1$ is imposed in estimation, yielding

$$F_\eta(\eta) = \frac{1}{1 + \exp(-\eta)}, \tag{9}$$

Since this normalization is different for different choices for the distribution of η, care must be taken when comparing estimates of $g'(\cdot)$ and $g''(\cdot)$ that are obtained using different distributional assumptions for η, as their scale will be different. In the case where η has the normalized Logistic distribution, it can be shown (see Appendix) that the second derivative of $m(\cdot)$ is at most

$$|m''(R)| \leq \left| \frac{1}{4} g''(R) + \frac{1}{6\sqrt{3}} [g'(R)]^2 \right|, \tag{10}$$

Suppose instead η is Normally distributed. For the same non-identification reasons as with the Logistic distribution, it is common in the literature to normalize the location and scale of η so that it has the Standard Normal distribution, in which case

$$m(R) = \Phi(g(R)), \tag{11}$$

where $\Phi(\cdot)$ denotes the CDF of a Standard Normal random variable. In this case it can be shown that (see Appendix)

$$|m''(R)| \leq \left| \frac{1}{\sqrt{2\pi}} g''(R) + \frac{1}{\sqrt{2\pi e}} [g'(R)]^2 \right|, \tag{12}$$

Note that $\frac{\sqrt{2\pi e}}{\sqrt{2\pi}} = \sqrt{e} \approx 1.649$ and $\frac{6\sqrt{3}}{4} \approx 2.598 > 1.649$, so that the assumption of Normally distributed η places slightly less weight on the second derivative of $g(\cdot)$ than the assumption that η has the Logistic distribution. If $g(\cdot)$ is a linear function, then since the coefficients from a linear index Logit model are $\sqrt{1.6}$ times larger than those from a Probit model with a linear index estimated on the same data (due to the difference in scale normalizations Train (2009)), the ratio of upper bounds on the second derivative is $\frac{1.6}{6\sqrt{3}} \Big/ \frac{1}{\sqrt{2\pi e}} \approx 0.64$, so that the assumption that η has the Logistic distribution when $g(\cdot)$ is linear results in a tighter upper bound for the second derivative of $m''(\cdot)$ and hence narrower confidence intervals than the assumption that η has the Standard Normal distribution. This is a direct consequence of the fact that the maximal first derivative of the Standard Normal pdf is $\frac{1}{\sqrt{2\pi e}} \approx 0.24$ whereas the maximal first derivative of the Logistic pdf with location parameter $\mu = 0$ and scale parameter $s_\eta = 1$ is $\frac{1}{6\sqrt{3}} \approx 0.096$, which in turn is consistent with Proposition 1 (above) that this maximum is decreasing in the scale of the distribution for log-concave random variables.

The remaining inputs into the uniform bound on the second derivative of $m(.)$ are the upper bounds on the first and second derivative of the index function $g(\cdot)$. We leverage prior results from the literature on semiparametric identification in discrete choice models using shape restrictions to motivate data driven choices for these bounds. In particular, Matzkin (1991) shows that the following assumptions are jointly sufficient to identify $g(\cdot)$ from the observed data.

Assumption A8. *The distribution of η is known up to some finite-dimensional parameter vector.*

Assumption A9. *$g(.)$ is monotone in R.*[8]

Assumption A10. *$g(\cdot)$ is weakly concave in R.*

Proposition 2. If the foregoing three assumptions are all met, then $g(\cdot)$ is identified up to a location and scale normalization.

Proof. See Matzkin (1991). ■

A further result in Matzkin (1992) shows that it is in principle possible to achieve fully nonparametric and distribution-free identification of $g(.)$ and F_η, but only under the additional, highly restrictive assumption that $g(\cdot)$ is homogeneous of degree one. The assumption of homogeneity of degree one may be implausible in many contexts of interest. For example, if R is age in years and Y is a binary indicator for retirement status, this assumption implies that the difference in the marginal value of leisure for the same individual at any two adjacent ages is the same regardless of whether we are comparing e.g. age 60 vs. 61 or age 69 vs. 70 (net of any discontinuities or kinks generated by policy rules surrounding retirement and/or retirement benefits). As a consequence, we restrict attention in the rest of the paper to identifying $g(\cdot)$ conditional on assuming that η is drawn from a known probability distribution.

5 Estimation

This section describes the empirical implementation of the theoretical results above. To avoid additional parametric assumptions, we can estimate a piecewise linear spline with

[8] Two other recent approaches to regression discontinuity designs that exploit monotonicity can be found in Kwon and Kwon (2020) and Babii and Kumar (2021). Kwon and Kwon (2020) construct confidence intervals that have the nominal rate of coverage when the true conditional expectation function is monotone and Lipschitz continuous. This is similar to the bias-aware approach in this paper, as the researcher has to specify the Lipschitz constant, which then bounds the first derivative of the CEF in absolute value if it is differentiable. Babii and Kumar (2021) proceed from previous work in the statistics literature on isotonic regression estimators and assume that the population conditional expectation function is Hölder continuous with known Hölder constant. Neither paper explicitly considers the case where the outcome variable is discrete, as their methods apply equally to continuous and discrete outcome variables; this paper's contribution is to exploit the fact that common assumptions made in discrete-choice models and economic theory can be used to motivate the shape restrictions and distributional choices that allow an upper bound for the second derivative of the function that generates the conditional expectation function.

$J+1$ intercepts and $J+1$ slope parameters δ^j, where J is the number of knots in the spline and δ^j is the estimated slope on the spline connecting knots j and $j+1$. This is done by maximizing the likelihood

$$\mathcal{L} = \prod_{i=1}^{N} F_\eta(\sum_{j=1}^{J}(\alpha_j + \delta_j R))^{Y_i}(1 - F_\eta(\sum_{j=1}^{J}(\alpha_j + \delta_j R))^{Y_i}))^{1-Y_i}$$

where F_η corresponds to the CDF of the assumed distribution for η (in our simulations and empirical application (in the Appendix), either the Logistic or Normal CDF, as these are the two most popular choices in the discrete choice literature). It is then straightforward to estimate, under the same assumptions as above, the maximal slope as $\max_{k=1,\dots,J+1} \left| \sum_{j=1}^{k} \delta^j \right|$ and maximal second derivative as either equal to the maximal slope (as in the conservative approach) or the maximal difference in subsequent slope estimates, $\max_{j=2,\dots,J+1} \delta^j - \delta^{j-1}$. An alternative approach to estimation which we do not pursue in this paper is given by Matzkin (1991); details regarding the computation of this estimator can be found in Matzkin (1999). Her estimator incorporates the monotonicity and concavity restrictions on the index function $g(R)$; this estimator is likely to be more efficient than using splines that do not incorporate the identifying assumptions directly, but since the objective of our estimation is to recover an upper bound, rather than a least upper bound, the main benefit to using this alternative estimator would be to make the resulting Bounded Second Derivative confidence intervals less conservative than the ones that we estimate (cf. Section 6, below).

Using a piecewise linear spline to approximate the underlying monotone and concave function $g(\cdot)$ leaves open the question of the optimal number of knots in the spline. We leave this to the discretion of the researcher, as these will vary from context to context. In this paper, we undersmooth so as to err on the side of overestimating the second derivative of the underlying function $m(\cdot)$, since Bounded Second Derivative confidence intervals are Honest with respect to the Hölder class of functions that have second derivative no larger than the assumed value K. Hence overestimating the upper bound on the second derivative K is less costly than underestimating it, since any upper bound will guarantee at least nominal coverage of the resulting BSD confidence intervals (though overestimating K by a large enough amount is likely to result in conservative confidence intervals, as will be evident in the simulation results below).

Our estimation strategy can be summarized via the following steps:

1. Check whether monotonicity and concavity of the index function $g(\cdot)$ are plausible according to theory or the institutional setting; if so, then the first and second derivatives of $g(\cdot)$ can be semiparametrically identified from the data conditional on a distributional assumption for η by the results in Matzkin (1991). In practice, we use piecewise linear splines to approximate $g(\cdot)$.

2. Estimate Bounded Second Derivative confidence intervals for the treatment effect of interest τ using the implied uniform bound K on the second derivative of $m(\cdot)$ and local linear regression, which can be implemented using standard software. Optimal bandwidths are automatically calculated in the RDHonest package in R conditional on the assumed value, K, of the uniform bound on the second derivative of $m(\cdot)$, in line with the formulae in Kolesár and Rothe (2018).

6 Simulation Evidence

In this section, we compare the results from using the identifying assumptions from the previous section to semiparametrically identify the second derivative of the conditional expectation function $E[Y|R]$ from the data to the "automatic" selection procedure used in Kolesár and Rothe (2018), as well as the rule of thumb recommended in Imbens and Wager (2019). We vary two dimensions of the estimation: whether the distribution of η has (i) the Logistic distribution, or (ii) the Normal distribution, and whether the index function $g(.)$ is one of

1. Partially linear, monotone and concave with $g(R) = 0.25R \times 1[R < 0] + (\log(1 + 0.25R) + 0.05) \times 1[R \geq 0]$, (ii)

2. Monotone and partially linear but not concave, with $g(R) = (\exp(0.25R) - 1) \times 1[R < 0] + (0.05 + 0.25R) \times 1[R \geq 0]$,

3. Concave, with bounded first and second derivatives, but not monotone, with $g(R) = -\frac{R^2}{32} \times 1[R \in [-4, 4]] + 0.05 \times 1[R \geq 0] + 0.25R \times 1[R < -4] - 0.25R \times 1[R > 4]$,

4. Neither monotone nor concave with bounded first and second derivatives, with $g(R) = \sin(0.25R) + 0.05 \times 1[R \geq 0]$.

This provides eight different scenarios in which the second derivative of $F_\eta(g(R))$ is bounded in absolute value, but the assumptions for semiparametric identification of the derivatives of $F_\eta(g(.))$ are satisfied to different extents. In each simulation, the running variable, R, is drawn from a discrete Uniform distribution with points of support $\{-5, -4, ..., 0, 1, ..., 5\}$, with the cutoff value $R^0 = 0$. We use 1000 replications per scenario, with $N = 1100$ in each case (giving approximately 100 observations per point of support of the running variable). Outcomes of interest are the average estimated upper bound on the second derivative K, Type I error rate, and the average confidence interval estimated across the 1000 replications. The simulation scenarios are designed so that for each of the cases where the first and second derivatives of $g(.)$ are bounded in absolute value, these bounds are $|g'(R)| \leq \frac{1}{4}$ and $|g''(R)| \leq \frac{1}{16}$. When η has the Logistic distribution, the analytical upper bound for the second derivative of $F_\eta(g(.))$ in each case is $\frac{1}{4} \times \frac{1}{16} + \frac{1}{6\sqrt{3}} \times \frac{1}{16} \approx 0.022$; when η has the Standard Normal distribution, the upper bound is $\frac{1}{\sqrt{2\pi}} \times \frac{1}{16} + \frac{1}{\sqrt{2\pi e}} \times \frac{1}{16} \approx 0.040$. These values for the upper bound on the second derivative of the CEF are within the range of values considered in previous studies of Bounded Second Derivative confidence intervals (cf. Kolesár and Rothe (2018), who use 0.02 as their preferred upper bound, and 0.04 as a conservative upper bound, in one of their empirical applications, or Noack and Rothe (2021)).

The design of the DGPs for the simulations also allows us to control the size of the true effect τ. In each case, we have $g_0(0) = 0$ and $g_1(0) = 0.05$, so that $\tau = F_\eta(0.05) - F_\eta(0)$. When $\eta \sim Logistic(0, 1)$, $\tau = 0.0125$, and when $\eta \sim Normal(0, 1)$, $\tau = 0.0199$.

We compare three approaches to bounding the second derivative from above: (i) the semiparametric approach proposed in this paper, (ii) the RDHonest package's default, proposed in Armstrong and Kolesár (2020), and (iii) the approach proposed by Imbens and Wager (2019), which involves multiplying the coefficient on the quadratic term following estimation of a local or global polynomial by a "suitably large constant" - we use local quadratic regressions, and multiply the maximum of the two quadratic terms by 2.

Table 1 displays the results of the Monte Carlo experiments. It appears that the semiparametric approach to estimating an upper bound on the second derivative of $m(\cdot)$ is less conservative than the Armstrong-Kolesar approach that is built into the RDHonest *R* package, and more conservative than the Imbens-Wager approach when the estimated second derivative from local quadratic regressions is multiplied by 2. The latter approach does not have the nominal rate of coverage in all of the scenarios considered: when the index function $g(.)$ is monotone but not concave, and η has the Standard Normal distribution, the Type I error rate using the Imbens-Wager bound is 0.057. When $g(.)$ is neither monotone nor concave, the Type I error rate from using the Imbens-Wager approach is 0.047. The approach proposed in this paper appears by contrast to be better protected from the failure of concavity by its relative conservatism - its Type I error rate does not exceed 0.039 in any of the simulations. Perhaps surprisingly, it does not appear to necessarily perform best when both monotonicity and concavity are satisfied, despite the necessity of these assumptions for identification. In the case of a non-monotone concave $g(.)$, our method appears to be much more conservative than in other scenarios. In every scenario, estimation that assumes that unobserved heterogeneity is Logistic is more conservative than estimation that assumes that it is Normally distributed. This differs from the case where $g'' = 0$, where estimation assuming Normally distributed unobserved heterogeneity η is more conservative. This shows that whether g'' is assumed to be zero or not can matter a great deal for how the choice of the distribution of η affects the construction of Bounded Second Derivative confidence intervals.

Overall, it appears that overestimating the maximal bound on the second derivative K

is less costly than underestimating it. While the actual second derivative of $F_\eta(g(R))$ in any particular sample need not attain this maximum, and so underestimating K need not result in worse-than-nominal coverage (as in the case of the Imbens-Wager bound when η is Normally distributed and $g(\cdot)$ is both monotone and concave), the conditions under which it does so are not easy to anticipate (cf. the Imbens-Wager bound with Normal η and monotone, but not concave $g(\cdot)$). By contrast, the combination of undersmoothing the spline that approximates $g(\cdot)$ and sampling variability produces "enough of an overestimate" of K to produce greater-than-nominal coverage of τ without the extreme conservatism of the default Armstrong-Kolesar selection method for K built into the RDHonest R package.

The general pattern seen in the Monte Carlo simulations also holds in an empirical application that we consider (see Appendix): our method is less conservative than the default Armstrong-Kolesar bound, but more conservative than the Imbens-Wager bound. Point estimates in the application differ depending on whether the Logistic or Normal bound is employed. This follows from point estimates differing depending on the bandwidth and the direct relationship between the optimal bandwidth and the Uniform bound on the second derivative that is chosen.

Table 1: Simulated Estimated Second Derivative Bounds, Average 95% Confidence Intervals, and Type I Error Rates

			$\eta \sim Logistic(0,1)$					$\eta \sim Normal(0,1)$				
Semiparametric Bound	$g(.)$ Monotonic?	$g(.)$ Concave?	True K	$\frac{1}{S}\sum_{s=1}^{S} K_s$	Type I	Ave. 95% CI		True K	$\frac{1}{S}\sum_{s=1}^{S} K_s$	Type I	Ave. 95% CI	
	Y	Y	0.022	0.143	0.003	-0.346	0.381	0.04	0.043	0.030	-0.166	0.265
	Y	N	0.022	0.157	0.008	-0.352	0.407	0.04	0.13	0.031	-0.28	0.386
	N	Y	0.022	0.21	0.001	-0.442	0.446	0.04	0.224	0.009	-0.457	0.455
	N	N	0.022	0.139	0.007	-0.339	0.379	0.04	0.061	0.039	-0.187	0.299
			$\eta \sim Logistic(0,1)$					$\eta \sim Normal(0,1)$				
Imbens-Wager Bound	$g(.)$ Monotonic?	$g(.)$ Concave?	True K	$\frac{1}{S}\sum_{s=1}^{S} K_s$	Type I	Ave. 95% CI		True K	$\frac{1}{S}\sum_{s=1}^{S} K_s$	Type I	Ave. 95% CI	
	Y	Y	0.022	0.023	0.024	-0.167	0.211	0.04	0.025	0.041	-0.150	0.232
	Y	N	0.022	0.024	0.027	-0.158	0.225	0.04	0.026	0.057	-0.134	0.252
	N	Y	0.022	0.048	0.020	-0.239	0.248	0.04	0.058	0.019	-0.258	0.265
	N	N	0.022	0.026	0.024	-0.167	0.223	0.04	0.03	0.046	-0.154	0.251
			$\eta \sim Logistic(0,1)$					$\eta \sim Normal(0,1)$				
Armstrong-Kolesar Bound	$g(.)$ Monotonic?	$g(.)$ Concave?	True K	$\frac{1}{S}\sum_{s=1}^{S} K_s$	Type I	Ave. 95% CI		True K	$\frac{1}{S}\sum_{s=1}^{S} K_s$	Type I	Ave. 95% CI	
	Y	Y	0.022	0.699	0	-0.998	1.029	0.04	0.626	0	-0.894	0.956
	Y	N	0.022	0.719	0	-1.011	1.063	0.04	0.671	0	-0.936	1.021
	N	Y	0.022	0.759	0	-1.082	1.086	0.04	0.764	0	-1.089	1.092
	N	N	0.022	0.707	0	-1.004	1.040	0.04	0.633	0	-0.899	0.968

Notes: Each panel displays the true upper bound on the second derivative of the conditional expectation function, average estimated upper bound, Type I error rate, and average confidence interval for each of four DGPs with differing properties of the index function $g(\cdot)$ in the data-generating process $E[Y|R] = F_\eta(g(R))$, where F_η denotes the CDF of either the standard Logistic or Standard Normal distributions. Each row is the result of 1000 replications, with $N = 1100$ and the running variable $R \in \{-5, -4, ..., 0, 1, ..., 5\}$. Average confidence intervals are calculated by separately averaging lower limits and upper limits for estimated confidence intervals across replications. The top panel displays results from simulations using the method derived in this paper to bound the second derivative of $m(\cdot)$ from above, the middle panel results using the larger coefficient (in absolute value) of the squared terms of a local quadratic specification multiplied by 2, as recommended in Imbens and Wager (2019), and the bottom panel uses the nonparametric method derived in Armstrong and Kolesár (2020), which is the default selection method in the RDHonest R package.

7 Conclusion

This paper has provided the first analysis of Honest inference for regression-discontinuity designs where both the assignment variable and the outcome of interest is discrete. Researchers who use "Bounded Second Derivative" Honest confidence intervals often have little guidance for how to choose the *a priori* uniform bound on the second derivative of an unknown function that these confidence intervals require. We proposed a procedure for identifying this second derivative based on the literature on the identification of discrete-choice models. A key advantage of using this procedure is that it allows the researcher to be transparent and precise about the assumptions necessary for identifying this second derivative from the same data that is used to estimate the treatment effect of interest. The most common alternative when constructing Honest confidence intervals is to rely on prior evidence from the literature, which may not be available in novel applications. Economic theory also typically does not make predictions about the variability of outcomes of interest.

Our simulation results provide evidence that when the necessary assumptions are satisfied, the second derivative of the conditional expectation function can be semiparametrically identified and used to construct confidence intervals that are Honest with respect to the set of possible functions that all have a second derivative that is no larger than the maximum inferred from the data. Our empirical application in the Appendix shows that estimating this second derivative under different assumptions can produce confidence intervals that differ in their substantive implications. A natural extension of our procedure for "sharp" regression discontinuity designs is to derive companion results for "fuzzy" designs (as studied in Noack and Rothe (2021)). We leave this extension to future research.

References

Almond, D., Doyle Jr, J. J., Kowalski, A. E., and Williams, H. (2010). Estimating Marginal Returns to Medical Care: Evidence from At-Risk Newborns. *The Quarterly Journal of Economics*, 125(2):591–634.

Andrews, I., Gentzkow, M., and Shapiro, J. M. (2020). On the informativeness of descriptive statistics for structural estimates. *Econometrica*, 88(6):2231–2258.

Armstrong, T. B. and Kolesár, M. (2018). Optimal Inference in a Class of Regression Models. *Econometrica*, 86(2):655–683.

Armstrong, T. B. and Kolesár, M. (2020). Simple and Honest Confidence Intervals in Nonparametric Regression. *Quantitative Economics*, 11(1):1–39.

Babii, A. and Kumar, R. (2021). Isotonic Regression Discontinuity Designs. *Journal of Econometrics*.

Barreca, A. I., Lindo, J. M., and Waddell, G. R. (2016). Heaping-induced bias in regression-discontinuity designs. *Economic inquiry*, 54(1):268–293.

Card, D., Dobkin, C., and Maestas, N. (2008). The Impact of Nearly Universal Insurance Coverage on Health Care Utilization and Health: Evidence from Medicare. *American Economic Review*, 98(5):2242–2258.

Card, D., Dobkin, C., and Maestas, N. (2009). Does Medicare Save Lives? *Quarterly Journal of Economics*, 124(2):597–636.

Cattaneo, M. D., Frandsen, B. R., and Titiunik, R. (2015). Randomization inference in the regression discontinuity design: An application to party advantages in the us senate. *Journal of Causal Inference*, 3(1):1–24.

Cattaneo, M. D., Idrobo, N., and Titiunik, R. (2019). *A practical introduction to regression discontinuity designs: Foundations*. Cambridge University Press.

Cattaneo, M. D. and Titiunik, R. (2022). Regression discontinuity designs. *Annual Review of Economics*, 14:821–851.

Cattaneo, M. D., Titiunik, R., and Vazquez-Bare, G. (2017). Comparing inference approaches for rd designs: A reexamination of the effect of head start on child mortality. *Journal of Policy Analysis and Management*, 36(3):643–681.

Chernozhukov, V., Newey, W. K., and Singh, R. (2023). A simple and general debiased machine learning theorem with finite-sample guarantees. *Biometrika*, 110(1):257–264.

Chetverikov, D., Santos, A., and Shaikh, A. M. (2018). The Econometrics of Shape Restrictions. *Annual Review of Economics*, 10:31–63.

Dong, Y. (2015). Regression Discontinuity Applications with Rounding Errors in the Running Variable. *Journal of Applied Econometrics*, 30(3):422–446.

Fan, J. and Gijbels, I. (2018). *Local Polynomial Modelling and its Applications*. Routledge.

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, 69(1):201–209.

Hastings, J. S. (2004). Vertical relationships and competition in retail gasoline markets: Empirical evidence from contract changes in southern california. *American Economic Review*, 94(1):317–328.

Heckman, J. J., Matzkin, R. L., and Nesheim, L. (2010). Nonparametric Identification and Estimation of Nonadditive Hedonic Models. *Econometrica*, 78(5):1569–1591.

Imbens, G. and Wager, S. (2019). Optimized Regression Discontinuity Designs. *Review of Economics and Statistics*, 101(2):264–278.

Kaboski, J. P. and Townsend, R. M. (2011). A structural evaluation of a large-scale quasi-experimental microfinance initiative. *Econometrica*, 79(5):1357–1406.

Kleven, H. J. and Waseem, M. (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from pakistan. *The Quarterly Journal of Economics*, 128(2):669–723.

Kolesár, M. and Rothe, C. (2018). Inference in Regression Discontinuity Designs with a Discrete Running Variable. *American Economic Review*, 108(8):2277–2304.

Kwon, K. and Kwon, S. (2020). Inference in Regression Discontinuity Designs under Monotonicity. Unpublished manuscript.

Lee, D. S. and Card, D. (2008). Regression Discontinuity Inference with Specification Error. *Journal of Econometrics*, 142(2):655–674.

Lee, D. S. and Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2):281–355.

Li, K.-C. (1989). Honest Confidence Regions for Nonparametric Regression. *The Annals of Statistics*, 17(3):1001–1008.

Manoli, D. and Weber, A. (2016). Nonparametric evidence on the effects of financial incentives on retirement decisions. *American Economic Journal: Economic Policy*, 8(4):160–182.

Matzkin, R. L. (1991). Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models. *Econometrica: Journal of the Econometric Society*, pages 1315–1327.

Matzkin, R. L. (1992). Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models. *Econometrica: Journal of the Econometric Society*, pages 239–270.

Matzkin, R. L. (1999). Computation of nonparametric concavity restricted estimators. Technical report. Unpublished manuscript.

Noack, C. and Rothe, C. (2021). Bias-Aware Inference in Fuzzy Regression Discontinuity Designs. Unpublished manuscript.

Oreopoulos, P. (2006). Estimating Average and Local Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter. *American Economic Review*, 96(1):152–175.

Petrin, A. and Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of marketing research*, 47(1):3–13.

Shigeoka, H. (2014). The Effect of Patient Cost Sharing on Utilization, Health, and Risk Protection. *American Economic Review*, 104(7):2152–84.

Todd, P. E. and Wolpin, K. I. (2006). Assessing the impact of a school subsidy program in mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *American Economic Review*, 96(5):1384–1417.

Todd, P. E. and Wolpin, K. I. (2023). The best of both worlds: combining randomized controlled trials with structural modeling. *Journal of Economic Literature*, 61(1):41–85.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Xu, K.-L. (2017). Regression Discontinuity with Categorical Outcomes. *Journal of Econometrics*, 201(1):1–18.