Beckert, Walter; Kaliski, Daniel

Working Paper

Honest inference for discrete outcomes

cemmap working paper, No. CWP67/19

Provided in Cooperation with:
The Institute for Fiscal Studies (IFS), London

Suggested Citation: Beckert, Walter; Kaliski, Daniel (2019) : Honest inference for discrete outcomes, cemmap working paper, No. CWP67/19, Centre for Microdata Methods and Practice (cemmap), London,
https://doi.org/10.1920/wp.cem.2019.6719

This Version is available at:
https://hdl.handle.net/10419/241870

cemmap
centre for microdata methods and practice

Honest inference for discrete outcomes

Walter Beckert
Daniel Kaliski

The Institute for Fiscal Studies
Department of Economics,
UCL

cemmap working paper CWP67/19

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

Honest Inference for Discrete Outcomes

Walter Beckert and Daniel Kaliski*

December 1, 2019

Abstract

We investigate the consequences of discreteness in the assignment variable in regression-discontinuity designs for cases where the outcome variable is itself discrete. We find that constructing confidence intervals that have the correct level of coverage in these cases is sensitive to the assumed distribution of unobserved heterogeneity. Since local linear estimators are improperly centered, a smaller variance for unobserved heterogeneity in discrete outcomes actually requires *larger* confidence intervals, since standard confidence intervals become narrower around a biased estimator, leading to a higher-than-nominal false positive rate. We provide a method for mapping structural assumptions regarding the distribution and variance of unobserved heterogeneity to the construction of "honest" confidence intervals that have the correct level of coverage. An application to retirement behavior reveals that the spike in retirement at age 62 in the United States can be reconciled with a wider range of values for the variance of unobserved heterogeneity (due to reservation wages or offers) than the spike at age 65.

1 Introduction

Most empirical researchers are aware that their inferences depend on unverifiable assumptions. Usually these assumptions can be written in terms of conditional independence between two random variables Y and Z so that $E[Y|Z] = 0$ with probability 1. It is less common for an inference

*Birkbeck, University of London. Corresponding author: d.kaliski@bbk.ac.uk. We are grateful to seminar participants at Birkbeck for comments and advice.

to be made conditional on some variable Y being Normally distributed, or that its variance has a particular magnitude. It is likely that this is because economic theory tends to make strong predictions about the signs of parameters of interest, but few predictions regarding their magnitude or the distribution of unobservables.

We nonetheless show that distributional assumptions regarding unobservables, and the magnitude of their variance, are necessary for inference in at least one common empirical setting: regression-discontinuity designs where the assignment variable is discrete and the outcome variable is also discrete.[1] The missing data problem arises from the impossibility of observing how close individuals are to the margin of choosing one option over another if the outcome is a discrete choice and that of observing individuals infinitely close to the cutoff in a regression-discontinuity setting where the running variable is discrete. In these cases, researchers need to specify both a prior distribution for the unobserved heterogeneity term, and a prior value for the variance of the unobserved heterogeneity, in order to obtain hypothesis tests that have the correct size.

The necessity of specifying the size of the variance term has an analogy in the literature on the distribution of reservation wages and extensive-margin labor supply elasticities (Attanasio, 2012). If the distribution of reservation wages is diffuse, then even a large exogenous change in the offered wage will result in a small mass of individuals moving in or out of employment. Since this distribution is not observed by the econometrician, it is implicit in a test of the null hypothesis that an observed mass of individuals moving into or out of employment at a discontinuous policy threshold is statistically significant. For a sufficiently concentrated distribution of reservation wages, even large movements in and out of employment could be present in the absence of the discontinuity, making it spurious to infer that a large change observed at the discontinuity is significant.

Formally, the importance of unobserved variability follows from the results of 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 the conditional expectation function. Armstrong and Kolesár (2018) moreover

[1] Papers that include both of these types of discreteness include, but are not limited to: Card et al. (2008), Card et al. (2009), Almond et al. (2010), and Shigeoka (2014).

show that any "data-driven" method for constructing confidence intervals necessarily implicitly assumes some value for K even if one isn't explicitly chosen. This formalizes the argument above that some notion of "usual variability" is implict in tests of the hypothesis that a discontinous change exists at some threshold. In discrete outcome models, what determines this usual variability are the distribution of unobserved heterogeneity and the functional forms present in the process that generates the outcomes. Hence the results in this paper, which show that choosing K in the context of discrete-outcome models is equivalent to choosing some set of distributional and functional-form assumptions that are consistent with prior evidence (even though economic theory typically provides little guidance on these questions).[2]

The necessity of assumptions regarding the distribution of unobserved heterogeneity has an earlier precedent in regression-discontinuity designs. Consistency of the regression-discontinuity estimator for the discontinuous change in mean outcomes at the cutoff for some assignment rule relies on the assumption that mean unobserved traits do not also discontinuously change at the cutoff (Lee and Lemieux, 2010). This motivated the McCrary test for whether individuals were sorting on unobservables around the cutoff, an important confounder if the assignment is based on a rule that can induce strategic behavior such as a minimum test score (McCrary, 2008). The insight of papers such as Lee and Card (2008) and Kolesár and Rothe (2018) is that even with this identifying assumption satisfied, misspecification of the true relationship between outcomes and assignment when the assignment variable is discrete can lead to confidence intervals that are inappropriately narrow. Our contribution is to connect an explicit distributional assumption regarding unobserved heterogeneity in the utility of alternative choices to construction of confidence intervals which have at least the nominal coverage probability when the outcome variable is also discrete.

Our results are analogous to the parallel literature on the necessity of distributional assumptions in identifying the taxable income elasticity (Blomquist et al., 2018). More broadly, this paper is in the tradition of studies of how quasi-experimental results depend on theoretical assumptions (Rosenzweig and Wolpin, 2000, Keane, 2010a,b, Wolpin, 2013). It is most closely connected with the

[2] A notable exception is the Roy Model of occupational choice. If individuals pursue their comparative advantage, and skills are log-concave, then earnings inequality is reduced relative to random assignment of individuals to occupations (Heckman and Honore, 1990). The Roy Model moreover makes testable predictions regarding the means and variances of truncations of the overall earnings distribution.

recent papers by Andrews et al. (2017, 2018), which examine how "sensitive" structural parameters are to moments estimated from data and how "informative" reduced-form results are for the identification of structural parameters respectively. Our paper asks the inverse of the former question: how sensitive is inference regarding a particular reduced-form result to the assumed underlying structure?

We answer this question in one much-studied context: the strength of retirement incentives at ages 62 and 65 in the United States. We estimate one parameter from a simple latent index model of retirement, and construct confidence intervals that are conditional on the assumed value for the other, unidentified parameter, the variance of unobserved heterogeneity. We find that the range of assumptions consistent with a statistically significant spike in retirement at age 62 is broader than the range consistent with a spike at age 65. Proceeding in this way allows researchers to commit to a single value of a parameter of the data-generating process when performing different analyses so that assumptions do not vary arbitrarily across analyses. It also allows researchers to analyse the sensitivity of their results to economically interpretable assumptions.

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. It also does not necessarily allow for the sensitivity analyses of the kind we perform in this paper, since if the variance-covariance matrix is not consistently estimated in a nonlinear model the estimators of the coefficients will also be inconsistent (Wooldridge, 2010).

The rest of this paper is organized as follows. Section 2 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, as well as our more general result for models where the index function is linear in the running variable, which finds a previously-unremarked general relationship between log-concave random variables' maximal first derivative and their variance.

Section 3 applies our method to a common empirical setting that admits a regression-discontinuity design with both a discrete running variable and a discrete outcome of interest. Section 4 concludes.

2 Theoretical Results

In this section, we first present some results for specific distributions of unobserved heterogeneity. Then we provide the more general result that underlies these specific results: for random variables with log-concave density functions, the maximum of the density function's first derivative is inversely proportional to the variance of the random variable.

2.1 Set-Up and Notation

We randomly sample N individuals, and for each $i \in \{1, ..., N\}$ we observe all the pairs of the outcome of interest, Y and some covariate R, (Y_i, R_i) for every i. We will be concerned with these joint observations of Y_i and R_i within an interval of length $2h$, where $h > 0$ defines the bandwidth, and there is a cutoff for assignment to the "treatment" group R^* (normalised to be 0 throughout, as is standard practice in the literature), so that we examine observations in the window $R_i \in [-h, h]$.

Denote the Conditional Expectation Function (CEF) $E[Y|R] = m(R)$. Suppose $D_i = 1$ if an individual is exposed to treatment, and 0 otherwise, with the potential outcomes of individual i being written $Y_i(D_i) = Y_i(1)$ for the treated state and $Y_i(0)$ for the untreated state. The object of interest is

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

We will assume that $m(R)$ is twice continuously differentiable. The following assumption has been shown (Armstrong and Kolesár, 2018, Kolesár and Rothe, 2018) to be important for the construction of confidence intervals that have the correct level of coverage for inference on τ (esp. in regression discontinuity designs):

Assumption A1. $|m''(R)| < K$ *for some* $K > 0$.

Let $\mathcal{M}$ denote the set of all $m(.)$ that satisfy this assumption. Assumption A1 is shown by Kolesár and Rothe (2018) to be sufficient for confidence intervals $C^{1-\alpha}$, where α is the nominal significance level, to have the property that

$$\liminf_{N\to\infty} \inf_{m\in\mathcal{M}} P_m(\tau \in C^{1-\alpha}) \geq 1-\alpha, \tag{2}$$

which is known as "honesty", with confidence intervals that satisfy this property being known as honest confidence intervals (Li et al., 1989). Confidence intervals that are honest by virtue of having been derived under Assumption A1 are called by Kolesár and Rothe (2018) "Bounded Second Derivative" confidence intervals. As they point out, honesty is a desirable property for confidence intervals to have, as "it guarantees good coverage properties even when facing 'the worst possible' CEF that still satisfies the postulated constraints" (Kolesár and Rothe, 2018). Details of how to construct confidence intervals for τ that are honest conditional on a pre-specified value for K can be found in Kolesár and Rothe (2018).

Now suppose that $Y \in \{0,1\}$ and whether $Y_i = 1$ or 0 is determined by the rule

$$Y_i = 1[g(R_i) \leq \eta_i] \tag{3}$$

where η is unobserved, and $g(R)$ is twice continuously differentiable. In this case

$$m(R_i) = E[Y_i|R_i] = F_\eta(g(R_i)) \tag{4}$$

where $F_\eta(.)$ denotes the CDF of η. The derivatives of $m(R_i)$ in this case are

$$m'(R_i) = f_\eta(g(R_i))g'(R_i), \tag{5}$$

$$m''(R_i) = f_\eta(g(R_i))g''(R_i) + f'_\eta(g(R_i))[g'(R_i)]^2, \tag{6}$$

so the assumption of interest becomes

$$|m''(R_i)| = |f_\eta(g(R_i))g''(R_i) + f'_\eta(g(R_i))[g'(R_i)]^2| < K, \tag{7}$$

To illustrate the consequences of this condition for some common distributional choices for η, and build intuition, we now consider some common distributional choices for η and their effect on this assumption. At the end of this section, we provide a general result for the case $g(R) = \delta R$: in this case, the condition reduces to

$$|f'_\eta(g(R_i))|\delta^2 < K, \tag{8}$$

and the general result is that the maximum value for $|f'_\eta(g(R_i))|$ is decreasing in the variance of η for log-concave random variables.

2.2 Uniform Unobserved Heterogeneity

Suppose $\eta \sim U(\omega_1, \omega_2)$. Then $f_\eta(g(R_i)) = \frac{1}{\omega_2 - \omega_1}$ for all R_i, and $f'_\eta(g(R_i)) = 0$ for all R_i. So we require that

$$|\frac{g''(R_i)}{\omega_2 - \omega_1}| < K, \tag{9}$$

which, using the formula for the variance of the Uniform distribution, can also be written

$$\frac{|g''(R_i)|}{2\sqrt{3}\sigma_\eta} < K \tag{10}$$

with σ_η denoting the standard deviation of η. This will provide a connection with the formulae for the distributions considered later: the less dispersion there is in the distribution of η, the harder it is to satisfy the condition under which the confidence intervals are constructed.

Consider the null hypothesis that is being tested in these cases: that there is no change (especially a discontinuous change) that could not have been produced by chance alone. The tighter is the distribution of unobserved heterogeneity η, the more common are fluctuations in $E[Y|R]$ since the mass of individuals close to their reservation utilities is greater at any one time (there are more

individuals "at the margin" for every value of R_i). In consequence, we need to observe a larger discontinuous change to reject the null, and the range of fluctuations that could be explained by chance expands. Hence, the confidence interval that has the correct level of coverage becomes wider for smaller values of σ_η.

If $g(.)$ is an affine function, then the confidence intervals are honest for any value of K. Though sufficient, this combination of Uniform η and affine $g(.)$ is not necessary, though it is plain that the necessary condition $f_\eta(g(R_i))g''(R_i) + f'_\eta(g(R_i))[g'(R_i)]^2 = 0$ requires a fortuitous coincidence of the shapes of $g(.)$ and the distribution of η. It follows that assuming confidence intervals have the correct level of coverage with a discrete outcome variable in these empirical settings amounts to implicitly making restrictive functional form and distributional assumptions.

2.3 Normal Unobserved Heterogeneity

Let the distribution of unobserved heterogeneity be distributed $N(0, \sigma_\eta^2)$. We then have $f_\eta(g(R_i)) = \sigma_\eta^{-1}\phi(\frac{g(R_i)}{\sigma_\eta})$ and $f'_\eta(g(R_i)) = \sigma_\eta^{-1}\phi(\frac{g(R_i)}{\sigma_\eta}) \times (\frac{-g(R_i)}{\sigma_\eta^2})$. $\phi(\frac{g(R_i)}{\sigma_\eta}) \leq 1$ since the Standard Normal density function is bounded from above by 1, so we obtain

$$|g''(R_i) - [g'(R_i)]^2 \frac{g(R_i)}{\sigma_\eta^2}| < K \tag{11}$$

if $g(.)$ is a linear function $g(R_i) = \delta R_i$ then this simplifies to

$$|\frac{\delta^3 R_i}{\sigma_\eta^2}| < K, \forall R_i \tag{12}$$

equivalently, $\max(0, \frac{\delta^3 R_i}{\sigma_\eta^2}) < K$.

2.4 Logistic Unobserved Heterogeneity

Let the distribution of unobserved heterogeneity be Type I extreme value with location $\mu = 0$ and scale $s_\eta = \frac{\sqrt{3}}{\pi}\sigma_\eta$. Then if g is linear, condition (4) becomes

$$\frac{2\sqrt{3}\delta^2}{\pi\sigma_\eta} < K \tag{13}$$

In all of the above cases, the variance of the unobserved heterogeneity plays a key role in inference. This seems to confirm that distributional assumptions are necessary for inference - and not just external validity, as is commonly assumed. In particular, we need some assumption regarding the variance of η conditional on a distributional assumption for η.

The next subsection presents the general result for random variables that have log-concave densities: the maximum of their first derivative is bounded from above by the inverse of their variance. This shows that the inverse relationship between K and σ_η that holds for the three specific distributions considered here is in fact a more general relationship that will hold for any log-concave random variable and linear $g(.)$.

2.5 General Result for Log-Concave Density Functions

The general result in this section says that if we rescale a log-concave random variable by a factor $\alpha > 1$, then the maximum of the first derivative of its probability density function will decrease. Accordingly, if a log-concave random variable's variance increases, all else equal, the maximum of the first derivative of its pdf will decrease. This has the consequence that if $g(.)$ is affine, the uniform bound K on the second derivative of the CEF is required to be larger for distributions of η for which the variance is smaller (since this will increase f'_η).

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. Consider a continuously distributed scalar random variable X with support $\mathcal{X} \subseteq \mathbb{R}$. Suppose the distribution of X has a density that is log-concave, i.e. it is of the form $f_X(x) = \exp(g(x))$, $x \in \mathcal{X}$, where $g(\cdot)$ is concave. The family of log-concave densities includes the exponential family, the uniform and also the extreme value distribution $\text{EV}(\mu, \sigma)$.

Log-concavity implies that the density $\exp(g(x))$ is unimodal.[3] The mode $x^\star$ satisfies $g'(x^\star) = 0$, provided $x^\star \in \text{int}(\mathcal{X})$.[4]

The sharpness of the peak of the density that corresponds to $x^\star$ is determined by $g''(x^\star) \leq 0$. Clearly, the location (parameter) of the density is irrelevant for the sharpness of the peak. The scale, however, matters. Consider $Z = \alpha X$, $\alpha \neq 0$. Then, Z has density $f_Z(z) = \frac{1}{\alpha}\exp(g(z/\alpha)) = \frac{1}{\alpha}f_X(z/\alpha)$; the mode $z^\star$ of the density of Z is $z^\star = \alpha x^\star$; and $f''_Z(z^\star) = \frac{1}{\alpha^2}f_X(z^\star/\alpha)g''(z^\star/\alpha) = \frac{1}{\alpha^2}f''_X(x^\star)$. Hence, the larger the scale $|\alpha|$, the flatter the peak.

In fact, for $|\alpha| > 1$, the density of X is more peaked about its mode $x^\star$ than the density of Z is about $z^\star$: For any $\tau \geq 0$,

$$\begin{aligned}
\int_{x^\star-\tau}^{x^\star+\tau} f_X(x)dx - \int_{z^\star-\tau}^{z^\star+\tau} f_Z(z)dz &= \int_{x^\star-\tau}^{x^\star+\tau} f_X(x)dx - \int_{z^\star-\tau}^{z^\star+\tau} \frac{1}{\alpha}f_X(z/\alpha)dz \\
&= \int_{x^\star-\tau}^{x^\star+\tau} f_X(x)dx - \int_{x^\star-\tau/\alpha}^{x^\star+\tau/\alpha} f_X(x)dx \\
&= \int_{x^\star-\tau}^{x^\star-\tau/\alpha} f_X(x)dx + \int_{x^\star+\tau/\alpha}^{x^\star+\tau} f_X(x)dx \\
&\geq 0.
\end{aligned}$$

Define $\check{x} = \arg\max_{x\in\mathcal{X}} f'_X(x) = \arg\max_{x\in\mathcal{X}} f_X(x)g'(x)$, and analogously for $\check{z}$. Then, $\check{z} = \alpha\check{x}$, and so

$$\begin{aligned}
f'_Z(\check{z}) &= \frac{1}{\alpha^2}\exp(g(\check{z}/\alpha))g'(\check{z}/\alpha) \\
&= \frac{1}{\alpha^2}\exp(g(\check{x}))g'(\check{x}) \\
&= \frac{1}{\alpha^2}f'_X(\check{x}),
\end{aligned}$$

and therefore $f'_Z(\check{z}) < f'_X(\check{x})$ for $|\alpha| > 1$. ■

Corollary. *If X and Z are log-concave random variables, identical except for $Var(Z) > Var(X)$, $\max f'_Z < \max f'_X$.*

[3] It is not necessarily symmetric; e.g. the extreme value distribution has $\mathcal{X} = \mathbb{R}$ and is not symmetric.

[4] Here and in what follows, $g'(x^\star) = 0$ is shorthand notation for $\frac{d}{dx}g(x)|_{x=x^\star} = 0$. Interiority places restrictions on the parameterization; e.g., for the Gamma density $f(x) \propto x^{\nu-1}\exp(-\frac{\nu}{\mu}x)$, it is required that $\nu > 1$ (and $\mu > 0$).

2.6 Consequences for Inference of Distributional Assumptions

Distributions of unobserved heterogeneity with higher variance allow for honest inference with lower values of K and hence smaller critical values than do those with lower variance. The intuition behind this result is both econometric and economic.

The econometric intuition is that while for a correctly centered estimator, a smaller standard error allows for more power with no sacrifice in the size of the test, for an improperly centered estimator, the distribution of the estimator becomes more tightly distributed around a value that deviates from the true value of the population parameter. This increases the number of false positives (equivalently, worsens the size of the test).

The economic intuition can be derived from the role of reservation wages in the estimation of extensive margin labor supply elasticities. If the unobserved distribution of reservation wages is diffuse, even large changes in wages do not lead to large movements in and out of the labor market. If the distribution is tightly concentrated, small changes in wage offers can produce large movements in and out of the labor market (Attanasio, 2012). Since the null hypothesis concerns the degree of discontinuity in outcomes that could be obtained by chance, the discreteness of the outcome variable necessitates some assumption regarding the closeness or otherwise of outcomes to being perturbed away from the discontinuity examined at the cutoff.

For a fixed variance, the Uniform distribution is significantly more forgiving than the more commonly used Normal or Logistic distributions. To take a simple example: suppose $R \in [-4, 4]$, fix $g(R) = 0.1R^2$, $\max_{R \in [-4,4]} |R| = 4$, and $\sigma_\eta = 1$. Then the minimal allowable K for Normal η is 0.824, and Logistic η it is 0.906, whereas for Uniformly distributed η it is 0.058. If instead we have $g(R) = 0.9R$, then any $K > 0$ bounds the second derivative of the CEF from above if η is Uniform, whereas we require $K > 2.916$ if η is Normal and $K > 0.893$ if η is Logistic. From this example it is clear that even for a given assumption regarding the variance of η, σ_η, the distribution η is assumed to follow can have a large influence on the size of the confidence intervals that are necessary for honest inference in the discrete outcome case. Whether the Logistic or Normal distributions allows for lower values of K seems likely to vary from case to case.

In the next section, we examine the consequences of different distributional assumptions for the conclusions of a recent empirical study that used both a discrete running variable and a discrete outcome variable in a regression-discontinuity design.

3 Empirical Application: Spikes in Retirement at Ages 62 and 65 in the United States

In this section, we use our mapping from assumptions regarding unobserved heterogeneity to confidence intervals to study retirement behaviour. In this context, unobserved heterogeneity will correspond to unobserved determinants of labor force participation that vary across individuals. Accordingly, whether the distribution of unobserved heterogeneity has low or high dispersion will correspond to whether the distribution of reservation wages has high or low dispersion.

3.1 Institutional Background

In the United States, Social Security is the main source of income in retirement for low-income individuals. Individuals become eligible for Social Security at age 62, producing a spike in retirement at 62 that is not typical of other countries (Gustman and Steinmeier, 2005). It is a matter of some debate as to whether the spike in retirement at age 65 that existed before the abolition of mandatory retirement in 1986 still remains (Von Wachter, 2002). Studies that examine the effect of qualifying for health insurance at age 65 via Medicare on United States residents' health outcomes and health behaviors check for the spike in retirement at this age as a possible confounder (Card et al., 2008, 2009, Dave and Kaestner, 2009, Kaliski, 2019), and typically find that from the late 90s onwards there is little evidence of a spike at age 65 (though the spike at 62 remains).

3.2 Data and Estimation

The data are drawn from the Health and Retirement Study (HRS), a nationally representative household survey run by the University of Michigan since 1992. Individuals are followed every two

years. New cohorts were added in 1998, the fourth wave of the survey, to increase the sample size and survey a broader range of cohorts in the study. We use a cleaned version of the HRS data informally titled the "RAND HRS" data set (for details of its construction, see Chien et al. (2013)).

Taking data on individuals' age in years, gender, and self-identified retirement status (dropping the partially retired, or those who say the question is irrelevant from the sample), we obtain 158326 nonmissing person-year observations from the 12 waves of the survey spanning the period 1992-2014. The "effective" sample size is typically much smaller than this since only observations within a narrow window of the cutoffs of 62 or 65 are used in estimation, but is larger than in many applications of the regression-discontinuity design, with around 3000 observations within a two-year radius of the cutoff per regression.

We use the `R` package `RDHonest` to compute confidence intervals according to the method described in Kolesár and Rothe (2018). The package is freely available online at the URL https://github.com/kolesarm/RDHonest/. We compute, for each value of $\sigma_\eta \in \{0.25, 0.5, 0.75, 1\}$, the associated implied minimal feasible K from the formulae in Equations 12 and 13.[5] We then estimate confidence intervals that have coverage rates of 95% given the assumed uniform bound on the second derivative of the conditional expectation function, K.

The estimating equation is the local linear specification

$$Y_i = \alpha + \tau \times 1[R_i \geq R^*] + \beta R_i + \gamma R_i \times 1[R_i \geq R^*] + \varepsilon_i, \frac{|R_i - R^*|}{h} < 1, \tag{14}$$

where $Y_i = 1$ if an individual responds that they are retired in the HRS data, and 0 otherwise, R_i is the respondent's age in years, R^* is one of $\{62, 65\}$ in each regression, $h > 0$ is the bandwidth and ε_i is a combination of unobserved heterogeneity and specification error. We pool the years of data together, ignoring period effects. Throughout we assume that away from the cutoff the data are generated by the process

$$Y_i = 1[-\delta R_i \leq \eta_i], \tag{15}$$

where $g(R_i) = -\delta R_i$, so that a higher estimated value for δ corresponds to a higher probability

[5] We do not include results for $\eta \sim U(\omega_1, \omega_2)$ since as discussed above when $g(.)$ is affine this allows for $K = 0$ regardless of the value of σ_η.

of $Y_i = 1$. The sign of δ is irrelevant in the formula for the minimal feasible K given in Equations 12 and 13.

3.3 Results

As a preliminary, we estimate the Logit and Probit coefficients on age in years that are relevant for calculating the value of K, the upper bound on the second derivative of the conditional expectation function. The results are displayed in Table 2. Men and women's age profiles within a model that has a linear index function do not seem to differ markedly, but as is to be expected the Logit and Probit coefficient estimates differ. In what follows we use $\widehat{\delta}_{probit}$ for δ in the calculations involving Normal unobserved heterogeneity and $\widehat{\delta}_{logit}$ for the calculations involving logistic unobserved heterogeneity. This allows us to produce 95% confidence intervals that vary with a single parameter, σ_η, the standard deviation of η, by mapping between σ_η, conditional on δ, and K, the uniform bound on the second derivative of the conditional expectation function (CEF). For comparison with the case where the researcher assumes a value for $\frac{\delta^2}{\sigma_\eta}$ or $\frac{|\delta^3|}{\sigma_\eta^2}$, the interested reader is referred to Table 1.

The main empirical results are displayed in Tables 3, 4, 5 and 6, which contain the results for men and women's self-identified retirement rates at ages 62 and 65, respectively. We vary the distribution and magnitude of σ_η and generate the confidence intervals corresponding to the minimal allowable value for K that is implied by the assumed σ_η, the relevant value for δ, and the $\max |R_i|$ chosen by the `RDHonest` package's chosen bandwidth, which is chosen so as to be optimal conditional on K. The values of σ_η are chosen from the set $\{0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 5\}$, with 5 providing an extreme case for contrast with the rest of the values ranging between 0.25 and 2.

Turning to Table 3, it appears that even as well-established an empirical regularity as the spike in retirement at age 62 for men in the U. S. (1) cannot be supported for all values of σ_η, given an assumed distribution, and (2) the choice of distribution implies different values of σ_η that are compatible with 95% confidence intervals that do not contain zero. For both men and women (Tables 3 and 5, the 95% confidence interval conditional on Normally distributed η does not contain zero for all $\sigma_\eta \geq 0.5$, whereas for Logistic η the higher values of $\sigma_\eta \geq 1$ for men and $\sigma_\eta \geq 1.25$

for women are required. It appears that in this relatively simple set-up, σ_η of at least 1.25 gives the most credible results, since it is consistently associated with a statistically significant spike in retirement for both men and women at age 62, the Social Security eligibility age.

The spike at age 65 requires the researcher to commit to a narrower range of assumptions regarding σ_η. From Tables 4 and 6, we can observe that the 95% confidence intervals only fail to contain zero with Normally distributed η and $\sigma_\eta = 5$, an order of magnitude greater than the value required to recover a statistically significant spike in retirement behavior for men at age 62.

When η is assumed to be Logistic, none of the values for σ_η, even the largest considered of $\sigma_\eta = 5$, produces a statistically significant spike in retirement for men or women. It follows that the spike in retirement at age 65 is furthermore more sensitive to implicit distributional assumptions than is the spike in retirement at age 62.

The results imply that statistical significance of the spike in retirement at age 62 is consistent with a broader range of assumptions regarding σ_η than is the spike in retirement at age 65. There are two observationally equivalent explanations for this outcome. The first is that the distribution of unobserved determinants of reservation wages and wage offers changes between 62 and 65 so that σ_η is larger at age 65 than it is at age 62. The other is that over the majority of the sample period the apparent spike in retirement at age 65 is an artifact rather than a truly discontinuous jump in retirement behavior.

The first explanation requires the individuals who choose to retire at age 62 to be drawn from a different part of the distribution of unobserved heterogeneity to those who choose to remain in the labor force. In the limit, if they are drawn at random from the distribution, σ_η should be unchanged between 62 and 65, absent especially strong age trends in the unobservables (if η is allowed to vary with age). If the distribution is skewed, and the mass of individuals are left after those in the tail of the distribution of η retire at 62, then σ_η can actually be *lower* at 65 than 62, making transitions in and out of the labor force more frequent between 62 and 70 than before age 62, raising the second derivative of the population conditional expectation function $E[Y|R]$.

The second explanation is simply that the evidence in favor of the spike in retirement at age 65, for an age-invariant σ_η, is weaker than for the spike at age 62, the Social Security eligibility age.

We leave which of these arguments is most plausible as a question for future research.

4 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. In doing so, it provides a test case where an explicit assumption regarding the magnitude of an unknown parameter is necessary for inference. This is in distinction with the majority of preceding literature, where the assumptions needed for point identification and inference typically require some population moments to be zero.[6] The best-known exception is the original regression-discontinuity design itself, which relies on the conditional distribution of the error term's continuity at the cutoff. Since McCrary (2008) a variant of this assumption - that there is no strategic sorting around the cutoff - has been testable. The standard RDD assumptions do not require any restrictions on the magnitudes of parameters governing the error term, nor anything regarding its distribution beyond continuity at a particular point. We have shown that when the running variable and outcome variable are both discrete, the apparently undemanding assumption of the continuity of unobserved outcomes at the cutoff R^* has to be supplemented by significantly stronger assumptions regarding the distribution of unobserved heterogeneity and the magnitude of its variance. We expect similar results to be formalized in the future for other settings where a quasi-experimental method is used in combination with a discrete dependent variable.

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.

[6] If one is willing to forego point identification, then moment inequalities will suffice for set identification of the population quantity of interest (Bontemps and Magnac, 2017).

Andrews, I., Gentzkow, M., and Shapiro, J. M. (2017). Measuring the Sensitivity of Parameter Estimates to Estimation Moments. *The Quarterly Journal of Economics*, 132(4):1553–1592.

Andrews, I., Gentzkow, M., and Shapiro, J. M. (2018). On the Informativeness of Descriptive Statistics for Structural Estimates. Technical report, National Bureau of Economic Research.

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

Attanasio, O. (2012). Comment on "Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities". In *NBER Macroeconomics Annual 2012, Volume 27*, pages 57–77. University of Chicago Press.

Blomquist, S., Kumar, A., Liang, C.-Y., and Newey, W. K. (2018). Identifying the Effect of Taxes on Taxable Income. Unpublished manuscript.

Bontemps, C. and Magnac, T. (2017). Set Identification, Moment Restrictions, and Inference. *Annual Review of Economics*, 9:103–129.

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.

Chien, S., Campbell, N., Hayden, O., et al. (2013). *RAND HRS Data Documentation, Version M.* Santa Monica, CA: RAND Center for the Study of Aging.

Dave, D. and Kaestner, R. (2009). Health Insurance and Ex Ante Moral Hazard: Evidence from Medicare. *International Journal of Health Care Finance and Economics*, 9(4):367.

Gustman, A. L. and Steinmeier, T. L. (2005). The Social Security Early Entitlement Age in a Structural Model of Retirement and Wealth. *Journal of public Economics*, 89(2-3):441–463.

Heckman, J. J. and Honore, B. E. (1990). The Empirical Content of the Roy Model. *Econometrica*, 58(5):1121–1149.

Kaliski, D. (2019). Does Insurance for Treatment Crowd Out Prevention? Evidence from Diabetics' Insulin Usage. Unpublished manuscript.

Keane, M. P. (2010a). A Structural Perspective on the Experimentalist School. *Journal of Economic Perspectives*, 24(2):47–58.

Keane, M. P. (2010b). Structural vs. Atheoretic Approaches to Econometrics. *Journal of Econometrics*, 156(1):3–20.

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

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. et al. (1989). Honest Confidence Regions for Nonparametric Regression. *The Annals of Statistics*, 17(3):1001–1008.

McCrary, J. (2008). Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of econometrics*, 142(2):698–714.

Rosenzweig, M. R. and Wolpin, K. I. (2000). Natural "Natural Experiments" in Economics. *Journal of Economic Literature*, 38(4):827–874.

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

Von Wachter, T. (2002). *The End of Mandatory Retirement in the US: Effects on Retirement and Implicit Contracts*. Center for Labor Economics, University of California, Berkeley.

Wolpin, K. I. (2013). *The Limits of Inference Without Theory*. MIT Press.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT press.

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

Tables and Figures

Table 1: Minimal Implied Bounds on Second Derivatives Given Distributions and Variances of Unobserved Heterogeneity, Assuming Linear Index Function

(1) Uniform	(2) K	(3) Normal	(4) K	(5) Logistic	(6) K
$\frac{g''}{2\sqrt{3}\sigma_\eta} = 0$	0	$\lvert\frac{\delta^3 R_i}{\sigma_\eta^2}\rvert = 0$	0	$\frac{\delta^2}{\sigma_\eta} = 0$	0
$\frac{g''}{2\sqrt{3}\sigma_\eta} = 0$	0	$\lvert\frac{\delta^3 R_i}{\sigma_\eta^2}\rvert = 0.25$	0.25	$\frac{\delta^2}{\sigma_\eta} = 0.25$	0.276
$\frac{g''}{2\sqrt{3}\sigma_\eta} = 0$	0	$\lvert\frac{\delta^3 R_i}{\sigma_\eta^2}\rvert = 0.5$	0.5	$\frac{\delta^2}{\sigma_\eta} = 0.5$	0.551
$\frac{g''}{2\sqrt{3}\sigma_\eta} = 0$	0	$\lvert\frac{\delta^3 R_i}{\sigma_\eta^2}\rvert = 0.75$	0.75	$\frac{\delta^2}{\sigma_\eta} = 0.75$	0.827
$\frac{g''}{2\sqrt{3}\sigma_\eta} = 0$	0	$\lvert\frac{\delta^3 R_i}{\sigma_\eta^2}\rvert = 1$	1	$\frac{\delta^2}{\sigma_\eta} = 1$	1.103
$\frac{g''}{2\sqrt{3}\sigma_\eta} = 0$	0	$\lvert\frac{\delta^3 R_i}{\sigma_\eta^2}\rvert = 1.25$	1.25	$\frac{\delta^2}{\sigma_\eta} = 1.25$	1.378
$\frac{g''}{2\sqrt{3}\sigma_\eta} = 0$	0	$\lvert\frac{\delta^3 R_i}{\sigma_\eta^2}\rvert = 1.5$	1.5	$\frac{\delta^2}{\sigma_\eta} = 1.5$	1.654

Notes: K refers to uniform bound on second derivative of the conditional expectation function (CEF) $E[Y|R]$, while δ and σ_η are the linear index parameter and standard deviation of unobserved heterogeneity η in a linear index model of discrete choice. R is the assigment variable and Y the outcome variable in a regression-discontinuity design.

Table 2: Probit and Logit Estimates of the Linear Index Parameter δ

(1)	(2)	(3)	(4)
Normal		Logistic	
Men	Women	Men	Women
0.128***	0.124***	0.233***	0.225***
(0.001)	(0.001)	(0.002)	(0.001)

Notes: Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: The Spike in Retirement at Age 62: Men

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Normal				Logistic		
σ_η	Implied K	95% CI		σ_η	Implied K	95% CI	
0.250	0.067	-0.006	0.224	0.250	0.239	-0.206	0.424
0.500	0.017	0.053	0.165	0.500	0.120	-0.067	0.285
0.750	0.007	0.061	0.146	0.750	0.080	-0.020	0.239
1.000	0.004	0.068	0.139	1.000	0.060	0.003	0.215
1.250	0.003	0.088	0.151	1.250	0.048	0.017	0.201
1.500	0.002	0.091	0.148	1.500	0.040	0.026	0.192
1.750	0.001	0.107	0.160	1.750	0.034	0.033	0.186
2.000	0.001	0.119	0.169	2.000	0.030	0.037	0.181
5.000	0.000	0.214	0.246	5.000	0.012	0.058	0.160

Notes: Formulae for mapping assumed values for σ_η to the uniform bound on the second derivative of the conditional expectation function K can be found in Equations 12 and 13. 95% confidence intervals are computed using the R package RDHonest, according to the method described in Kolesár and Rothe (2018).

Table 4: The Spike in Retirement at Age 65: Men

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Normal				Logistic		
σ_η	Implied K	95% CI		σ_η	Implied K	95% CI	
0.250	0.067	-0.084	0.153	0.250	0.239	-0.283	0.352
0.500	0.017	-0.026	0.095	0.500	0.120	-0.145	0.214
0.750	0.007	-0.026	0.065	0.750	0.080	-0.099	0.168
1.000	0.004	-0.065	0.013	1.000	0.060	-0.076	0.145
1.250	0.003	-0.060	0.008	1.250	0.048	-0.062	0.131
1.500	0.002	-0.043	0.018	1.500	0.040	-0.053	0.122
1.750	0.001	-0.041	0.016	1.750	0.034	-0.047	0.115
2.000	0.001	-0.020	0.033	2.000	0.030	-0.042	0.110
5.000	0.000	0.135	0.169	5.000	0.012	-0.036	0.074

Notes: Formulae for mapping assumed values for σ_η to the uniform bound on the second derivative of the conditional expectation function K can be found in Equations 12 and 13. 95% confidence intervals are computed using the R package RDHonest, according to the method described in Kolesár and Rothe (2018).

Table 5: The Spike in Retirement at Age 62: Women

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Normal				Logistic		
σ_η	Implied K	95% CI		σ_η	Implied K	95% CI	
0.250	0.061	-0.018	0.191	0.250	0.223	-0.206	0.378
0.500	0.015	0.035	0.138	0.500	0.112	-0.077	0.249
0.750	0.007	0.038	0.116	0.750	0.074	-0.034	0.206
1.000	0.004	0.044	0.110	1.000	0.056	-0.012	0.185
1.250	0.002	0.065	0.123	1.250	0.045	0.001	0.172
1.500	0.002	0.067	0.120	1.500	0.037	0.009	0.163
1.750	0.001	0.080	0.129	1.750	0.032	0.015	0.157
2.000	0.001	0.103	0.149	2.000	0.028	0.020	0.152
5.000	0.000	0.202	0.232	5.000	0.011	0.039	0.133

Notes: Formulae for mapping assumed values for σ_η to the uniform bound on the second derivative of the conditional expectation function K can be found in Equations 12 and 13. 95% confidence intervals are computed using the R package RDHonest, according to the method described in Kolesár and Rothe (2018).

Table 6: The Spike in Retirement at Age 65: Women

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Normal				Logistic		
σ_η	Implied K	95% CI		σ_η	Implied K	95% CI	
0.250	0.061	-0.108	0.110	0.250	0.223	-0.294	0.296
0.500	0.015	-0.056	0.058	0.500	0.112	-0.166	0.168
0.750	0.007	-0.020	0.065	0.750	0.074	-0.124	0.126
1.000	0.004	-0.052	0.020	1.000	0.056	-0.102	0.104
1.250	0.002	-0.047	0.015	1.250	0.045	-0.089	0.091
1.500	0.002	-0.032	0.025	1.500	0.037	-0.081	0.083
1.750	0.001	-0.030	0.023	1.750	0.032	-0.075	0.077
2.000	0.001	-0.014	0.036	2.000	0.028	-0.070	0.072
5.000	0.000	0.142	0.174	5.000	0.011	-0.030	0.074

Notes: Formulae for mapping assumed values for σ_η to the uniform bound on the second derivative of the conditional expectation function K can be found in Equations 12 and 13. 95% confidence intervals are computed using the R package RDHonest, according to the method described in Kolesár and Rothe (2018).