

# Regression Discontinuity Applications with Rounding Errors in the Running Variable\*

Yingying Dong  
University of California Irvine

First draft August 2011, revised January 2012

## Abstract

Many empirical applications of regression discontinuity (RD) models use a running variable that is rounded and hence is discrete, e.g., age in years, or birth weight in ounces. This paper shows that standard RD estimation using a rounded discrete running variable leads to inconsistent estimates of treatment effects, even when the true functional form relating the outcome and the running variable is known and is correctly specified. This paper provides simple formulas to correct for this discretization bias. The proposed approach does not require instrumental variables, but instead uses information regarding the distribution of rounding errors, which is easily obtained and often close to uniform. The proposed approach is applied to estimate the effect of Medicare on insurance coverage in the US, and to investigate the retirement-consumption puzzle in China, utilizing the Chinese mandatory retirement policy.

*JEL Codes:* C21, J26, I18

*Keywords:* Regression discontinuity, Rounding, Rounding errors, Discrete running variable

---

\*Correspondence: Department of Economics, 3151 Social Science Plaza, University of California Irvine, Irvine, CA 92697-5100, USA. Phone: (949)-824-4422. Email: [yyd@uci.edu](mailto:yyd@uci.edu). <http://yingyingdong.com/>.

# 1 Introduction

Regression discontinuity (RD) models identify local average treatment effects in cases where a treatment is determined by whether a continuous running variable  $X^*$  crosses a fixed threshold  $c$ . The RD treatment effect is given by the difference in the mean outcome of interest just above and just below the threshold. It can be obtained by separately estimating regressions of the outcome  $Y$  on  $X^*$  above and below the threshold point, and differencing the fitted value of these two regressions at the threshold (this can be equivalently done using a single regression that includes a dummy variable indicating whether  $X^*$  exceeds the threshold).

Many empirical applications of regression discontinuity (RD) methods use a running variable that is rounded and hence involves rounding or discretization errors. Examples include applications based on age in years, birth weight in ounces, an integer valued test score, calendar year or quarter etc. Age in years, calendar year, or similarly quarter involves rounding down to the nearest integer, whereas birth weight in ounces or an integer test score involves ordinary rounding, i.e., the true birth weight or test score is rounded either up or down to the nearest integer. Rounding issues exist frequently due to data availability. For example, many survey data sets only report an individual's age in years at the time of the survey. So in this case the reported age is an individual's actual age rounded down to the nearest integer.

As noted by Card and Lee (2008), when the observed running variable  $X$  is rounded and hence is discrete, it is impossible to compare outcomes for observations "just above" and "just below" the treatment threshold, and requires researchers to choose a functional form for the relationship between the treatment variable and the outcome of interest. Essentially, the RD treatment effect is not nonparametrically identified in this case. Standard practice is therefore to estimate parametric regressions of  $Y$  on reported age  $X$  (generally low order polynomial models; see, e.g., Card and Shore-Sheppard, 2004, DiNardo and Lee, 2004, Kane, 2003, and Lee, 2008) above and below the RD treatment threshold.

In this paper I use age in years as a leading motivating example and focus on the case of rounding down. I then extend the results to other typical forms of rounding. I first show that the standard method for dealing with a rounded discrete running variable leads to inconsistent estimates of the RD treatment effect, even if the true functional form relating the outcome to the running variable is known and is correctly specified. This inconsistency exists for the same reason that measurement errors in regressors yield biased and inconsistent estimates of regression coefficients, even when the functional form of the regression is correctly specified. In this case the observed rounded running variable can be taken as a mismeasure of the true exact running variable.

I next provide a formula for the size of the bias in the standard RD model estimators that use a rounded running variable based on rounding down, and describe the restrictive conditions under which this bias will be zero. For example, a sufficient condition for the rounding or discretization bias to be zero is that the slope and higher derivatives of the outcome as a function of age do not change at either side of the cutoff. Roughly, the more the slope and higher derivatives change at the cutoff, the greater will be the size of the bias due to rounding. The bias can be either positive or negative, depending on how the slope and higher derivatives change at the threshold.

I also show how to correct this rounding bias and thereby obtains consistent estimates of the true RD treatment effect. These corrections are very simple to implement in practice, and it is also simple to obtain standard errors for the bias corrected estimator of the treatment effect. These bias formulas and resulting corrections are not the same as the corrections for standard measurement error models.

When the running variable is only observed at discrete points, like age in years, what appears to be a discontinuity might actually only be a kink, i.e., a change in slope. For example, estimates based on rounded data may suggest a significant positive or negative treatment effect when the true treatment effect is zero. The reverse can also occur, that is, the true treatment effect could be large and significant while the estimate based on rounded discrete data is small and insignificant. The  $t$  statistic on the bias corrected estimator provides a valid test for the presence of a true discontinuity, and hence a nonzero true treatment effect.

A convenient feature of the proposed bias correction is that it does not require an instrument. It instead assumes that one has some information regarding the distribution of the rounding error within the discretized variable cell. In the case of age in years, it means that one can obtain the distribution of ages within a year, i.e., the distribution of birthdays for the underlying population the data are drawn from, which is readily available from census data. Alternatively, in most applications this distribution can be well approximated by a uniform distribution.

Two empirical applications of these results are provided. The first looks at the effect of Medicare eligibility on medical insurance coverage in the US. For this application, higher frequency (monthly versus yearly) age data are available. Estimates based on these monthly age data provide a benchmark. I show that the proposed methodology works well and produces estimates that are consistent with having and using data where age is more accurately measured. In particular, both the benchmark estimates and the bias corrected estimates imply that rounding leads to an overestimate of the impact of Medicare eligibility on the medical insurance coverage rate. The bias corrected estimates are similar to those estimates using monthly age data and departure further away from the uncorrected estimates based on age in years.

The second investigates the retirement-consumption puzzle in China. The puzzle (see, e.g., Banks, Blundell, & Tanner, 1998) refers to the empirical finding that, in many data sets of the developed Western countries, consumption (typically food consumption) drops significantly at retirement, which is inconsistent with consumption smoothing by the standard life cycle model. In China the official retirement age for male workers is 60. This mandatory retirement rule yields a significant jump in the retirement rate at age 60, which helps cleanly identify the impact of retirement on various outcomes of interest. In this paper, I focus on food consumption to investigate whether a similar retirement-consumption puzzle holds in China.

Since age is recorded in years in the available Chinese dataset, accurately estimating the retirement impact using RD models requires correcting for the rounding bias. I show that the slope of the food consumption profile changes substantially at the threshold age 60, resulting in a relatively large rounding bias. Ignoring this rounding bias produces overestimates of the impact of retirement on food consumption. I find that there is a significant drop in food consumption at retirement among male household heads, but the drop is not as large as one would calculate from a standard RD analysis that ignores the rounding bias. Overall, the empirical results in China are consistent with evidence from many developed Western countries.

I extend these results based on rounding down to cases involving non-integer cutoffs or other typical forms of rounding, such as ordinary rounding or rounding up to the nearest integer .

The rest of the paper proceeds as follows. Section 2 briefly reviews the literature. Section 3 provides the main identification result for the sharp design RD based on rounded and hence discrete data, as well as a numerically simple correction to the standard discrete data RD estimation. Section 4 describes how to estimate the true RD treatment effect with rounded data. Section 5 extends the approach to fuzzy design RD models. Sections 6 and 7 present two empirical applications. Sections 8 and 9 discuss some extensions to the basic setup. Short concluding remarks are provided in Section 10. Proof of the main result is provided in the Appendix.

## **2 Literature Review**

RD methods have been widely used in the treatment effect program evaluation literature in recent years. Recent surveys include Imbens and Lemieux (2008), and Lee and Lemieux (2010). This paper discusses dealing with rounding or discretization errors in RD running variables. I cast the discussion mostly in the context of RD applications based on age. Age is commonly used as a running variable in RD applica-

tions. Examples include Behaghel, Crepon, and Sedillot (2008), Card, Dobkin, and Maestas (2008), Card, Dobkin, and Maestas (2009), Carpenter and Dobkin (2009), Chen and van der Klaauw (2008), De Giorgi (2005), Edmonds (2004), Edmonds, Mammen, and Miller (2005), Ferreira (2007), Lalive, Van Ours, and Zweimller (2006), Lalive (2007), Lalive (2008), Lee and McCrary (2005), Lemieux and Milligan (2008), and Leuven and Oosterbeek (2004). This paper's main results may also be similarly applied to RD applications with calendar years as a running variable. For example, Oreopoulos (2006) uses birth year as a running variable to estimate returns to education in the UK. In addition, the proposed approach is readily extended to RD applications involving other types of rounding errors.

One paper that specifically considers discreteness of an RD running variable is Lee and Card (2008). They model deviations of the true regression function from a given approximating function, i.e., the chosen parametric (e.g., polynomial) function for the discrete running variable, as random specification errors. They then discuss the impact of these random specification errors on inference. In particular, they recommend use of cluster-consistent standard errors instead of conventional heteroskedasticity-consistent standard errors. They also give an additional correction that is needed when the specification errors for regression functions above and below the cutoff are not assumed to be identical.

As noted by Lee and Card (2008), when the random specification errors are not identical for the regressions above and below the cutoff, the true parameter of interest, i.e., the true RD treatment effect is not the same as the simple difference of the expected values of these two regressions at the cutoff. In particular, the true treatment effect equals the latter plus the expected difference between the two random specification errors at the cutoff.

If one interprets their specification errors as rounding errors due to discretization of the running variable, then it can be shown that the expected difference between these two rounding errors is not necessarily zero. In this paper, I show that under general conditions, this bias term can be identified and estimated, using information on the distribution of the rounding error within the discretized cell, e.g., the distribution of age (birthdays) within a year. One can then use this estimated bias to fix the naive RD estimator based on discretized age, that is, one can fix the bias in the simple difference of fitted values of the discrete data regressions above and below the cutoff evaluated at the cutoff.

The biases associated with rounding, or more generally the use of interval data, have been examined in the context of standard regression models. See, e.g., Manski and Tamer (2002) and references therein. In the regression context, this has traditionally been dealt with by setting observations equal to the midpoint of the interval, and assuming instruments are available that correlate with the underlying continuously distributed variable while being uncorrelated with the rounding error. In this way rounding is handled

in the same way as other measurement error. Alternatively, discretization or rounding leads to a loss of identification, and only bounds on model parameters are obtained as in Manski and Tamer (2002).

Other papers that consider treatment effect estimation in the presence of measurement errors in either covariates or indicators of treatment include Cochran and Rubin (1973), Mahajan (2006), Lewbel (2007), Hu (2008), Molinari (2008), Babanezhad et al. (2010), Hu and Lewbel (2010), Battistin and Sianesi (2011), and Battistin and Chesher (2011). None of these specifically handle rounding bias in RD models.

### 3 Identification

Begin with the standard RD set up. I consider sharp design RD first, and later provide the straightforward extension to fuzzy designs. Let  $T$  be the indicator of treatment, so an individual has  $T = 1$  if treated and  $T = 0$  otherwise. Let  $c$  denote the cut-off or threshold age for treatment, which is assumed to be an integer. Let  $X^*$  be an individual's exact age at the time of the survey minus  $c$ . Age  $X^*$  is the underlying continuous running or forcing variable.

Follow Rubin (1974) to let  $Y(1)$  and  $Y(0)$  denote an individual's potential outcomes of interest from being treated or not, respectively. The observed outcome  $Y$  can then be written as  $Y = Y(0) + [Y(1) - Y(0)]T$ . An individual's potential outcome can depend on  $X^*$ . Define conditional means of the potential outcomes conditioning on the true age as  $g_t(X^*) = E(Y(t) | X^*)$  for  $t = 0, 1$ . The conditional mean of the observed outcome is then  $E(Y | X^*, T) = g_0(X^*) + [g_1(X^*) - g_0(X^*)]T$ .

Define the dummy indicating crossing the threshold as  $T^* = I(X^* \geq 0)$ , where  $I(\cdot)$  is the indicator function that equals one if its argument is true and zero otherwise. Sharp design RD means that  $T = T^*$ , so for now, an individual is assumed to be treated if and only if  $X^* \geq 0$ , that is, when his true age equals or exceeds the cutoff  $c$ . It follows that when  $X^* \geq 0$ ,  $E(Y | X^*) = E(Y | X^*, T^* = 1) = g_1(X^*)$  and when  $X^* < 0$ ,  $E(Y | X^*) = E(Y | X^*, T^* = 0) = g_0(X^*)$ . By assuming continuity of  $g_0(X^*)$  and  $g_1(X^*)$  at the threshold point  $X^* = 0$ , the standard sharp design RD local average treatment effect is given by  $\tau = g_1(0) - g_0(0)$ . Outcomes could also depend on other covariates, which are suppressed for now. All the statements and theorems in this paper can be assumed to hold conditioning on the values of other covariates, though it should be noted that a generic virtue of the RD approach is that inclusion of other covariates generally only affects efficiency but not consistency of estimated RD treatment effects.<sup>1</sup>

---

<sup>1</sup>Given that  $T = T^*$  is a deterministic function of  $X^*$ , even though  $Y$  may depend on other observed covariates, if  $E(Y | X^*, T = t) = g_t(X^*)$  for  $t = 0, 1$  are nonparametric regression functions or correctly specified parametric conditional mean functions conditioning on  $X^*$  only, then  $E(Y | X^*) = E(Y | X^*, T) = g_0(X^*) + [g_1(X^*) - g_0(X^*)]T$  will

Let  $\widehat{g}_1(X^*)$  be a consistent estimator of  $g_1(X^*)$  for  $X^* \geq 0$ , obtained by regressing  $Y$  on  $X^*$  either nonparametrically, or by a correctly specified parametric model, using observations of data having  $X^* \geq 0$ . Similarly let  $\widehat{g}_0(X^*)$  be a consistent estimator of  $g_0(X^*)$  for  $X^* < 0$ . The sharp design RD treatment effect  $\tau$  is consistently estimated by  $\widehat{\tau} = \widehat{g}_1(0) - \widehat{g}_0(0)$ .

Now suppose one does not observe continuous  $X^*$ . Instead one observes  $X$ , defined as the largest integer less than or equal to  $X^*$ . That is,  $X$  is  $X^*$  rounded down and is an individual's reported age minus  $c$  at the time of the survey. This  $X$  is what is usually available in survey datasets, since surveys typically report individuals' age in (integer) years up to their most recent birthday at the time of the survey. The same analysis can be applied for finer or coarser reporting of age, e.g., if age is recorded in months, then the same analysis can apply by taking the units of  $X$  and  $X^*$  as measured by month rather than by year.

Suppose one were to ignore the fact that the reported data are rounded, and attempted to construct the standard RD estimator  $\widehat{\tau}$  as described above, but using the rounded discrete data  $X$ . Refer to this as the naive discrete data RD treatment effect estimator and denote it  $\widehat{\tau}'$ , of which the probability limit is denoted as  $\tau'$ . Call  $\tau'$  as the naive discrete data RD treatment effect. If  $\tau' = \tau$ , then the naive discrete data estimator  $\widehat{\tau}'$  is a consistent estimator of the true RD treatment effect, otherwise if  $\tau' \neq \tau$  then the discrete data RD treatment effect  $\tau'$  does not equal the true RD treatment effect, and so  $\widehat{\tau}'$  will be an inconsistent estimator of  $\tau$ . Define the bias in the discrete data RD treatment effect as  $\tau' - \tau$ .

Let  $h_t(X)$  be the discrete data analog of  $g_t(X^*)$  for  $t = 0, 1$ , so  $h_1(X) = E(Y | X, T^* = 1)$  when  $X \geq 0$  and that  $h_0(X) = E(Y | X, T^* = 0)$  when  $X < 0$ . By construction,  $E(Y | X, T) = E(Y | X, T^*) = h_0(X) + [h_1(X) - h_0(X)]T^*$ . Let  $\widehat{h}_1(X)$  be a consistent estimator of  $h_1(X)$  for  $X \geq 0$ , obtained by regressing  $Y$  on  $X$  in a correctly specified parametric model using observations of data having  $X \geq 0$ . Similarly let  $\widehat{h}_0(X)$  be a consistent estimator of  $h_0(X)$  for  $X < 0$ , obtained by regressing  $Y$  on  $X$  in a correctly specified parametric model using observations of data having  $X < 0$ .

Access to only rounded age data will make local nonparametric estimation in the neighborhood of the threshold impossible, because one does not observe data anywhere in the neighborhood of zero except at zero itself. It will therefore be necessary to assume parametric models for  $g_1(X^*)$  and  $g_0(X^*)$ . I will follow the usual practice in the literature of specifying these models as polynomials with possibly unknown but finite degrees. Given the parametric model for the true continuous age, I then derive the corresponding estimation equations for the rounded age. In particular it will follow from the assumptions below (as proven in Corollary 1) that the correct specifications for  $h_0(X)$  and  $h_1(X)$  will also be polynomials. When

---

be the correct conditional mean function and hence  $\tau = g_1(0) - g_0(0)$  will correctly identify the RD local average treatment effect at the cutoff.

estimating the model, standard covariate selection tests can be employed to determine the degree of these polynomials. I assume polynomials because they are commonly used in empirical practice, and because any sufficiently smooth (i.e., analytic) function can be approximated arbitrarily well by a polynomial.

The goal now is to provide assumptions that will allow one to identify and consistently estimate the true RD local treatment effect  $\tau$  using what one can identify from rounded data. Equivalently, these assumptions will permit one to quantify and correct for the bias in the discrete data RD treatment effect  $\tau'$ .

Let  $e = X^* - X$ , so  $e$  is the measurement error in the reported rounded age, and has  $0 \leq e < 1$ . Define  $\mu_k = E(e^k)$ . The following lists the identifying assumptions.

#### ASSUMPTIONS

A1:  $T = I(X^* \geq 0)$ .

A2:  $g_0(X^*)$  and  $g_1(X^*)$  are continuous at  $X^* = 0$ .

A3: The conditional mean functions  $g_0(X^*)$  for  $X^* < 0$  and  $g_1(X^*)$  for  $X^* \geq 0$  are polynomials of possibly unknown degree  $J$ .

A4:  $h_0(X)$  is identified for all  $-(J+1) \leq X < 0$ , and  $h_1(X)$  is identified for all  $0 \leq X \leq J$ .

A5:  $I(X \geq 0) = I(X^* \geq 0)$ .

A6: For all integers  $k \leq J$ ,  $E(e^k | X) = \mu_k$  where  $e = X^* - X$ , and these moments are identified.

Assumption A1 is the standard sharp design RD identifying assumption that treatment occurs if and only if age exceeds the threshold  $c$ , and hence when  $X^*$  crosses zero. Assumption A2 is the standard identifying assumption of RD design that the conditional means of the potential outcomes are continuous at  $X^* = 0$ , so the discontinuity in the observed conditional mean of  $Y$  at the threshold is attributable to the treatment.

Assumption A3 is a functional form restriction. As noted previously, unlike the standard RD, local nonparametric estimation is not possible with rounded discrete data, so estimation requires some assumed functional form. I choose polynomials because they are the most commonly used functional forms in empirical applications of RD with rounded data, and because any sufficiently smooth function can be approximated arbitrarily well by a polynomial. Note that Assumption A3 is an assumption only about observables, not counterfactuals, and so could be tested (using, e.g., a validation sample where true age  $X^*$  is reported). Assumption A3 can be easily extended to allow  $g_0(X^*)$  and  $g_1(X^*)$  to be polynomials of different degrees. Here for simplicity, I assume that they both are polynomials of  $J$ th degree. One can take  $J$  to be the maximum of the degree of these two polynomials.



Assumptions A4, A5 and A6 are the only assumptions that are imposed on the observed age  $X$  (as opposed to the true age  $X^*$ ). Assumption A4 says that one can identify the mean of  $Y$  in each observed age cell of  $X$ . So, for example,  $h_1(X)$  is just the mean of  $Y$  across everyone with reported age  $X$  above the threshold, that is, the mean outcome of everyone who's real age is between  $X$  and  $X + 1$ . If  $g_0(X^*)$  and  $g_1(X^*)$  are polynomials of different degrees  $J_0$  and  $J_1$ , then obviously  $h_0(X)$  and  $h_1(X)$  only need to be identified for  $-(J_0 + 1) \leq X < 0$  and for  $0 \leq X \leq J_1$ , respectively. In practice, they would be estimated simply by averaging  $Y$  across everyone with reported age  $X$ . There are always minimum and maximum ages at which sufficient numbers of people can be observed, which puts a limit on the degree  $J$  of the polynomials that could be identified in practice.

Assumption A5 says that the crossing threshold indicator  $T^*$ , when defined in terms of  $X$  rather than  $X^*$  is not mismeasured. This holds automatically for the type of rounding considered here, where the observed age is the true age rounded down to the nearest integer and the cutoff  $c$  is also an integer.

Assumption A6 says that the moments of the rounding error  $e$  do not depend upon  $X$ , and that these moments are identified. These are essentially moments of the distribution of birthdays within a year (among individuals in the survey population), so A6 will hold if birthdays are uniformly distributed. In this case  $\mu_k = \int_0^1 e^k de = 1/(k + 1)$  is known. Note that regardless of the distribution of  $e$ , all moments  $\mu_k$  are finite, because  $e$  is bounded between zero and one by construction.

There exists evidence of small but statistically significant seasonal departures from uniformity in the distribution of births within a year. See, e.g., Beresford (1980) or Murphy (1996). However, this seasonal variation appears to have very little impact on the lower order moments  $\mu_k$ . For example, Murphy (1996) provides birthdays of 480,040 life insurance applicants. The first four empirical moments  $\mu_1$  to  $\mu_4$  in his data are 0.506, 0.339, 0.254, and 0.203, which are numerically quite close the corresponding moments of a true uniform distribution, 0.500, 0.333, 0.250, and 0.200.

Small departures from uniformity could also arise among very old populations, where those with birthdays earlier in the year may be slightly underrepresented due to the higher mortality risk. However, even without assuming a uniform distribution, these distributions may be estimated using data from other sources such as a census, so it is not restrictive to assume that moments from these distributions are identified.

Recall that  $\tau = g_1(0) - g_0(0)$  is the true local RD treatment effect.

**THEOREM 1.** Let assumptions A1 to A6 hold. Then  $\tau$  is identified even if  $X^*$  is not observed.

Theorem 1 says that the above assumptions are sufficient to identify the true local RD treatment effect

$\tau$ . The only data these assumptions require are age cell means  $E(Y | X = x)$  and moments  $\mu_k = E(e^k)$ . Given just these data one can consistently estimate  $\tau$ .

Corollary 1 below describes the bias in the discrete data RD treatment effect  $\tau'$ , and provides a general method for constructing a consistent estimator for  $\tau$  using rounded age based on Theorem 1. The next section provides an example, showing that this method is very simple to apply in practice.

Write the true data model as  $Y^* = E(Y | X^*, T) + \varepsilon^* = g_0(X^*) + [g_1(X^*) - g_0(X^*)]T + \varepsilon^*$ , where  $\varepsilon^* = Y^* - E(Y | X^*, T)$ . Given Assumptions A1 and A3, this model reduces to

$$Y = \sum_{j=0}^J a_j X^{*j} + \sum_{j=0}^J b_j X^{*j} T^* + \varepsilon^* \quad (1)$$

where  $g_0(X^*) = \sum_{j=0}^J a_j X^{*j}$  and  $g_1(X^*) - g_0(X^*) = \sum_{j=0}^J b_j X^{*j}$ . Let  $A$  be the vector of unknown coefficients  $a_0, a_1, \dots, a_J$ , and let  $B$  be the vector of unknown coefficients  $b_0, b_1, \dots, b_J$ . The true treatment effect at  $X^* = 0$  is then  $\tau = b_0$ , which is the first element of  $B$  or the coefficient of  $T^*$  in equation (1), or the difference at the threshold of the intercepts of  $g_1(X^*)$  and  $g_0(X^*)$ . The second element of  $B$  (the coefficient  $b_1$ ), is the difference in slopes of  $g_1(X^*)$  and  $g_0(X^*)$  at the threshold  $X^* = 0$ , and similarly the other elements of  $B$  equal the differences in higher order derivatives.

It is shown in Corollary 1 below that the rounded data model  $Y = E(Y | X, T) + \varepsilon = h_0(X) + [h_1(X) - h_0(X)]T + \varepsilon$ , with  $\varepsilon = Y - E(Y | X, T)$ , has the same functional form as equation (1) but with different coefficients, so

$$Y = \sum_{j=0}^J d_j X^j + c_j X^j T^* + \varepsilon. \quad (2)$$

Let  $D$  be the vector of coefficients  $d_0, d_1, \dots, d_J$  of  $h_0(X) = \sum_{j=0}^J d_j X^j$  and  $C$  be the vector of coefficients  $c_0, c_1, \dots, c_J$  of  $h_1(X) - h_0(X) = \sum_{j=0}^J c_j X^j$ . Equation (2) is the regression model that can be estimated with the available rounded age data, and the naive discrete data RD treatment effect is then given by  $\tau' = c_0$ . Given equation (2), the identification and estimation problem is to recover  $b_0$ , and more generally all the coefficients  $A$  and  $B$  in equation (1), from the coefficients  $D$  and  $C$  in (2).

Let  $\binom{j}{k}$  denote the binomial coefficient  $\frac{j!}{k!(j-k)!}$ . Define the upper triangular  $J+1$  by  $J+1$  matrix  $M$  as having the element  $\binom{j}{k} \mu_{j-k}$  in row  $k+1$  and column  $j+1$  for all  $j, k$  satisfying  $J \geq j \geq k \geq 0$ , recalling that  $\mu_k = E(e^k)$ . All elements of  $M$  below the diagonal are zero. The proof of Theorem 1 shows that  $M$  is nonsingular.

**COROLLARY 1:** Let assumptions A1 to A6 hold. Then:

(i) Equation (2) holds, with  $D$  identified as the vector of coefficients of the polynomial  $h_0(x)$  that goes through the points  $x = -1, -2, \dots, -J, -(J+1)$ , and  $C$  is identified by  $C = C_1 - D$  where  $C_1$  is the vector of coefficients of the polynomial  $h_1(x)$  that goes through the points  $x = 0, 1, \dots, J$ .

- (ii) the coefficients in the true underlying model  $A$  and  $B$  are identified by  $A = M^{-1}D$  and  $B = M^{-1}C$ .
- (iii) the true treatment effect is  $\tau = b_0$ , the naive discrete data treatment effect is  $\tau' = c_0$ , and the bias is  $\tau' - \tau = \sum_{j=1}^J b_j \mu_j$ .

Corollary 1 shows that the rounded data model is itself a polynomial, and that the matrix  $M$  connects the polynomial coefficients  $D$  and  $C$  in the rounded data model  $h_0(X)$  and  $h_1(X)$  to the coefficients  $A$  and  $B$  in the true model  $g_0(X^*)$  and  $g_1(X^*)$  by  $MA = D$  and  $MB = C$ .

To illustrate the potential size of the rounding bias, and hence the size of the proposed correction, consider the case where the  $e$  distribution is uniform so  $\mu_k = 1/(k+1)$ . It then follows immediately from Corollary 1 that the bias in the discrete data estimator  $\tau'$  is

$$\tau' - \tau = \frac{1}{2}b_1 + \frac{1}{3}b_2 + \dots + \frac{1}{J+1}b_J.$$

As noted earlier, the coefficient  $b_1$  is the change in the slope of the conditional mean of the outcome  $E(Y | X^*)$  at the threshold. Similarly,  $b_j$  for  $j = 2, 3, \dots, J$  is the change in the  $j$ th derivatives of  $E(Y | X^*)$  at the threshold.

This bias formula shows that if the slope and higher derivatives of the outcome as a function of age do not change at the threshold, meaning that the treatment effect is locally constant, then the bias from rounding will be zero. Otherwise, the larger the changes in slope and higher derivatives are at the threshold, the larger the rounding bias tends to be. Note that the rounding bias can result in either an over- or an underestimate of the true RD treatment effect, depending on the changes in the polynomial coefficients crossing the threshold.

In particular, what appears to be a discontinuity in the rounded discrete data may not exist with continuous data. That is, the naive discrete data estimator  $\hat{\tau}' = \hat{c}_0$  can be significantly different from zero, while the true treatment effect is  $\tau = b_0$  is zero. Alternatively, the discrete data treatment effect estimator  $\hat{\tau}' = \hat{c}_0$  could be small even when the true treatment effect  $\tau = b_0$  is substantial. The next section describes how to estimate the true treatment effect from rounded discrete data, and test the significance of a true discontinuity.

The analysis in this section assumes that the threshold  $c$  is an integer, which means that by observing  $X$  instead of  $X^*$  one can still determine  $T^*$ , whether one is above or below the threshold without error. In particular,  $X$  is nonnegative if and only if  $X^*$  is nonnegative, so  $T^* = I(X^* \geq 0) = I(X \geq 0)$ . For example, in Card, Dobkin, and Maestas (2008) the treatment threshold is defined to be age 65 (the age of near universal Medicare eligibility), so their data correctly sorts individuals into those who are above 65

and hence are eligible for Medicare from those who are not, even though they only observe age in years. Note that when the threshold is not an integer,  $I(X^* \geq 0) = I(X \geq 0)$  may not hold. Cases like this are discussed in the extension Section 8.

These results can be extended immediately to applications involving other types of discretization or rounding, as long as they maintain this property of no mismeasurement in the crossing threshold indicator. Extensions to rounding involving mismeasurement of the crossing threshold status at the cutoff, i.e.,  $I(X^* \geq 0) \neq I(X \geq 0)$  are discussed in the extension Section 8.

## 4 Estimation

This section describes how to apply Theorem 1 and Corollary 1 to estimate the true treatment effect  $\tau$  with rounded data of the running variable. For simplicity, I first present the estimator for the case where the polynomial regressions are fourth (or lower) order polynomials, which should cover most actual empirical applications. I then describe the general estimation method for polynomials of any degree.

Let  $Y_i$  be the observed outcome for an individual  $i$ , and let  $X_i$  be  $i$ 's reported age in years subtracting off the threshold age  $c$ . Let  $T_i = T_i^* = 1$  for any individual who has  $X_i \geq 0$ , otherwise  $T_i = T_i^* = 0$ . Then the first step is to estimate the following polynomial model

$$Y_i = d_0 + d_1X_i + d_2X_i^2 + d_3X_i^3 + d_4X_i^4 + (c_0 + c_1X_i + c_2X_i^2 + c_3X_i^3 + c_4X_i^4)T_i^* + \varepsilon_i. \quad (3)$$

Equation (3) is essentially equation (2) with  $J = 4$ . Assuming all the assumptions hold conditioning on covariates, one can add other covariates to the model if desired.

The naive discrete data treatment effect  $\tau'$  will just be  $c_0$  in this regression. However, as shown below by applying Theorem 1 and Corollary 1, if the distribution of ages within a year (for individuals in the population from which the data are drawn) is uniform, then the correct treatment effect is given by

$$\tau = c_0 - \frac{1}{2}c_1 + \frac{1}{6}c_2 - \frac{1}{30}c_4. \quad (4)$$

More generally, the true treatment effect is

$$\tau = c_0 - \mu_1 c_1 + \left(2\mu_1^2 - \mu_2\right) c_2 + \left(-6\mu_1^3 + 6\mu_2\mu_1 - \mu_3\right) c_3 + \left(24\mu_1^4 - 36\mu_1^2\mu_2 + 8\mu_3\mu_1 + 6\mu_2^2 - \mu_4\right) c_4 \quad (5)$$

where  $\mu_j = E(e^j)$  for  $j = 1, \dots, 4$ , and the distribution of  $e$  is the distribution of ages within a year, on a scale of zero to one, where zero means a birth at the beginning of the first day of the year, and one means

a birth at the end of the last day. For lower order polynomial models these same formulas can be used, by just setting the higher order coefficients equal to zero.

Equation (4) or more generally (5) can be used to test for the presence of rounding bias, by applying an ordinary Wald t-test to the hypothesis that  $\tau - c_0 = 0$ . For example, when  $e = X^* - X$  has a uniform distribution and the polynomial is degree  $J \leq 4$ , by equation (4) one would just need to test if the linear combination of regression coefficients  $-c_1/2 + c_2/6 - c_4/30$  equals zero. A sufficient condition for no rounding bias when the polynomial  $J \leq 4$  is that  $c_1$ ,  $c_2$ , and  $c_4$  equal zero, so even if we did not know moments of the distribution of  $e$ , one could still do a standard F test of the joint hypothesis that these three regression coefficients equal zero.

More generally, for higher order polynomials, based on equation (2) one may estimate the regression

$$Y_i = \sum_{j=0}^J d_j X_i^j + \sum_{j=0}^J c_j X_i^j T_i^* + \varepsilon_i$$

by ordinary least squares. Then let  $C$  be the vector of coefficients  $c_0, c_1, \dots, c_J$ . Next construct the matrix  $M$  as the upper triangular  $J + 1$  by  $J + 1$  matrix having the element  $\binom{j}{k} \mu_{j-k}$  in row  $k + 1$  and column  $j + 1$  for all  $j, k$  satisfying  $0 \leq k \leq j \leq J$ , and having all elements below the diagonal be zero. For the special case in which  $e$  is uniformly distributed on the range zero to one, the element of  $M$  in row  $k + 1$  and column  $j + 1$  for all  $j, k$  satisfying  $0 \leq k \leq j \leq J$  will be

$$\binom{j}{k} \mu_{j-k} = \binom{j}{k} \frac{1}{j-k+1} = \frac{j}{k(j-k+1)}$$

while all other elements of  $M$  (those below the diagonal) are zero.

Recall that the regression of the outcome  $Y_i$  on the underlying continuous running variable  $X_i^*$  can be written as

$$Y_i = g_0(X_i^*) + [g_1(X_i^*) - g_0(X_i^*)] T_i^* + \varepsilon_i^* = \sum_{k=0}^J a_k X_i^{*k} + \sum_{j=0}^J b_j X_i^{*j} T_i^* + \varepsilon_i^*,$$

where the true treatment effect is given by  $\tau = g_1(0) - g_0(0) = b_0$ . Although one cannot estimate this regression, by Corollary 1,  $B = M^{-1}C$ , and so the estimate of  $\tau$  is given by the first element of  $M^{-1}C$ .

For example, if  $J = 4$  then  $B = M^{-1}C$  has the form

$$\begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} = \begin{pmatrix} 1 & \mu_1 & \mu_2 & \mu_3 & \mu_4 \\ 0 & 1 & 2\mu_1 & 3\mu_2 & 4\mu_3 \\ 0 & 0 & 1 & 3\mu_1 & 6\mu_2 \\ 0 & 0 & 0 & 1 & 4\mu_1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} \quad (6)$$

Notice that for each  $j = 0, \dots, J$ , the coefficients in column  $j + 1$  of  $M$  are just the binomial coefficients of  $(X + e)^j = \sum_{k=0}^j \binom{j}{k} X^{j-k} e^k$ . The true treatment effect  $\tau = b_0$  is given by the first element of the right-hand side of equation (6), which is equation (5).

For the case where  $e$  has a uniform distribution,  $\mu_k = 1/(k + 1)$  and  $M$  becomes

$$M = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & \frac{3}{2} & 2 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{so} \quad \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} = \begin{pmatrix} 1 & -\frac{1}{2} & \frac{1}{6} & 0 & -\frac{1}{30} \\ 0 & 1 & -1 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & -\frac{3}{2} & 1 \\ 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix}$$

which then gives equation (4).

## 5 Fuzzy Designs

Continue to let  $X_i^*$  equal the true continuous age minus  $c$  for individual  $i$ , let  $T_i$  be the indicator of whether person  $i$  is treated or not, and let  $T_i^* = I(X_i^* \geq 0)$  be the crossing threshold dummy. In the RD literature an individual  $i$  is defined to be a complier if for that person  $T_i = I(X_i^* \geq 0)$ , so a complier is treated if and only if his true age equals or exceeds the threshold. Sharp design RD assumes everyone is a complier. Fuzzy design RD relaxes this assumption, so unlike in the sharp design, it is no longer assumed that  $T_i = T_i^*$  for all individuals.

Still define  $g_t(X^*) = E(Y(t) | X^*)$  for  $t = 0, 1$ . Note that in this case  $g_t(X^*) \neq E(Y | X^*, T^* = t)$ .

Also define

$$\tau_Y = E(Y | X^* = 0, T^* = 1) - E(Y | X^* = 0, T^* = 0),$$

so in the special case of sharp design  $\tau_Y = g_1(0) - g_0(0)$ , which equals the RD treatment effect  $\tau$ .

Similarly define

$$\tau_T = E(T | X^* = 0, T^* = 1) - E(T | X^* = 0, T^* = 0),$$

so  $\tau_T = 1$  for sharp design RD, and  $0 < \tau_T < 1$  for fuzzy design RD.

Under well known standard conditions, the fuzzy design RD local treatment effect  $\tau_f$  is given by

$$\tau_f = \frac{\tau_Y}{\tau_T}, \quad (7)$$

where the numerator  $\tau_Y$  is the size of the jump or discontinuity in the mean outcome at the threshold, while the denominator  $\tau_T$  is the size of the jump in the probability of the treatment at the threshold. Both the numerator and the denominator may involve rounding bias.

The estimator of the previous sections can therefore be immediately extended to estimation of fuzzy design treatment effects. First apply the exact same estimator as in the sharp design case to obtain a consistent estimator  $\widehat{\tau}_Y$  of  $\tau_Y$ . Then replace  $Y$  with  $T$  and apply the exact same estimator again to obtain a consistent estimator  $\widehat{\tau}_T$  of  $\tau_T$ . The fuzzy design treatment effect estimator is then  $\widehat{\tau}_Y/\widehat{\tau}_T$ .

To illustrate the fuzzy design estimator, consider the case where the functional forms for  $E(Y | X^*, T^* = t)$  and  $E(T | X^*, T^* = t)$  are fourth order polynomials, so  $J = 4$ . Then one can use least squares to estimate the coefficients in the models

$$Y_i = d_0 + d_1X_i + d_2X_i^2 + d_3X_i^3 + d_4X_i^4 + \left(c_0 + c_1X_i + c_2X_i^2 + c_3X_i^3 + c_4X_i^4\right) T_i^* + \varepsilon_i. \quad (8)$$

and

$$T_i = r_0 + r_1X_i + r_2X_i^2 + r_3X_i^3 + r_4X_i^4 + \left(s_0 + s_1X_i + s_2X_i^2 + s_3X_i^3 + s_4X_i^4\right) T_i^* + \tilde{\varepsilon}_i. \quad (9)$$

Then, assuming a uniform distribution for  $e$ , the fuzzy design treatment effect will be given by

$$\tau_f = \frac{c_0 - \frac{1}{2}c_1 + \frac{1}{6}c_2 - \frac{1}{30}c_4}{s_0 - \frac{1}{2}s_1 + \frac{1}{6}s_2 - \frac{1}{30}s_4} \quad (10)$$

which corresponds to equation (4) in the numerator and in the denominator. This is in contrast to the incorrect discrete data treatment effect  $\tau'_f = \tau'_Y/\tau'_T = c_0/s_0$ . The order of the polynomials in these two equations need not be the same, and these fourth order polynomial formulas cover lower order polynomials as special cases by setting higher order coefficients equal to zero.

## 6 Medicare and Insurance Coverage

This section provides an empirical application examining the impact of qualifying for Medicare at age 65 on the health insurance coverage rate in the United States. Others have applied RD analyses to Medicare qualification (see, e.g., Card, Dobkin, and Maestas, 2008). I use this example because the presence of a discontinuity in the insurance rate due to the treatment (Medicare eligibility) is uncontroversial, and because the available data can be used to verify the accuracy of the proposed method for correcting rounding bias.

The data used are from the US Health and Retirement Study (HRS). The HRS is a national panel survey of individuals over age 50 and their spouses in the US. It has extensive information on health insurance, health, employment, and demographics etc. Data have been collected every two years since 1992. Nine waves of data have been released so far. The HRS is suitable, as it focuses on the elderly, covering an

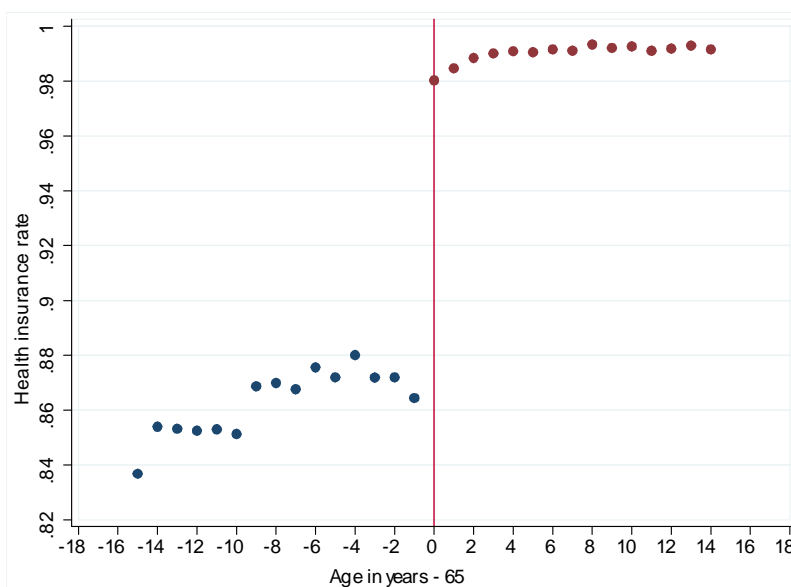


Figure 1: The age (in years) profile of health insurance coverage rates, HRS 1992 - 2008

age period particularly relevant to the empirical analysis here. More importantly, the HRS has available both age in months and age in years, so one can compare the bias corrected estimates based on yearly age data with estimates based on monthly age data, and thereby empirically evaluate how well the proposed correction works. I pool together all waves of data to have a sufficiently large sample. After observations with missing values deleted, the final samples have 60,290 to 135,582 observations, depending on the age ranges examined.

Let the outcome  $Y$  be the dummy indicating whether one has any health insurance, the cutoff  $c$  be age 65, and  $X$  be the reported age minus 65. The (sharp design) treatment  $T = T^*$  then corresponds to crossing age 65, and thereby becoming eligible for Medicare.

Figures 1 and 2 show the age profiles of health insurance rates, i.e., the age cell means of  $Y$  against age in years and in months, respectively. These figures clearly show a jump in insurance coverage at age 65. To model this treatment effect, I fit second, third, and fourth order polynomials to annual data. The quadratic model appears to underfit the model, while the fourth order polynomial tends to overfit especially for the narrower age ranges considered (both graphically and in terms of statistical significance of higher order terms as well as the adjusted  $\bar{R}^2$  of the regressions). I therefore focus on the third order polynomial as the preferred model (i.e., equation (3) with  $c_4$  and  $d_4$  set to zero), though estimates using both third and fourth order polynomials are reported for comparison. Attempts to include terms of degree five or more are completely insignificant and including these higher order terms does improve overall fit of the model.



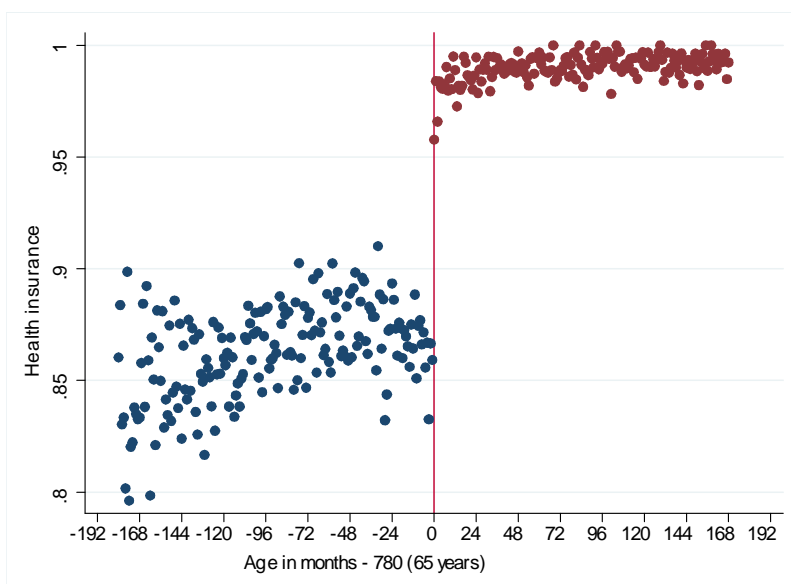


Figure 2: The age (in months) profile of health insurance coverage rates, HRS 1992 - 2008

In practice, there is a tradeoff regarding how many years of data around the threshold to include in the model. More years provides more observations, thereby adding to the precision with which the model coefficients can be estimated. However, the further included ages are from the threshold, the more likely it is that the correct model specification for these distant observations will differ from the correct specification near the threshold, risking specification errors. I consider four ranges of data, specifically, 6, 9, 12, and 15 years below and above the threshold, corresponding to age ranges 59 - 70, 56 - 73, 53 - 76, and 50 - 79. Note that the smallest window width here is less than half the largest window width.

Another specification issue is inclusion of covariates. To assess the impact of covariates, I estimate models that include year of survey dummies with or without additional demographic characteristics such as gender, race (white/non-white), ethnicity (hispanic/non-hispanic), and education levels. Three education levels represent less than high school (the default), high school or GED, and college or above.

The results are reported in Table 1. For each specification, the top panel in Table 1 presents the naive discrete data estimates corresponding to  $\tau' = c_0$ , and the bottom panel represents the bias corrected estimates corresponding to  $\tau$  in equation (4). Estimates based on monthly age data are reported in the middle panel. Standard errors for the corrected estimator are obtained by the delta method, but it would alternatively be numerically trivial to obtain standard errors by bootstrapping the data. All of the reported estimates are statistically significant at the 1% level.

For the preferred third order polynomial, the results are similar across different specifications, i.e.,

Table 1 Estimated increases in the health insurance coverage rate at the Medicare eligibility age 65

	(I)			(II)		
	(1)	(2)	(3)	(1)	(2)	(3)
[-6, +6)	0.128 (0.013)***	0.124 (0.013)***	0.125 (0.013)***	0.107 (0.031)***	0.106 (0.030)***	0.107 (0.031)***
[-9, +9)	0.128 (0.009)***	0.127 (0.008)***	0.128 (0.009)***	0.124 (0.014)***	0.119 (0.014)***	0.121 (0.014)***
[-12, +12)	0.126 (0.007)***	0.126 (0.007)***	0.127 (0.007)***	0.120 (0.010)***	0.119 (0.010)***	0.119 (0.010)***
[-15, +15)	0.124 (0.006)***	0.126 (0.006)***	0.126 (0.006)***	0.129 (0.008)***	0.128 (0.008)***	0.129 (0.008)***
[-6, +6)	0.119 (0.009)***	0.118 (0.009)***	0.119 (0.009)***	0.112 (0.011)***	0.113 (0.010)***	0.113 (0.011)***
[-9, +9)	0.119 (0.008)***	0.120 (0.008)***	0.120 (0.008)***	0.116 (0.010)***	0.115 (0.009)***	0.115 (0.010)***
[-12, +12)	0.119 (0.006)***	0.120 (0.006)***	0.121 (0.006)***	0.116 (0.008)***	0.117 (0.008)***	0.116 (0.008)***
[-15, +15)	0.119 (0.006)***	0.121 (0.006)***	0.120 (0.006)***	0.119 (0.007)***	0.120 (0.007)***	0.120 (0.007)***
[-6, +6)	0.118 (0.009)***	0.117 (0.009)***	0.117 (0.009)***	0.112 (0.012)***	0.113 (0.012)***	0.113 (0.012)***
[-9, +9)	0.117 (0.007)***	0.117 (0.007)***	0.118 (0.007)***	0.116 (0.009)***	0.115 (0.009)***	0.116 (0.009)***
[-12, +12)	0.118 (0.006)***	0.119 (0.006)***	0.119 (0.006)***	0.114 (0.008)***	0.113 (0.008)***	0.114 (0.008)***
[-15, +15)	0.117 (0.006)***	0.118 (0.005)***	0.119 (0.006)***	0.119 (0.007)***	0.120 (0.007)***	0.120 (0.007)***

Note: (I) 3rd order polynomial, (II) 4th order polynomial; (1) does not include additional covariates; (2) controls for year dummies; (3) controls for year dummies and additional demographic variables. Top panel, estimates based on age in years; Middle panel, estimates based on age in months; Bottom panel, discretization bias-corrected estimates. Standard errors are in parentheses; \*Significant at the 10% level; \*\* Significant at the 5% level; \*\*\*Significant at the 1% level.

controlling for different covariates or using different ranges of data. Note that the monthly data estimates are systematically smaller than the yearly data estimates, which implies that rounding of age by years in this data results in an overestimate of the impact of the Medicare program. In contrast, the bias corrected annual data estimates are all close to, and slightly smaller than the monthly data estimates. So going from annual to monthly data appears to correct most but not all of the bias associated with rounding. This is what one would expect if the proposed model and bias correcting methodology are valid.

Across the annual data models, the discrete data treatment effect  $\tau'$  is estimated to be in the range of 0.124 to 0.128. These numbers represent the estimated increase in the fraction of individuals possessing health insurance as a result of qualifying for Medicare. In contrast, the estimates based on age in months are on average about 5% lower, in the range of 0.118 to 0.121. The bias corrected estimates are in a similar range of 0.117 to 0.119, averaging about 6% lower than estimates using yearly age data.

In this application, the bias due to rounding, about 6%, is relatively small in percentage terms. This is not surprising given the fact that the slope of  $E(Y | X)$ , the health insurance profile, changes very little at the threshold (as clearly shown by figures 1 and 2, and by the coefficient estimates). As a result, the leading term  $c_1$  in the correction expression  $\tau' - \tau = -\frac{1}{2}c_1 + \frac{1}{6}c_2 - \frac{1}{30}c_4$ , is quite small. In absolute terms, failing to correct for the rounding bias results in an overestimate of insurance coverage of 0.7 to 1.0 percent of the relevant population. The current population of the US that is over age 65 and hence qualifies for Medicare is approximately 38 million (according to the US Census), so even half of one percent of this total is a large number of people.

## 7 The Retirement Consumption Puzzle in China

This section applies the proposed approach to investigating consumption changes around retirement in China. Standard Life cycle models suggest that rational people smooth consumption over the life cycle, which implies that consumption should not change at retirement when retirement is expected. However, many empirical studies find that consumption (typically food consumption) drops significantly at retirement. This finding is referred to as the "retirement-consumption puzzle."

Evidence of this puzzle has been mostly obtained from developed Western countries, including the United Kingdom (Banks, Blundell, & Tanner, 1998), the United States (Bernheim, Skinner, and Weinberg, 2001; Aguila, Attanasio, and Meghir, forthcoming; Ameriks, Caplin, and Leahy, 2007; Haider and Stephens, 2007; Hurd and Rohwedder, 2008), Canada (Robb and Burbridge, 1989), Germany (Schwerdt, 2005), and Italy (Battistin, Brugiavini, Rettore and Weber, 2010; Miniaci, Monfardini, and Weber, 2009;

Borella, Moscarola, and Rossi, 2011). Evidence from developing countries is scanty.

Most analyses of the retirement-consumption puzzle depend on structural models. One exception is Battistin, Brugiavini, Rettore and Weber (2009), who estimate RD models that exploit pension eligibility rules in Italy.

This section conducts an RD analysis of the retirement-consumption puzzle in China, taking advantage of the Chinese mandatory retirement rule. Since age is reported in years in the dataset used here, I apply the proposed approach to correct the associated rounding bias. The Chinese case is interesting due to its unique social and cultural environment, which differs in many ways from developed Western countries, and hence may shed additional light on the underlying mechanism of consumption changes around retirement. One way China differs from these other countries is that China has very high savings rates, so most households may have saved enough to avoid significant drops in consumption at retirement. In addition, cash transfers from adult children to retirees are a common practice in urban China. This may also help to prevent consumption declines that would otherwise result from inadequate accumulated wealth.

In China, the official retirement age is 60 for male workers, 55 for white-collar female workers, and 50 for blue-collar female workers, with some exceptions applying to certain occupations and to disabled workers.<sup>2</sup> These mandatory retirement ages have not changed ever since the retirement system was founded in the 1950s. Compared with pension eligibility rules, the mandatory retirement policy in China may induce a sharper change in the retirement probability and hence helps more precisely identify the causal impact of retirement on outcomes of interest.

The analysis here focuses on male workers, because female workers' labor supply is more complicated and their mandatory retirement age depends on the types of their work. I look at food consumption, as food is nondurable and is more likely than other consumption categories to change immediately at retirement. The sample includes all urban male household heads who are labor force participants, so, for example, homemakers are not included. Some workers may retire earlier than the mandatory retirement age, and some may be re-employed after the official retirement. Also, the mandatory retirement policy may not be strictly enforced in the private sector compared to the state sector, including the state-owned enterprises (SOE's) and the government units. As a result, the change in the retirement rate is less than one at 60, which entails fuzzy design RD models.

---

<sup>2</sup>Those who have jobs that are risky, harmful to their health, or extremely physically demanding can retire 5 years before the official retirement ages, i.e., 45 for blue-collar female workers and 55 for male workers. Male workers who become disabled and hence are unable to do their work can apply to retire at 50, while disabled female workers can retire at 45. Civil servants also qualify for early retirement if they have worked for 30 years and are within 5 years of their retirement age.

Data in this analysis are from the China Urban Household Survey (UHS), which are collected by the National Bureau of Statistics (NBS) every year to monitor consumption in China and to construct consumer price index (CPI). Complete data from five provinces and one municipality are used.<sup>3</sup> The pension system in China changed in 1997. In particular, the Chinese government adopted a combination of individual accounts and social pooling as the uniform pension system in 1997. Before that, the welfare of urban retirees was determined and provided entirely by their employers. To ensure that all retirees in the sample are subject to similar pension rules, this analysis is based on data from 1997 to 2006.

Typically, eligible male workers can start their retirement application at the beginning of the month they turn 60, and it may take more than one months to have the application approved, so a worker's true age at retirement  $X^*$  is typically between 60 and 61. Also, in the UHS dataset used here, the recorded age (in years)  $X$  and retirement status  $T$  are determined at the end of the survey year.<sup>4</sup> As a result, those who follow the retirement rule and start to apply to retire when they turn 60 may not have their retirement finalized until their recorded age is 61. Consistent with these observations, the data show that the retirement rate jumps significantly at both 60 and again at 61. I therefore exclude the cutoff age 60 in the estimation, assuming that the retirement change induced by the mandatory retirement policy is fully realized at 61. This ensures that all individuals who are observed below the cut off age of 60 are drawn from the pre mandatory retirement age profile  $h_0(X)$ , and all the individuals who are observed above the cut off age are drawn from the post mandatory retirement age profile  $h_1(X)$ . I then estimate the polynomial models using data from ages 59 and below and ages 61 and above, and evaluate changes at 60 by extrapolating these regression curves to the cutoff age of 60.<sup>5</sup>

Figures 3 and 4 show the age profiles (age cell means) of the retirement rate and the logarithm of household food consumption. The retirement profile shows an obvious jump and a mild slope change crossing the retirement cutoff age 60 (normalized to 0 in the figures). The jump represents an exogenous change in the retirement rate induced by the retirement policy and provides identification of the retirement impact on food consumption. The modest slope change implies that rounding bias in the retirement rate change may not be zero, but is possibly rather small.

In contrast to the retirement profile, the food consumption profile shows an obvious drop crossing the

---

<sup>3</sup>The five provinces are Liaoning, Zhejiang, Guangdong Shanxi, Sichuan, and the one city is Beijing.

<sup>4</sup>In particular, for years 2002 - 2006, employment information was collected in the first month of the survey year and updated every month for any changes, so the final recorded retirement status reflects the status at the end of the year.

<sup>5</sup>Note that this is somewhat similar to the alternative cases of discretization discussed in the extension section 7. The similarity is that observations at 60 do not uniquely belong to either the pre- or the post- cutoff outcome profile, i.e.,  $h_0(X)$  or  $h_1(X)$ .

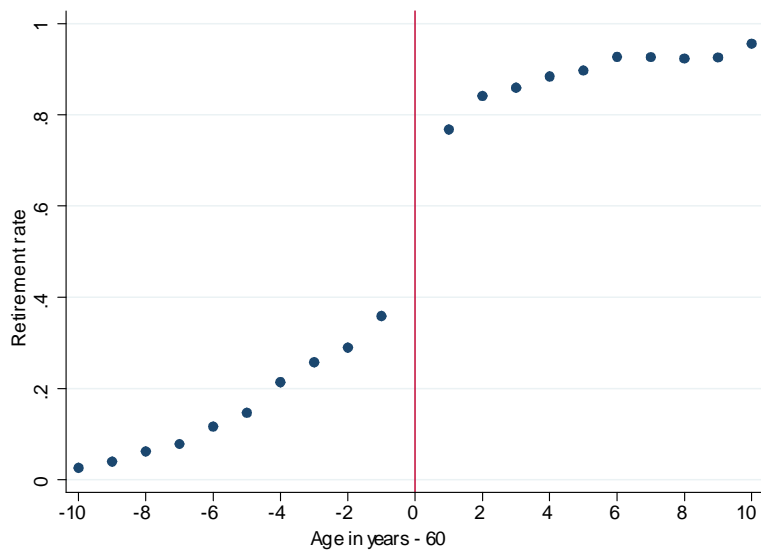


Figure 3: The age (in years) profile of retirement rates for male household heads, UHS 1997 - 2006

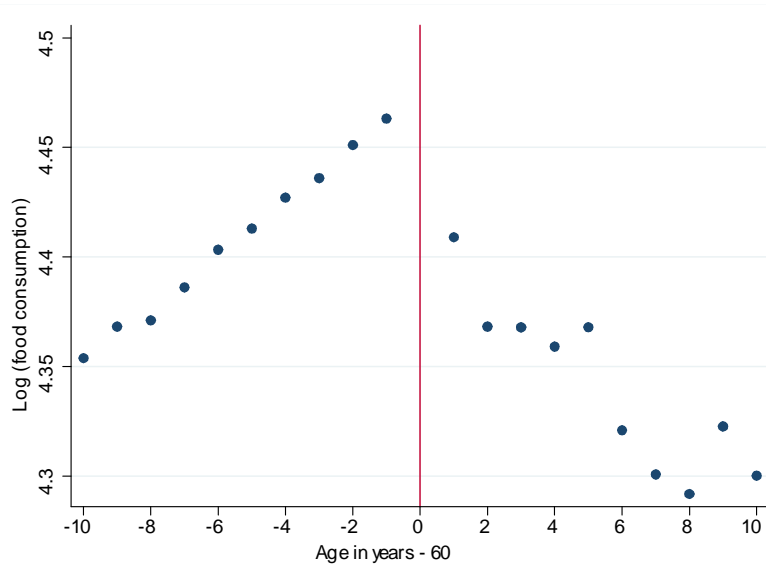


Figure 4: The age (in years) profile of log (food consumption), UHS 1997 - 2006

threshold age, along with a substantial change in slope. Before 60, food consumption increases steadily with age, while after 60 food consumption declines rapidly. This large change in slope means that one should expect substantial rounding bias in the estimated change in food consumption at the retirement age of 60.

Define the outcome  $Y$  as the logarithm of household food expenditure. Let  $T$  be a dummy indicating whether a household head has retired or not. Let  $c$  be the cutoff age 60, and  $X$  be the recorded age in years minus 60. Log food consumption  $Y$  and retirement  $T$  are specified as polynomial models as in equations (8) and (9). These models are estimated using three different window widths, i.e., 6, 10, and 15 years above and below the cutoff. The sample sizes corresponding to the three window widths are 12,634, 22,628, and 34,133, respectively. In particular, linear models (setting  $d_p$  and  $c_p$  for  $p = 2, 3, 4$  in equation (8) to zero) are adopted for log food consumption, while third order polynomials (setting  $r_4$  and  $s_4$  in equation (9) to zero) are used for the retirement rate, except that when using the short 6 years window, a quadratic model is adopted in that case.<sup>6</sup> These polynomial orders are chosen based on goodness of fit measures and significance of the coefficients on higher order terms.

The estimation results are reported in Table 2. Bootstrapped standard errors are reported for the estimated retirement effects on food consumption ((a)/(b) in Table 2). The naive biased and corrected retirement effects in this case are given by  $\tau'_f = c_0/s_0$  and  $\tau_f = (c_0 - \frac{1}{2}c_1) / (s_0 - \frac{1}{2}s_1 + \frac{1}{6}s_2)$ , respectively. I also try controlling for different covariates. The estimates on the left side of Table 2 (noted as (1) in the Table) control for survey year dummies, family size, family size squared, and education levels, including college or above, high school, and less than high school (the default). As a comparison, the estimates on the right half of the table (noted as (2) in Table 2) control only for year dummies.

The preferred specification is the one that uses data 10 years above and below the cutoff age 60 and controls for the full set of covariates as discussed above. Note that household food consumption crucially depends on family size and permanent income (proxied by education levels here), so including these covariates helps reduce a large fraction of the sample variation in log food consumption, and hence provides estimates that are more precise and more robust to variation in window widths.

The top and the bottom panels in Table 2 present the biased and the corrected estimates, respectively. For the specification controlling for year dummies along with other covariates, the uncorrected estimates of the retirement effect range from -0.165 to -0.186, representing a drop of 16.5% to 18.6% in food consumption at retirement among those male workers who retire due to the mandatory retirement policy.

---

<sup>6</sup>Adopting a quadratic model for the retirement equation in this case means also setting  $s_3$  to zero.

In contrast, the bias corrected estimates indicate a smaller 13.0% to 15.9% drop in food consumption at retirement, so failing to account for rounding of age leads to an average overestimate of the impact of retirement on food consumption by 3 percentage points.

Put it differently, bias correction results in roughly a decrease of 14% to 21% in the estimated retirement effects on food consumption. For the alternative specification when only controlling for year dummies, RD estimates that do not correct for rounding bias indicate a consumption drop of 14.2% to 19.3%, while after correcting for the bias the estimated drop is 10.5 % to 15.6%. Again, the difference is about 3 percentage points on average, representing a decrease of 16% to 26% in the estimated effects of retirement.

Table 2 Effects of retirement on food consumption at the mandatory retirement age 60

	(1)			(2)		
	(a)	(b)	(a)/(b)	(a)	(b)	(a)/(b)
[-6,+6]	-0.053 (0.020)***	0.320 (0.030)***	-0.166 (0.064)**	-0.061 (0.021)***	0.318 (0.031)***	-0.193 (0.068)***
[-10,+10]	-0.055 (0.014)***	0.295 (0.028)***	-0.186 (0.055)***	-0.055 (0.015)***	0.293 (0.028)***	0.188 (0.059)***
[-15,+15]	-0.048 (0.012)***	0.297 (0.020)***	-0.165 (0.04)***	-0.042 (0.012)***	0.296 (0.017)***	0.142 (0.042)***
[-6,+6]	-0.041 (0.020)***	0.317 (0.033)***	-0.130 (0.67)*	-0.049 (0.021)***	0.315 (0.033)***	0.156 (0.071)**
[-10,+10]	-0.044 (0.015)***	0.279 (0.032)***	-0.159 (0.059)***	-0.048 (0.015)***	0.276 (0.032)***	-0.158 (0.062)**
[-15,+15]	-0.039 (0.012)***	0.301 (0.018)***	-0.131 (0.04)***	-0.032 (0.012)***	0.304 (0.018)***	-0.105 (0.042)**

Note: (a), change in log food consumption at 65; (b), change in the retirement rate at 65; (a)/(b) effect of retirement on food consumption. (1) controls for year dummies family size, family size squared, and education levels; (2) only controls for year dummies. Standard errors are in the parentheses; \*significant at the 10% level; \*\* significant at the 5% level, \*\*\* significant at the 1% level.

Overall, results from China are largely consistent with the existing evidence documented for many developed Western countries, i.e., food consumption drops significantly when male household heads retire



at the mandatory retirement age, and that households do not seem to smooth food expenditures at retirement even though the age of retirement is fully anticipated. RD models using age in years as the running variable overestimate the food consumption drop at retirement. Because of the substantial change in the slope of the food consumption profile around the mandatory retirement age, correcting for the rounding bias has sizable effects on the estimated retirement effects in this case. Applying the proposed correction appears to be both statistically and economically important.

## 8 Extensions: Other Forms of Rounding or Non-integer Threshold

So far the analysis has focused on rounding down to the nearest integer, as in the case of how age is typically reported. However, Theorem 1 and Corollary 1 do not actually specify or require  $X$  to be  $X^*$  rounded down to the nearest integer. In particular, the assumptions that involve  $X$  are Assumptions A4, A5, and A6. While these assumptions are plausible for discretization based on rounding down, they do not require this type of rounding, and they may be applied to other types of rounding.

Still assumption A5 requires that there should be no mismeasurement in the crossing threshold dummy  $I(X \geq 0)$ . This may not hold in other common types of rounding, such as rounding up or ordinary rounding, i.e., rounding either up or down, whichever is closer. In the following I discuss these alternative forms of rounding and provide simple extensions of the previous approach to handle these cases. In particular, I show that one can simply discard observations at the cutoff, because the crossing threshold dummy is mismeasured only at that point, i.e., the true crossing threshold dummy  $I(X^* \geq 0|X = 0)$  could be 0, while the observed crossing threshold dummy  $I(X \geq 0|X = 0)$  is always 1. In these cases, observations at the cutoff contains both above and below threshold outcomes, i.e., they contains data generated by both the pre- and post-cutoff regression functions,  $h_0(X)$  and  $h_1(X)$ .

To illustrate, suppose that age is now recorded by ordinary rounding and that the threshold  $c$  is still age 62. Then individuals who are over 61.5 and under 62 will have their true crossing threshold status  $I(X^* \geq 0) = 0$ , while at the same time they will have their recorded age be 62 (based on ordinary rounding), and hence their observed crossing threshold status  $T^* = I(X \geq 0) = 1$ . These individuals are misclassified regarding their location relative to the threshold. This will tend to bias downward the treatment probability change at the cutoff.

Note that discretization by ordinary rounding or rounding up with an integer cutoff can only cause  $I(X^* \geq 0) \neq I(X \geq 0)$  at  $X = 0$ . In the above example, by ordinary rounding, everyone over age 62.5 will have both their true age and their rounded age be above the cutoff, and hence both  $X^*$  and  $X$  positive.

Similarly, everyone strictly under age 61.5 will have both their true age and their rounded age be below the cutoff and hence both  $X^*$  and  $X$  negative. Discarding observations at the cutoff can ensure that one only uses observations truly above the cutoff to estimate  $h_0(X)$  and those truly below the estimate  $h_1(X)$ .

Another way in which rounding can cause the crossing threshold dummy to be mismeasured is when the running variable is rounded to integer values while the threshold  $c$  is not an integer. For example, the age at which people born in the years 1938 to 1942 qualify for full social security benefits in the United States (called the full retirement age by the social security administration) ranges from 65 years and 2 months to 65 years and 10 months. In particular, for those who were born in 1939, the full retirement age is 65 years and 4 months, i.e., 65.33 years. Individuals who are 65.33 to 66 years old will have passed the full retirement age, given their recorded age of 65 (assuming rounding down), yet they will be mistaken as still being below the cutoff.

Note that  $X$  is normalized by subtracting off the cutoff  $c$ , and so will be non-integer valued if the cutoff is a non-integer. With rounding down,  $I(X \geq 0)$  can fail to equal  $I(X^* \geq 0)$  only for the one observable value of  $X$  right under the cutoff, i.e., the one value that lies in the interval  $-1 < X < 0$ , because the observations that have  $X^*$  right above the non-integer cutoff will be rounded down to below the cutoff. Similarly, if  $X$  is discretized by always rounding up, then  $I(X \geq 0)$  can fail to equal  $I(X^* \geq 0)$  only for the one observable value of  $X$  right above the cutoff, i.e. the one value that lies in the interval  $0 < X < 1$ . Further, if  $X$  is discretized by ordinary rounding, then  $I(X \geq 0)$  can fail to equal  $I(X^* \geq 0)$  for the one observable value of  $X$  either right under or right above the cutoff, depending on whether the cutoff is positive or negative.

In all these cases, the observed outcomes for  $X$  at or next to the cutoff contains a mix of observations right above and right under the cutoff, so treating them as if they are all at or above the cutoff or all below the cutoff leads to a biased estimate of the true treatment effect, in addition to the rounding bias that involves all points away from the cutoff.

Another way to understand the problem of mismeasuring the crossing threshold dummy is to think of its role as an instrumental variable (IV) in RD models. It is well known that the standard fuzzy design RD estimator can be interpreted as a local IV estimator, using the crossing threshold dummy  $I(X^* \geq 0)$  as an instrument for the treatment  $T$ . With these alternative types of rounding, the observed crossing threshold dummy  $I(X \geq 0)$  is mismeasured relative to the true instrument  $I(X^* \geq 0)$ , and this mismeasurement will introduce bias in the estimated treatment effect, in addition to the bias caused by rounding as in Theorem 1.

To consistently handle all the above cases, consider the following simple extension to Theorem 1 and

Corollary 1.

COROLLARY 2: Let assumptions A1 to A4 and A6 hold. Assume that if  $X \geq 1$ , then  $X^* > 0$ , and that if  $X \leq -1$ , then  $X^* < 0$ , then the conclusions of Theorem 1 and Corollary 1 hold, replacing equation (2) with

$$Y = \sum_{j=0}^J d_j X^j + \sum_{j=0}^J c_j X^j T^* + \varepsilon \text{ for all } X \geq 1 \text{ or } X \leq -1. \quad (11)$$

Since these alternative forms of rounding cause trouble only for observations at one value of  $X$  such that  $-1 < X < 1$ , Corollary 2 says that one can fix the problem by just discarding those observations from the estimation. A similar approach, dropping observations for which the running variable is mismeasured in an RD model, has been proposed by Barreca et al. (2010). In particular, Barreca et al. (2010) find that birth weights are disproportionately represented at multiples of round numbers, which caused biased RD treatment effect estimates when using birth weight as a running variable. To deal with the problem, they suggest discarding observations corresponding to heaps in the running variable (i.e., 100 gram and ounce multiples).

Ordinary rounding may not be common for reporting age, but may be more likely in other applications, such as when the running variable is a test score. If one does have the ordinary problem rounding and apply Corollary 2, then  $e$  will range from  $-.5$  to  $+.5$ , instead of ranging from zero to one. If  $e$  is uniformly distributed in this interval then  $\mu_k = E(e^k) = \int_{-.5}^{.5} e^k de = [(.5)^k - (-.5)^k] / (k + 1)$  which is zero for all even values of  $k$ , so many more elements of the matrix  $M$  will be zero than before, and hence the bias from rounding in this case is likely to be smaller.

Corollary 2 can be extended to fuzzy designs in the same way as Corollary 1, by being applied in both the numerator and denominator of the fuzzy design treatment effect  $\tau = \tau_Y / \tau_T$ .

## 9 Other Extensions and Generalizations

Let  $\tau(x^*)$  denote the RD local treatment effect at the age  $c + x^*$ , so  $\tau(0)$  equals the treatment effect  $\tau$  one has been estimating, and  $\tau(x^*)$  is the treatment effect one would identify if the cutoff were  $c + x^*$  instead of  $c$ . In the sharp design case,  $\tau(x^*) = g_1(x^*) - g_0(x^*)$ . If Assumption A2 is generalized to assume that  $g_0(X^*)$  for  $X^* < q_0$  and  $g_1(X^*)$  for  $X^* > -q_1$  are polynomials for some positive values  $q_0$  and  $q_1$  (which means making untestable modeling assumptions regarding counterfactuals) then the treatment

effects  $\tau(x^*)$  will be identified for all  $X^*$  in the interval  $-q_1 \leq X^* \leq q_0$ . This is because one can identify and recover the coefficients of the correct model  $g_t(X^*)$  from the coefficients of the estimated model  $h_t(X)$ . So if the models  $g_t(X^*)$  remain correctly identified in the counterfactual regions, that is,  $g_0(X^*)$  for  $0 \leq X^* \leq q_0$  and  $g_1(X^*)$  for  $-q_1 \leq X^* < 0$ , then the true treatment effects can be identified and estimated in those regions.

Even without generalizing Assumption A2, one can still identify and estimate  $\partial \tau(x^*) / \partial x^*$  evaluated at the point  $x^* = 0$ . This is the Marginal Threshold Treatment Effect (MTTE) of Dong and Lewbel (2011). It describes the marginal change in the RD treatment effect that would result from making a marginal change in the cutoff threshold. With polynomial models, in the sharp design the true MTTE is just  $b_1$ , and so equals the second element of  $M^{-1}C$ , while the naive biased MTTE based on rounded data is  $c_1$ , the second element of  $C$ . In the earlier example where  $J = 4$ , given estimated coefficients  $c_k$  for  $k = 1, 2, 3, 4$  from the regression (3), the true MTTE is given by  $b_1 = c_1 - 2\mu_1 c_2 - (3\mu_2 - 6\mu_1^2) c_3 + (-24\mu_1^3 + 24\mu_2\mu_1 - 4\mu_3) c_4$ , and for the uniform distribution case, this MTTE simplifies to  $c_1 - c_2 + \frac{1}{2}c_3$ . As before, these same results can be applied to lower order polynomials by setting the coefficients  $c_k$  for higher values of  $k$  equal to zero.

Suppose  $E(T | X^*)$  has a kink (a discrete change in slope) instead of a discontinuity at  $X^* = 0$ . Recall that a t test can be used to test whether a discontinuity exists in the conditional treatment probability  $E(T | X^*)$ , i.e., to test the statistical significance of the bias corrected estimate  $\hat{\tau}_t$ . Dong (2011) shows that in the case of fuzzy design RD, the treatment effect can still exist when there is only a slope change, and can be identified from a slope change rather than a level change in the treatment probability  $E(T | X^*)$  along with a corresponding slope change in the conditional mean outcome equation  $E(Y | X^*)$  at the cutoff  $X^* = 0$ . In particular, the RD treatment effect in this case is given by the ratio of the two slope changes in the treatment and outcome regression functions at the cutoff. In this case, the first element of the vector  $B = M^{-1}C$ , i.e.,  $b_0$  or the jump is zero, and the slope change required for Dong (2011)'s kink based RD treatment effect estimator, is given by the second element of  $B = M^{-1}C$ , or  $b_1$ .

When  $J = 4$ , for the fuzzy design RD model given by equations (8) and (9), the naive discrete data kink based treatment effect is given by the ratio of the two slopes in these two equations, i.e.,  $\tau_k = c_1/s_1$ . Given that  $X$  is rounded, the slopes in the numerator and in the denominator may be biased, so one can similarly apply the bias correction formula to obtain the corrected kink-based treatment effect. Assuming the rounding error is uniformly distributed, the bias corrected kink-based RD treatment effect in Dong (2011) is given by

$$\tau'_k = \frac{c_1 - c_2 + \frac{1}{2}c_3}{s_1 - s_2 + \frac{1}{2}s_3},$$

and more generally, it is given by

$$b_1 = \frac{c_1 - 2\mu_1 c_2 - (3\mu_2 - 6\mu_1^2) c_3 + (-24\mu_1^3 + 24\mu_2\mu_1 - 4\mu_3) c_4}{s_1 - 2\mu_1 s_2 - (3\mu_2 - 6\mu_1^2) s_3 + (-24\mu_1^3 + 24\mu_2\mu_1 - 4\mu_3) s_4}.$$

## 10 Conclusions

In the context of RD models where the running variable is rounded and hence is discrete, the standard RD estimation yields biased estimates of the RD treatment effects, even if the functional form of the model is correctly specified. In practice, this rounding or discretization bias can be very easily corrected. This paper presents simple formulas to fix this bias and hence provides consistent estimates of RD treatment effects given only rounded data of the running variable. The proposed approach does not require instrumental variables, but instead uses information regarding the distribution of rounding errors within the discretization cell, e.g., the distribution of ages within a year in the case of using age in years as a running variable. This can be easily obtained from census data, and often close to uniform.

In one empirical application, I investigate the effect of Medicare eligibility at 65 on insurance coverage in the US. Higher frequency age data (age in months) are available, and so provide a benchmark. I show that the proposed method to correct the rounding bias works well and produces estimates that are consistent with having and using data where age is more accurately measured.

In another empirical application, I provide an RD analysis that exploits the mandatory retirement policy in China to test for the presence of, and estimate the magnitude of, a retirement-consumption puzzle in China. In this case, the food consumption profile around the mandatory retirement age 60 for male workers have relatively large slope changes, so the rounding bias is sizable, and the bias correction is empirically important.

Although this paper's main analysis focuses on the case of rounding down and an integer cutoff, I show that the proposed methodology can be immediately extended to cases involving non-integer cutoffs or other common forms of rounding, such as ordinary rounding or rounding up to the nearest integer. In addition, this paper uses age as the leading example, but the proposed approach can be used in other RD applications where the running variable is similarly rounded. Examples include using calendar years, birth weight in ounces or integer valued test scores as a running variable.

A recent development in the RD literature is estimation of quantile RD treatment effects (Frandsen, Frolich, and Melly 2012). One interesting topic for future research is then to extend the current results to the case of estimating quantile RD treatment effects.

## 11 Appendix

**Proof of Theorem 1:**  $X$  equals the integer part of  $X^*$ , and hence it is a deterministic function of  $X^*$ ; therefore,  $E(Y | X^*, T^*) = E(Y | X^*, X, T^*)$ . By this result and the law of iterated expectations, we have

$$E(Y | X, T^*) = E[E(Y | X^*, T^*) | X, T^*]$$

Therefore,

$$E(Y | X, T^* = t) = E[E(Y | X^*, T^* = t) | X, T^* = t] = E[g_t(X^*) | X] = E[g_t(X + e) | X].$$

The second equality follows from  $T^* = I(X \geq 0) = I(X^* \geq 0)$  being a deterministic function of  $X^*$  and  $E(Y | X^*, I(X^* \geq 0) = t) = g_t(X^*)$  for  $t = 0, 1$  given a sharp design.

Given Assumption A3, define the unknown polynomial coefficients  $b_{jt}$  by

$$g_t(X^*) = \sum_{j=0}^J b_{jt} X^{*j} \quad (12)$$

and let  $B_t$  be the column vector of elements  $b_{0t}, b_{1t}, \dots, b_{Jt}$  for  $t = 0, 1$ . Then

$$g_t(X + e) = \sum_{j=0}^J b_{jt} (X + e)^j = \sum_{j=0}^J \sum_{k=0}^j \binom{j}{k} b_{jt} e^{j-k} X^k = \sum_{k=0}^J \sum_{j=k}^J \binom{j}{k} b_{jt} e^{j-k} X^k$$

where  $\binom{j}{k}$  is the binomial coefficient  $\frac{j!}{k!(j-k)!}$ . Substituting this expression for  $g_t$  into the equation for  $E(Y | X, T^* = t)$  gives

$$E(Y | X, T^* = t) = \sum_{k=0}^J \sum_{j=k}^J \binom{j}{k} b_{jt} E(e^{j-k} | X) X^k$$

By A6  $E(e^k | X) = \mu_k = E(e^k)$  is known. By A1, A4, and the definition of  $h_t(X)$ , we have  $h_0(X) = E(Y | X, T^* = 0)$  when  $X < 0$  and  $h_1(X) = E(Y | X, T^* = 1)$  when  $X \geq 0$ . Putting these equations together gives

$$h_t(X) = \sum_{k=0}^J c_{kt} X^k \quad \text{where} \quad c_{kt} = \sum_{j=k}^J \binom{j}{k} \mu_{j-k} b_{jt}. \quad (13)$$

By A6, the value of  $h_t(X)$  is identified at  $J$  (or more) values of  $X$  for  $t = 0, 1$ . Since the above shows that  $h_t(X)$  is a polynomial of order  $J$ , and any polynomial of order  $J$  is uniquely identified by its values at  $J + 1$  points, it follows that the coefficients  $c_{kt}$  are identified. Note that one does not need to know the polynomial order  $J$  a priori, since given this paper's assumption the observed values of  $h_t(X)$  will trace out the polynomial of proper degree, thereby identifying  $J$ .

Equation (13) shows the connection between the coefficients in the discrete data regression  $b_{jt}$  and the true continuous data regression  $c_{kt}$  for  $j, k = 0, 1, \dots, J$ . The following express this relationship in matrix notation.

For  $t = 0, 1$ , define the upper triangular  $J + 1$  by  $J + 1$  matrix  $M$  as having the element  $\binom{j}{k}\mu_{j-k}$  in row  $k + 1$  and column  $j + 1$  for all  $j, k$  satisfying  $0 \leq k \leq j \leq J$ . All elements of  $M$  below the diagonal are zero. Recall that  $C_t$  consists of elements  $c_{kt}$  for  $k = 0, 1, \dots, J$  and that  $B_t$  consists of elements  $b_{jt}$  for  $j = 0, 1, \dots, J$ . Given the matrix  $M$ ,  $c_{kt} = \sum_{j=k}^J \binom{j}{k}\mu_{j-k}b_{jt}$ , for  $t = 0, 1$  can be rewritten as

$$C_t = MB_t,$$

The matrix  $M$  is identified because, by A6, each  $\mu_{j-k}$  is identified.

Note that the matrix  $M$  is nonsingular, because it is triangular with all ones on its diagonal, and all finite values off its diagonal, as  $e$  is bounded between 0 and 1. Then  $M$  can be inverted to solve for  $b_{jt}$  for  $j = 0, 1, \dots, J$  given the already identified constants  $c_{kt}$  for  $k = 0, 1, \dots, J$ . Therefore the constants  $b_{jt}$  are all identified, and  $B_t = M^{-1}C_t$ . For the special case that  $e = 0$  with probability one,  $b_{jt} = c_{jt}$  for all  $j$  and  $t$ . Then  $\tau = g_1(0) - g_0(0) = b_{01} - b_{00}$  is identified. Also the conditional mean functions  $g_0(X^*)$  for  $X^* < 0$  and  $g_1(X^*)$  for  $X^* \geq 0$  are identified, since  $g_t(X^*) = \sum_{j=0}^J b_{jt}X^{*j}$ .

**Proof of Corollary 1:** By construction  $E(Y | X, T^*) = h_0(X) + [h_1(X) - h_0(X)]T^*$ , and this equation along with equation (13) yields equation (2) with  $d_j = c_{j0}$  and  $c_j = c_{j1} - c_{j0}$  for  $j = 0, \dots, J$ , and hence part (i) of the Corollary holds given the property that any polynomial of degree  $J$  is identified by  $J + 1$  points.

By construction,  $A$  equals the vector of elements  $b_{00}, b_{10}, \dots, b_{J0}$ , and  $B$  equals  $B_1 - B_0$ , which equals the vector of elements  $b_{01} - b_{00}, b_{11} - b_{10}, \dots, b_{J1} - b_{J0}$ . Recall  $B$  is defined as a vector of elements  $b_0, b_1, \dots, b_J$ . So  $\tau = b_0$  is the first element of  $B$ . Similarly, by construction  $D$  equals the vector of elements  $c_{00}, c_{10}, \dots, c_{J0}$  and  $C$  equals  $C_1 - C_0$ , which equals the vector of elements  $c_{01} - c_{00}, c_{11} - c_{10}, \dots, c_{J1} - c_{J0}$ . Recall  $C$  is defined as a vector of elements  $c_0, c_1, \dots, c_J$ . So  $\tau' = c_0$  is the first element of  $C$ . By Theorem 1,  $C_t = MB_t$ , for  $t = 0, 1$ , which implies  $MA = D$  and  $MB = C$ , and  $M$  is invertible by Theorem 1,

which gives part (ii). Also,  $C_t = MB_t$  for  $t = 0, 1$  leads to  $C_1 - C_0 = M(B_1 - B_0)$ , i.e.,  $C = MB$ . The first row of this matrix equation is  $c_0 = b_0 + \sum_{j=1}^J b_j \mu_j$  and so part (iii) holds.

**Proof of Corollary 2:** The assumptions of Corollary 2 imply that  $T^* = I(X \geq 0)$  for all values of  $X \geq 1$  or  $X \leq -1$ , i.e., there is no mismeasurement error in  $T^*$ . So repeat the steps of the proofs of Theorem 1 and Corollary 1 using all values of  $X$  except those having  $-1 < X < 1$ .

## References

- [1] Aguila, E., O. Attanasio, and C. Meghir, (2011) "Changes in Consumption at Retirement: Evidence from Panel Data," *The Review of Economics and Statistics*, 93(3): 1094-1099
- [2] Ameriks, J., A. Caplin, and J. Leahy (2007), "Retirement Consumption: Insights from a Survey," *The Review of Economics and Statistics*, 89:2, 265-274.
- [3] Babanezhad, M., S. Vansteelandt, and E. Goetghebeur (2010), "Comparison of causal effect estimators under Exposure Misclassification," *Journal of Statistical Planning and Inference*, 140, 1306-1319.
- [4] Banks, James, R. Blundell, and S. Tanner (1998), "Is there a Retirement-Savings Puzzle?" *American Economic Review*, 88:4 769-788.
- [5] Barreca, A., M. Guldi, J. M. Lindo, and G. R. Waddell (2010), "Running and Jumping Variables in RD Designs: Evidence based on Race, Socioeconomic Status, and Birth Weights," IZA Discussion Papers 5106, Institute for the Study of Labor (IZA).
- [6] Battistin, E., A. Brugiavini, E. Rettore, and G. Weber (2009), "The Retirement Consumption Puzzle: Evidence from a Regression Discontinuity Approach," *American Economic Association*, 99(5): 2209-2226.
- [7] Battistin, E. and A. Chesher (2011), "Treatment Effect Estimation With Covariate Measurement Error," Unpublished Cemmap working paper.
- [8] Battistin, E. and B. Sianesi (2011), "Misreported Schooling and Returns to Education: Evidence from the UK," *Review of Economics and Statistics*, 93, 495-509.



- [9] Behaghel, L., B. Crepon, and B. Sedillot (2008), "The Perverse Effects of Partial Employment Protection Reform: The Case of French Older Workers." *Journal of Public Economics*, 92(3–4): 696–721.
- [10] Beresford, G. C. (1980), "The Uniformity Assumption in the Birthday Problem," *Mathematics Magazine*, 53, 286-288.
- [11] Bernheim, D., J. Skinner, and S. Weinberg (2001), "What Accounts for the Variation in Retirement Wealth Among U.S. Households," *American Economic Review*, 91(4): 832-857.
- [12] Borella, M., F. C. Moscarola, and M. Rossi (2011), "(Un)expected retirement and the consumption puzzle," unpublished manuscript.
- [13] Card, D., C. Dobkin, and N. Maestas (2008), "The Impact of Nearly Universal Insurance Coverage on Health Care Utilization: Evidence from Medicare." *American Economic Review*, 98(5): 2242–58.
- [14] Card, D., C. Dobkin, and N. Maestas (2009), "Does Medicare Save Lives?" *Quarterly Journal of Economics*, 124(2): 597–636.
- [15] Card, D. and L. Shore-Sheppard (2004), "Using discontinuous eligibility rules to identify the effects of the federal medicaid expansions on low income children," *Review of Economics and Statistics* 86, 752–766.
- [16] Carpenter, C. and C. Dobkin (2009), "The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age," *American Economic Journal: Applied Economics*, 1(1): 164–82.
- [17] Chen, S. and W. van der Klaauw (2008), "The Work Disincentive Effects of the Disability Insurance Program in the 1990s," *Journal of Econometrics*, 142(2): 757–84.
- [18] Cochran, W. G. and D. B. Rubin (1973), "Controlling Bias in Observational Studies: a Review," *Sankhya Series A*, 35, 417-466.
- [19] De Giorgi, G. (2005), "Long-Term Effects of a Mandatory Multistage Program: The New Deal for Young People in the UK," Institute for Fiscal Studies Working Paper 05/08.
- [20] DiNardo, J. and D. S. Lee (2004), "Economic impacts of new unionization on private sector employers: 1984–2001," *Quarterly Journal of Economics*, 119, 1383–1442.

- [21] Dong, Y. (2011), "Jumpy or Kinky? Regression Discontinuity Without the Discontinuity," Unpublished CSUF working paper.
- [22] Dong, Y. and A. Lewbel (2011), "Regression Discontinuity Marginal Threshold Treatment Effects," Boston College Working Paper #759.
- [23] Du, Y. and M. Wang, (2011), "Population Ageing, Domestic Consumption and Future Economic Growth in China," in *Rising China Global Challenges and Opportunities*, J. Golley and L. Song, eds., chap. 17, 301-314, Australian National University Press.
- [24] Edmonds, E. V (2004), "Does Illiquidity Alter Child Labor and Schooling Decisions? Evidence from Household Responses to Anticipated Cash Transfers in South Africa," National Bureau of Economic Research Working Paper 10265.
- [25] Edmonds, E. V., K. Mammen, and D. L. Miller (2005), "Rearranging the Family? Income Support and Elderly Living Arrangements in a Low-Income Country," *Journal of Human Resources*, 40(1): 186–207.
- [26] Ferreira, F. (2010), "You Can Take It With You: Proposition 13 Tax Benefits, Residential Mobility, and Willingness to Pay for Housing Amenities" *Journal of Public Economics*, 94(9-10): 661-673.
- [27] Frandsen, Frolich, and Melly (2012), "Quantile Treatment Effects in the Regression Discontinuity Design," Unpublished manuscript.
- [28] Haider, S. J. and M. Stephens (2007) "Is There a Retirement-Consumption Puzzle? Evidence Using Subjective Retirement Expectations," *The Review of Economics and Statistics*, 89(2): 247-264.
- [29] Hu, Y. (2008), "Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: A General Solution," *Journal of Econometrics*, 144, 27-61.
- [30] Hurd, M. D. and S. Rohwedder (2008), "The Retirement Consumption Puzzle: Actual Spending Change in Panel Data," RAND, Working Paper WR-563.
- [31] Hurst, E. (2008), "The Retirement of a Consumption Puzzle," NBER working paper 13789.
- [32] Imbens, G. W. and T. Lemieux (2008), "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142, 615–35.

- [33] Kane, T. J. (2003), A quasi-experimental estimate of the impact of financial aid on college-going. Working Paper, National Bureau of Economic Research.
- [34] Lalive, R. (2007), "Unemployment Benefits, Unemployment Duration, and Post-unemployment Jobs: A Regression Discontinuity Approach." *American Economic Review*, 97(2): 108–12.
- [35] Lalive, R. (2008), "How Do Extended Benefits Affect Unemployment Duration? A Regression Discontinuity Approach," *Journal of Econometrics*, 142(2): 785–806.
- [36] Lalive, R., J. C. van Ours, and J. Zweimuller (2006), "How Changes in Financial Incentives Affect the Duration of Unemployment," *Review of Economic Studies*, 73(4): 1009–38.
- [37] Lee, D. S. (2008), "Randomized experiments from non-random selection in U.S. house elections," *Journal of Econometrics*, 142: 675–697.
- [38] Lee, D. S. and D. Card, (2008), "Regression discontinuity inference with specification error," *Journal of Econometrics* 142, 655-674.
- [39] Lee, D. S. and T. Lemieux (2010), "Regression Discontinuity Designs in Economics," *Journal of Economic Literature* 48, 281–355.
- [40] Lee, D. S. and J. McCrary (2005), "Crime, Punishment, and Myopia," National Bureau of Economic Research Working Paper 11491.
- [41] Lemieux, T. and K. Milligan (2008), "Incentive Effects of Social Assistance: A Regression Discontinuity Approach," *Journal of Econometrics*, 142(2): 807–28
- [42] Leuven, E. and H. Oosterbeek (2004), "Evaluating the Effect of Tax Deductions on Training," *Journal of Labor Economics*, 22(2): 461–88.
- [43] Lewbel, A. (2007), "Estimation of Average Treatment Effects With Misclassification," *Econometrica*, 75, 537-551.
- [44] Mahajan, A. (2006), "Identification and Estimation of Regression Models with Misclassification," *Econometrica*, 74, 631-665.
- [45] Manski, C. F. and E. Tamer (2002), "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519-546.

- [46] Miniaci, R., C. Monfardini, and G. Weber (2003), "Is There a Retirement Consumption Puzzle in Italy?" Institute for Fiscal Studies, Working Paper W03/14.
- [47] Molinari, F. (2008), "Partial Identification of Probability Distributions with Misclassified Data," *Journal of Econometrics*, 144, 81-117.
- [48] Murphy, R. (1996), "An analysis of the distribution of birthdays in a calendar year," unpublished manuscript available as <http://www.panix.com/~murphy/bday.html>.
- [49] Oreopoulos, P. (2006), "Estimating Average and Local Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter," *American Economic Review*, 96(1): 152–75.
- [50] Robb, A. L. and J. B. Burbridge (1989), "Consumption, Income, and Retirement," *Canadian Journal of Economics*, 22(3): 522-42.
- [51] Schwerdt, G. (2005), "Why Does Consumption Fall at Retirement? Evidence from Germany," *Economics Letters*, 89(3): 300-305.