Robust Inference in Fuzzy Regression Discontinuity with Multiple Forcing Variables

Rong Ma* Ang Sun† Zhaoguo Zhan‡

October 5, 2016

Abstract

We propose a test on the causal effect in fuzzy regression discontinuity designs, where treatment assignment is allowed to depend on more than one forcing variable. The conventional t-test is not adequate for this purpose, due to its potential size distortion caused by non-substantial discontinuity in the probability of receiving treatment. The robust t-test of Feir et al. (2016) is also infeasible, since it is restricted to the one forcing variable setup. Unlike the conventional t-test, our test does not rely on the magnitude of discontinuity in receiving treatment, so it remains size-correct when this magnitude is not substantial. Different from the robust t-test, our regression-based test procedure remains easy to implement under multiple forcing variables. For illustration, we use the proposed test to study whether the awareness of hypertension decreases fat intake.

JEL Classification: C14; C18; C21; C26

Keywords: fuzzy regression discontinuity; multiple forcing variables; weak identification

*Department of Economics, School of Economics and Management, Tsinghua University, Beijing, China, 100084. Email: mar.14@sem.tsinghua.edu.cn.

†Central University of Finance and Economics, Beijing, China, 100081. Email: ang.sun@gmail.com.

‡Corresponding author. Department of Economics, Finance and Quantitative Analysis, Coles College of Business, Kennesaw State University, GA 30144, USA. Email: zzhan@kennesaw.edu. Phone: (470)578-6537. Fax: (470)578-9022.

1 Introduction

Regression Discontinuity (RD) identifies causal effects by exploiting the discontinuity in treatment assignment. Since the late 1990s, it has become a standard tool of economists. An introduction to the vast RD literature can be found in the survey article of Lee and Lemieux (2010), while empirical RD examples include, e.g., Angrist and Lavy (1999), Black (1999), Van Der Klaauw (2002), Lee (2008), Gagliarducci and Paserman (2012), Eggleston et al. (2016).

This paper is motivated by two recent advances in the RD literature. First, most existing RD studies concern the scalar case where a single forcing variable determines whether treatment is assigned, while more recently, there are also numerous RD examples with multiple forcing variables. See, e.g., Papay et al. (2011), Cohodes and Goodman (2014), Snider and Williams (2015), Egger and Wamser (2015). Second, Feir et al. (2016) point out that a large proportion of existing RD studies suffer a weak identification problem. Specifically, if discontinuity in the probability of receiving treatment is not sufficiently large in the so-called fuzzy RD designs, then inference on the causal effect by the conventional t-test is no longer reliable.

We consider the two advances above simultaneously. That is, the fuzzy RD design studied in this paper allows for d forcing variables with $d \geq 1$, so it covers the existing scalar RD designs as well as the emerging multi-dimensional ones; in addition, we allow the discontinuity in the probability of receiving treatment to be potentially small (weak), so the weak identification problem studied in Feir et al. (2016) is accounted for. Our objective is to develop an inference method for the causal effect in the fuzzy RD design, which is potentially associated with multiple forcing variables and non-substantial magnitude of discontinuity.

Feir et al. (2016) have provided a robust inference method for fuzzy RD. Their approach is to modify the standard error of the fuzzy RD estimator for the causal effect, after taking weak identification into account. Using the modified standard error to replace its conventional counterpart, Feir et al. (2016) come up with a robust t-test. When weak identification exists, they show the robust t-test has correct size, while the conventional t-test does not. As a result, the robust t-test is preferred to the conventional one for statistical inference. The robust t-test in Feir et al. (2016), however, is restricted to the scalar RD setup with $d = 1$, i.e., treatment assignment is determined by a single forcing variable.

We suggest a robust inference method in fuzzy RD with potentially multiple forcing variables, so $d \geq 1$. Like Feir et al. (2016), our insight also comes from the sizeable weak identification literature (Anderson and Rubin 1949, Bound et al. 1995, Staiger and Stock 1997, Kleibergen 2002, Moreira 2003). Unlike Feir et al. (2016), we do not rely on the fuzzy RD estimator or its robust standard error for conducting causal inference. Instead, our suggested method builds on a regression-based statistic suggested by Anderson and Rubin (1949).

Our approach utilizes an interesting coincidence, i.e., the fuzzy RD estimator can also be viewed as a Two-Stage-Least-Squares (TSLS) estimator, which was discovered by Hahn et al. (2001). Specifically, the binary variable of the treatment assignment rule serves as the instrument for the endogenous variable that describes whether treatment is received, so the TSLS model of fuzzy RD is just-identified. Under just-identification of TSLS, the weak identification literature has suggested that the Anderson and Rubin (1949) test is size-correct and does not suffer power loss. Combining all these inputs, it is thus natural to adopt the Anderson and Rubin (1949) test in fuzzy RD designs. In addition, since this test procedure is regression-based, it is easy to implement in practice. When there are multiple forcing variables instead of a scalar forcing variable, the only needed change for the test procedure is to include multiple covariates in linear regression analysis. Consequently, there is no burden of extending robust inference from the scalar RD to the multi-dimensional case, if the proposed approach in this paper is adopted.

We conduct simulation experiments to compare the conventional t-test, the robust t-test of Feir et al. (2016) and the proposed Anderson and Rubin (AR) test. As expected, we find the evidence that the proposed AR test performs well in size and power. By contrast, the conventional t-test exhibits size distortion, when the magnitude of discontinuity in the probability of receiving treatment is small. In addition, in the scalar RD simulation setup, we find that the robust t-test and the AR test have similar power.

For illustrative purposes, we use the proposed test to study the impact of hypertension diagnosis on food demand. How the provision of health information affects human behavior has already attracted sizeable attention, see, e.g., Chern et al. (1995), Roosen et al. (2009). For hypertension, its diagnosis is commonly known to rely on two blood pressures: the systolic pressure and the diastolic pressure. When one of the two pressures is above its cutoff ($systolic: 140\ mmHg$, $diastolic: 90\ mmHg$), hypertension is diagnosed. Consequently, hypertension diagnosis can serve as a simple RD example with $d = 2$ forcing variables, namely, the systolic and diastolic pressures. Since both doctors and patients might mess up the diagnosis, misunderstanding of the hypertension status is not rare, especially for those around the cutoff of blood pressures. As a result, we face a fuzzy RD design with multiple forcing variables, where the discontinuity in hypertension diagnosis at the cutoff is possibly weak. Does this diagnosis of hypertension change human behavior such as fat intake? This question has been studied in the literature by adopting existing RD approaches (see, e.g., Zhao et al. 2013), while we revisit the question by applying our robust approach.

The rest of the paper proceeds as follows. Section 2 starts with the description of the inference problem in a fuzzy RD design with multiple forcing variables, and then proposes a robust test as the solution. Section 3 provides simulation evidence for the proposed test. Section 4 presents the empirical application. Section 5 concludes.

2 Fuzzy RD, TSLS and Anderson-Rubin Test

2.1 Basics of Fuzzy RD

The objective of Regression Discontinuity designs is to study the impact of a binary treatment variable W_i on the outcome variable Y_i. In accordance with the existing literature, let Y_{0i} denote the outcome of the i-th unit in the absence of treatment, while Y_{1i} denotes the outcome with treatment, so:

$$Y_i = Y_{0i} + \tau_i W_i \tag{1}$$

where $\tau_i \equiv Y_{1i} - Y_{0i}$ reflects the individual causal effect, $i = 1, ..., n$.

Suppose treatment assignment depends on a vector of d covariates (forcing variables), denoted by $\boldsymbol{X}_i$. In particular, treatment is assigned when $\boldsymbol{X}_i$ lies in the treatment region denoted by $\mathbb{T}$, and likewise treatment is not assigned when $\boldsymbol{X}_i \in \mathbb{T}^c$, the compliment of $\mathbb{T}$. Let T_i be the binary variable for the treatment assignment rule, so:

$$T_i = \mathbf{1}(\boldsymbol{X}_i \in \mathbb{T}) \tag{2}$$

A subtle point is that the actual treatment variable W_i does not necessarily coincide with T_i, if there exists noncompliance with the treatment assignment rule. For instance, units that are not in the treatment region may also receive treatment.

The treatment region in RD designs is commonly related to some cutoff point, denoted by $\boldsymbol{x}$. In scalar RD designs, it is typical that $T_i = \mathbf{1}(\boldsymbol{X}_i \geq \boldsymbol{x})$. In multi-dimensional RD designs, the cutoff point is not unique, and the set of cutoff points forms a boundary. For example, consider hypertension diagnosis with $\boldsymbol{X}_i = (systolic_i, diastolic_i)'$, and treatment is assigned if either $systolic_i \geq 140\ mmHg$ or $diastolic_i \geq 90\ mmHg$. Points with either $systolic_i = 140\ mmHg$ or $diastolic_i = 90\ mmHg$ thus serve as cutoff points, and they form a boundary in the 2-dimensional space of $\boldsymbol{X}_i$.

In addition, let $N_\epsilon(\boldsymbol{x})$ denote the set of $\boldsymbol{X}_i$ that lies in the ϵ-neighborhood around a given cutoff point $\boldsymbol{x}$. So $N_\epsilon^+(\boldsymbol{x}) \equiv N_\epsilon(\boldsymbol{x}) \cap \mathbb{T}$ contains the units in the treatment region around $\boldsymbol{x}$, and likewise $N_\epsilon^-(\boldsymbol{x}) \equiv N_\epsilon(\boldsymbol{x}) \cap \mathbb{T}^c$ contains the units in the control region around $\boldsymbol{x}$. The notation here is adopted from Imbens and Zajonc (2011).

Assume the following limits exist:

$$W^+(\boldsymbol{x}) \equiv \lim_{\epsilon \to 0} \mathbb{E}[W_i | \boldsymbol{X}_i \in N_\epsilon^+(\boldsymbol{x})],\ W^-(\boldsymbol{x}) \equiv \lim_{\epsilon \to 0} \mathbb{E}[W_i | \boldsymbol{X}_i \in N_\epsilon^-(\boldsymbol{x})] \tag{3}$$

$$Y^+(\boldsymbol{x}) \equiv \lim_{\epsilon \to 0} \mathbb{E}[Y_i | \boldsymbol{X}_i \in N_\epsilon^+(\boldsymbol{x})],\ Y^-(\boldsymbol{x}) \equiv \lim_{\epsilon \to 0} \mathbb{E}[Y_i | \boldsymbol{X}_i \in N_\epsilon^-(\boldsymbol{x})] \tag{4}$$

with $W^+(\boldsymbol{x}) \neq W^-(\boldsymbol{x})$, i.e., the probability of receiving treatment is discontinuous at the cutoff point $\boldsymbol{x}$, which is essential for RD designs. We do not impose other extra condition on $W^+(\boldsymbol{x}) - W^-(\boldsymbol{x})$, so its magnitude is allowed to be small.

The common object of interest in RD designs reads

$$\tau_{RD}(\boldsymbol{x}) = \frac{Y^+(\boldsymbol{x}) - Y^-(\boldsymbol{x})}{W^+(\boldsymbol{x}) - W^-(\boldsymbol{x})} \tag{5}$$

which is interpreted as the average treatment effect at $\boldsymbol{x}$, i.e., $\tau_{RD}(\boldsymbol{x}) = \mathbb{E}(\tau_i | \boldsymbol{X}_i = \boldsymbol{x})$, if regularity conditions are satisfied. See, e.g., Hahn et al. (2001). Since the cutoff point $\boldsymbol{x}$ is unique in scalar RD designs, $\tau_{RD}(\boldsymbol{x})$ could also be written as τ_{RD} for ease of notation. In multi-dimensional RD, however, $\boldsymbol{x}$ is not unique, so $\tau_{RD}(\boldsymbol{x})$ is used to avoid confusion. The set of $\tau_{RD}(\boldsymbol{x})$ consequently reflects the treatment effect on the boundary of multi-dimensional RD designs.

In the so-called sharp RD design, all units comply with the treatment assignment rule, so $W^+(\boldsymbol{x}) = 1$, $W^-(\boldsymbol{x}) = 0$, and $\tau_{RD}(\boldsymbol{x})$ reduces to $Y^+(\boldsymbol{x}) - Y^-(\boldsymbol{x})$. More generally, if treatment is received in a non-deterministic manner with $1 \geq W^+(\boldsymbol{x}) > W^-(\boldsymbol{x}) \geq 0$, then we have the fuzzy RD design, which nests the sharp RD as a special case.

Furthermore, it is common in existing RD studies to consider $d = 1$, i.e., treatment depends on a single forcing variable. However, it is more flexible to allow $d \geq 1$, since treatment assignment could depend on multiple forcing variables.

In view of the above, we focus on the inference problem on $\tau_{RD}(\boldsymbol{x})$ in the fuzzy RD design with $d \geq 1$.

2.2 Local Linear Regression and TSLS

Although methods for estimating $\tau_{RD}(\boldsymbol{x})$ in (5) are not unique, Hahn et al. (2001) advocate the use of local linear regression, which helps address the boundary problem that arises when estimating $Y^+(\boldsymbol{x})$, $Y^-(\boldsymbol{x})$, $W^+(\boldsymbol{x})$ and $W^-(\boldsymbol{x})$ non-parametrically. The general discussion on local linear regression is provided in Fan and Gijbels (1996), and Porter (2003) studies its attractive bias and rate properties. For the kernel function of local linear regression, we follow Imbens and Lemieux (2008) to employ the uniform kernel, since using other types of kernels merely makes differences.[1]

By adopting the multivariate local linear regression and the uniform kernel with bandwidth denoted by h, $Y^+(\boldsymbol{x})$, $Y^-(\boldsymbol{x})$, $W^+(\boldsymbol{x})$ and $W^-(\boldsymbol{x})$ can be estimated as follows. Consider:

$$\min_{\alpha_{y1},\beta_{y1}} \sum_{i:\boldsymbol{X}_i \in N_h^+(\boldsymbol{x})} [Y_i - \alpha_{y_1} - (\boldsymbol{X}_i - \boldsymbol{x})'\beta_{y_1}]^2 \tag{6}$$

$$\min_{\alpha_{y0},\beta_{y0}} \sum_{i:\boldsymbol{X}_i \in N_h^-(\boldsymbol{x})} [Y_i - \alpha_{y_0} - (\boldsymbol{X}_i - \boldsymbol{x})'\beta_{y_0}]^2 \tag{7}$$

$$\min_{\alpha_{w1},\beta_{w1}} \sum_{i:\boldsymbol{X}_i \in N_h^+(\boldsymbol{x})} [W_i - \alpha_{w_1} - (\boldsymbol{X}_i - \boldsymbol{x})'\beta_{w_1}]^2 \tag{8}$$

$$\min_{\alpha_{w0},\beta_{w0}} \sum_{i:\boldsymbol{X}_i \in N_h^-(\boldsymbol{x})} [W_i - \alpha_{w_0} - (\boldsymbol{X}_i - \boldsymbol{x})'\beta_{w_0}]^2 \tag{9}$$

and $\hat{\alpha}_{y_1}$, $\hat{\alpha}_{y_0}$, $\hat{\alpha}_{w_1}$, $\hat{\alpha}_{w_0}$ resulting from these four minimization problems are the estimators of $Y^+(\boldsymbol{x})$, $Y^-(\boldsymbol{x})$, $W^+(\boldsymbol{x})$ and $W^-(\boldsymbol{x})$, respectively.

Note that we use the same bandwidth h above for (6) - (9), following the suggestion of Imbens and Lemieux (2008). In principle, however, different bandwidths could be adopted for estimating $Y^+(\boldsymbol{x})$, $Y^-(\boldsymbol{x})$, $W^+(\boldsymbol{x})$ and $W^-(\boldsymbol{x})$. The optimal choice of h has been extensively studied in the RD bandwidth literature. See, e.g., Imbens and Kalyanaraman (2012). We do not contribute to this literature in this paper, and rely on it to suggest the value of h.

With the notation above, an estimator of $\tau_{RD}(\boldsymbol{x})$ thus reads:

$$\hat{\tau}_{RD}(\boldsymbol{x}) = \frac{\hat{\alpha}_{y_1} - \hat{\alpha}_{y_0}}{\hat{\alpha}_{w_1} - \hat{\alpha}_{w_0}} \tag{10}$$

Interestingly, as pointed out by Hahn et al. (2001), $\hat{\tau}_{RD}(\boldsymbol{x})$ can be viewed as the commonly

[1] If different kernels cause substantially different RD findings, then such RD studies are unlikely to be reliable. See Imbens and Lemieux (2008).

used Two-Stage-Least-Squares estimator. Consider the auxiliary regression equation below:

$$Y_i = \boldsymbol{R}_i'\boldsymbol{\beta} + \epsilon_i \tag{11}$$

and let $\boldsymbol{S}_i$ be the instruments for $\boldsymbol{R}_i$ with

$$\boldsymbol{R}_i = \begin{bmatrix} 1 \\ W_i \\ T_i \cdot (\boldsymbol{X}_i - \boldsymbol{x}) \\ (1 - T_i) \cdot (\boldsymbol{X}_i - \boldsymbol{x}) \end{bmatrix}, \quad \boldsymbol{S}_i = \begin{bmatrix} 1 \\ T_i \\ T_i \cdot (\boldsymbol{X}_i - \boldsymbol{x}) \\ (1 - T_i) \cdot (\boldsymbol{X}_i - \boldsymbol{x}) \end{bmatrix} \tag{12}$$

The second element in the TSLS estimator of $\boldsymbol{\beta}$, computed by using the sample within bandwidth h of the cutoff point $\boldsymbol{x}$, reads:

$$\hat{\tau}_{RD}(\boldsymbol{x}) = \boldsymbol{e}_2'[\boldsymbol{R'S}(\boldsymbol{S'S})^{-1}\boldsymbol{S'R}]^{-1}\boldsymbol{R'S}(\boldsymbol{S'S})^{-1}\boldsymbol{S'Y} \tag{13}$$

where $\boldsymbol{e}_2$ is a $(2+2d)\times 1$ vector of zeros with the second element equal to one, $\boldsymbol{R} = (\boldsymbol{R}_1', ..., \boldsymbol{R}_{n_h}')'$, $\boldsymbol{S} = (\boldsymbol{S}_1', ..., \boldsymbol{S}_{n_h}')'$, $\boldsymbol{Y} = (Y_1', ..., Y_{n_h}')'$, n_h denotes the total number of units that lie within bandwidth h of the cutoff point $\boldsymbol{x}$ for the fuzzy RD design.

Although appearing unlikely, expressions of $\hat{\tau}_{RD}(\boldsymbol{x})$ in (10) and (13) are numerically equivalent, see Imbens and Lemieux (2008), Imbens and Zajonc (2011). This coincidence implies that the (conventional) standard error of $\hat{\tau}_{RD}(\boldsymbol{x})$ can be derived by the TSLS procedure. With $\hat{\tau}_{RD}(\boldsymbol{x})$ and its standard error, the conventional t-test on the causal effect is thus applicable.[2]

If a kernel function other than the uniform kernel is adopted, then (6) - (9) need to be adjusted to include such a kernel. Correspondingly, $\hat{\tau}_{RD}(\boldsymbol{x})$ in (13) will be re-written as the weighted TSLS estimator, where the weight is determined by the adopted kernel. See Imbens and Zajonc (2011). Throughout this paper, we use the uniform kernel for ease of exposition.

[2]The expression of this conventional standard error is documented in the literature, see, e.g., Imbens and Lemieux (2008) for the scalar case.

2.3 Weak Identification in Fuzzy RD

A well-established literature on TSLS concerns the weak identification issue. That is, if instruments are only poorly correlated with endogenous variables, then the performance of the TSLS procedure is unreliable. In particular, the size of the conventional t-test for TSLS is distorted. See, e.g., Staiger and Stock (1997). Since the fuzzy RD estimator may also be viewed as a TSLS estimator, a natural question arises: Does weak identification exist in fuzzy RD designs?

As found by Feir et al. (2016), the answer is Yes. Specifically, when the variable for the treatment assignment rule (denoted by T_i above) is not strongly correlated with the actual treatment variable (denoted by W_i above), then weak identification of the RD estimand $\tau_{RD}(\boldsymbol{x})$ takes place. In other words, weak identification is likely to exist in fuzzy RD studies, if non-compliance with treatment assignment is so severe that $W^+(\boldsymbol{x}) - W^-(\boldsymbol{x})$ is not substantially different from zero, i.e., the discontinuity in receiving treatment is not clearly visible at the cutoff. Astonishingly, Feir et al. (2016) find that a large number of existing fuzzy RD studies seem to suffer the weak identification problem.[3]

To resolve this problem, Feir et al. (2016) provide a robust standard error for the fuzzy RD estimator. The robust standard error takes the potential weak identification into account while the conventional one does not. As a result, the t-test on the causal effect in fuzzy RD studies is still applicable, if this robust standard error is employed. Feir et al. (2016) show that such a robust t-test remains trustworthy, no matter whether weak identification exists, while the conventional t-test is only valid under strong identification.

2.4 Anderson-Rubin Test with Multiple Forcing Variables

The robust standard error proposed by Feir et al. (2016), however, only concerns the $d = 1$ case. When treatment assignment is determined by more than one forcing variable so $d > 1$, a robust approach in line with that in Feir et al. (2016) is similarly in demand. We thus propose

[3] Empirical fuzzy RD studies often present graphic evidence for the strong relationship of W_i and Z_i. However, the seemingly convincing graph is often not quantitatively sufficient to rule out weak identification.

such a robust method in this paper, which is applicable under $d \geq 1$.

Our insight comes from the well-established weak identification literature on TSLS. To conduct inference in the TSLS model, various robust tests have been suggested in this literature, including, e.g., Anderson and Rubin (1949), Kleibergen (2002), Moreira (2003). In particular, the Anderson and Rubin (1949) test is known to be powerful, if the model is just identified. Since the fuzzy RD estimator is a TSLS estimator with T_i as the instrumental variable for W_i in (11), we encounter just-identification and consequently adopt the Anderson and Rubin (1949) test.

To be specific, our interest lies in testing $H_0 : \tau_{RD}(\boldsymbol{x}) = \tau_{RD,0}$ against $H_1 : \tau_{RD}(\boldsymbol{x}) \neq \tau_{RD,0}$. Under H_0, we rewrite the auxiliary Equation (11) as follows:

$$Y_i - \tau_{RD,0} W_i = \boldsymbol{S}_i' \boldsymbol{\gamma} + u_i \tag{14}$$

The second element in $\boldsymbol{\gamma}$ (denoted by γ_2 below) is the coefficient for T_i, which equals zero under the null. Thus we can test whether γ_2 equals zero for testing H_0, after regressing $Y_i - \tau_{RD,0} W_i$ on $\boldsymbol{S}_i$. Such a test is known as the Anderson-Rubin test.

The ordinary least squares estimator of γ_2 in (14) reads:

$$\hat{\gamma}_2 = \boldsymbol{e}_2' (\boldsymbol{S'S})^{-1} \boldsymbol{S'} (\boldsymbol{Y} - \tau_{RD,0} \boldsymbol{W}) \tag{15}$$

where $\boldsymbol{W} = (W_1', ..., W_{n_h}')'$, and $\boldsymbol{e}_2$, $\boldsymbol{S}$, $\boldsymbol{Y}$ are defined above. In addition, the least squares estimator of the asymptotic variance of $\hat{\gamma}_2$ by White (1980) is written as:

$$\widehat{\mathbb{V}}_{\hat{\gamma}_2} = \boldsymbol{e}_2' \left(\frac{\boldsymbol{S'S}}{n_h} \right)^{-1} \left(\frac{1}{n_h - 2d - 2} \sum_{i=1}^{n_h} \hat{u}_i^2 \boldsymbol{S_i S_i'} \right) \left(\frac{\boldsymbol{S'S}}{n_h} \right)^{-1} \boldsymbol{e}_2 \tag{16}$$

where $\hat{u}_i$ is the regression residual of (14).

With the pieces above, our proposed test statistic is provided in the theorem below.

Theorem 1 *For the fuzzy RD design described above, assume:* $(Y_i, W_i, \boldsymbol{X}_i)$, $i = 1, ..., n$, *are* *i.i.d. with a regular joint distribution function; the RD bandwidth* $h = (h_1, ..., h_d)'$ *satisfies* $h_j \propto n^{-r}$, $j = 1, ..., d$, *and* r *is chosen in the manner such that* $\hat{\alpha}_{y_1}$, $\hat{\alpha}_{y_0}$, $\hat{\alpha}_{w_1}$, $\hat{\alpha}_{w_0}$ *are consistent estimators as* $n \to \infty$. *Under* $H_0 : \tau_{RD}(\boldsymbol{x}) = \tau_{RD,0}$*:*

$$AR(\tau_{RD,0}) = n_h \frac{\hat{\gamma}_2^2}{\widehat{\mathbb{V}}_{\hat{\gamma}_2}} \xrightarrow{d} \chi_1^2 \tag{17}$$

Theorem 1 directly results from the asymptotic normal distribution of $\hat{\gamma}_2$, the ordinary least squares estimator for Equation (14). The conditions imposed in the theorem are regular conditions for RD designs, see Feir et al. (2016) for the related discussion. The magnitude of h is commonly chosen to minimize the mean squared error of the fuzzy RD estimator: e.g., in the scalar case with $d = 1$, it is suggested that $1/5 < r < 1/3$ in Feir et al. (2016); in the two-dimensional case with $d = 2$, Imbens and Zajonc (2011) use $r = 1/6$ to compute the optimal bandwidth. As $n \to \infty$, the number of observations within the bandwidth h denoted by n_h also gets large, hence $AR(\tau_{RD,0})$ asymptotically converges to the χ^2 distribution with one degree of freedom.

Given Theorem 1, the following test for $H_0 : \tau_{RD}(\boldsymbol{x}) = \tau_{RD,0}$ against $H_1 : \tau_{RD}(\boldsymbol{x}) \neq \tau_{RD,0}$ has the asymptotic size that equals α, $0 < \alpha < 1$: reject H_0, if $AR(\tau_{RD,0})$ exceeds the $1 - \alpha$ quantile of the χ_1^2 distribution.

Remark 1: If the magnitude of h guarantees the consistency of $\hat{\alpha}_{y_1}$, $\hat{\alpha}_{y_0}$, $\hat{\alpha}_{w_1}$ and $\hat{\alpha}_{w_0}$, then γ_2 in the auxiliary Equation (14) equals zero.

Remark 2: A subtle point here is that the Anderson-Rubin statistic in Theorem 1 does not directly involve the fuzzy RD estimator $\hat{\tau}_{RD}(\boldsymbol{x})$. Consequently, the optimal choice of h for the AR test need not coincide with the one that intends to minimize the mean squared error of $\hat{\tau}_{RD}(\boldsymbol{x})$. It is practically convenient, however, to adopt the bandwidth suggested by the existing literature when implementing the proposed AR test, since the use of the same bandwidth facilitates the comparison of the AR test with the commonly used t-test.

3 Simulation

In this section, we conduct simple simulation experiments to compare the conventional t-test, the robust t-test of Feir et al. (2016) and our proposed AR test in fuzzy RD designs.

3.1 Fuzzy RD with $d = 1$

The data generation process (d.g.p.) considers a constant regression discontinuity effect:[4]

$$y_i = \tau w_i + u_{y_i},\ \ i = 1, ..., n.$$

where y_i is the observed outcome variable, w_i is the treatment variable. Let x_i be the scalar forcing variable and the treatment w_i is determined by:

$$w_i = 1\{u_{x_i} \le 0\} \times 1\{x_i < 0\} + 1\{u_{x_i} \le c\} \times 1\{x_i \ge 0\}$$

In addition, the errors u_{x_i} and u_{y_i} are jointly normal with the correlation coefficient ρ, and x_i also follows a normal distribution:

$$\begin{pmatrix} u_{y_i} \\ u_{x_i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),\ \ x_i \sim N(0, \sigma_x^2)$$

This RD design is fuzzy, since w_i depends not only on whether the forcing variable x_i exceeds the cutoff 0, but also the value of u_{x_i}. In particular, $w^+(0) - w^-(0) = \Phi(c) - \Phi(0)$ by this d.g.p., where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Hence the value of c drives the magnitude of the discontinuity of RD, and consequently the strength of identification. A small value of c leads to non-substantial discontinuity, which consequently induces the weak identification problem.

[4]To facilitate comparison, the d.g.p. is adopted from Feir et al. (2016), with only minor changes to make the d.g.p. consistent with the discussion in Section 2.

We consider the following values of $c \in \{10, 1, 0.1\}$, to vary the strength of identification. In addition, we set $\tau = 0$, $\sigma_x^2 = 1$, $\rho \in \{0.5, 0.99\}$ in d.g.p.. When implementing the conventional t-test, the robust t-test and the proposed AR test, we keep using the uniform kernel function; in addition, we consider two different bandwidth $h \in \{0.5, 1\}$ to check the sensitivity to the choice of h. The sample size n equals 2000.

Table 1 presents the actual sizes of the three tests, based on 2000 Monte Carlo replications. When $c = 10$ so identification is not very weak, all three tests appear to perform well, since their actual sizes are close to the nominal 5% and 10%. As c decreases, the conventional t-test exhibits severe size distortion; by contrast, actual sizes of the robust t-test and the proposed AR test remain close to the nominal ones. These findings prevail, as the value of ρ and the choice of bandwidth h vary. Overall, Table 1, consistent with Theorem 1, suggests that the AR test is robust to the magnitude of discontinuity in fuzzy RD designs.

Figure 1 presents the power curves of the three tests: the conventional t-test (dashed), the robust t-test (dotted) and the AR test (solid). We consider two scenarios: strong identification with $c = 10$ and weak identification with $c = 0.1$. When identification is strong, all three tests have similar power (almost perfectly overlapping in Panel (a) of Figure 1) and correct size. By contrast, when identification is weak, the conventional t-test (dashed) suffers size distortion, while the robust t-test (dotted) and AR test (solid) remain size-correct and have similar power plots.

3.2 Fuzzy RD with $d = 2$

The d.g.p. used for the scalar RD design above can be modified for the multi-dimensional case with $d = 2$.

Consider the 2×1 vector of forcing variables $x_i = (x_{1i}, x_{2i})'$, and the treatment w_i now reads:

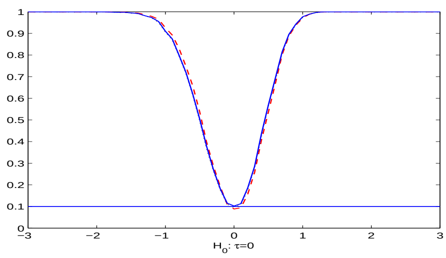
$$w_i = 1\{u_{x_i} \leq 0\} \times 1\{x_{1i} < 0, x_{2i} < 0\} + 1\{u_{x_i} \leq c\} \times 1\{x_{1i} \geq 0 \ or \ x_{2i} \geq 0\}$$

Table 1: Sizes of t-test, robust t-test and AR test in RD with $d = 1$

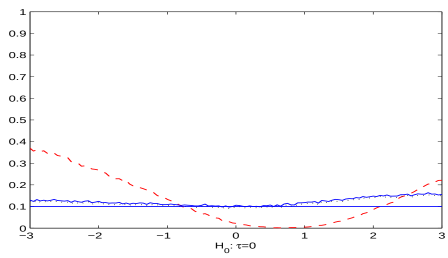
Panel A: $h = 0.5$							
		at 5%			at 10%		
ρ	c	t	robust t	AR	t	robust t	AR
0.5	10	0.044	0.048	0.057	0.105	0.108	0.115
0.5	1	0.035	0.041	0.044	0.069	0.091	0.093
0.5	0.1	0.010	0.048	0.056	0.022	0.100	0.106
0.99	10	0.044	0.047	0.052	0.087	0.092	0.099
0.99	1	0.044	0.042	0.047	0.074	0.088	0.094
0.99	0.1	0.116	0.056	0.057	0.172	0.110	0.112
Panel B: $h = 1$							
		at 5%			at 10%		
ρ	c	t	robust t	AR	t	robust t	AR
0.5	10	0.036	0.040	0.045	0.080	0.082	0.102
0.5	1	0.039	0.045	0.056	0.079	0.096	0.108
0.5	0.1	0.005	0.037	0.056	0.021	0.091	0.109
0.99	10	0.040	0.040	0.050	0.089	0.089	0.106
0.99	1	0.037	0.038	0.044	0.073	0.073	0.086
0.99	0.1	0.122	0.053	0.067	0.165	0.105	0.117

Note: This table reports the rejection frequencies of $H_0 : \tau = 0$ by three different tests (t, robust t and AR) in the scalar RD design, based on the average of 2000 Monte Carlo replications. The d.g.p. is described in the main text. Two choices of the bandwidth h are used in test implementation, with $h = 0.5$ for Panel A, $h = 1$ for Panel B. The value of c reflects the magnitude of discontinuity and thus the identification strength in the simulation. $\rho \in \{0.5, 0.99\}$ is the correlation coefficient of the used error terms in d.g.p..

Figure 1: Power curves of t-test, robust t-test and AR test in RD with $d = 1$



(a) Strong Identification with $c = 10$



(b) Weak Identification with $c = 0.1$

Note: This figure compares the power plots of the conventional t-test (dashed), the robust t-test (dotted) and the AR test (solid) at the nominal 10% for $H_0 : \tau = 0$ in the scalar RD design. The value of c reflects the magnitude of discontinuity and thus the identification strength: $c = 10$ for strong identification in (a), and $c = 0.1$ for weak identification in (b). The d.g.p. is described in the main text, with $h = 0.5$ and $\rho = 0.5$. The number of replications is 10000.

with

$$\begin{pmatrix} x_{1i} \\ x_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right)$$

and the other settings of d.g.p. remain the same as in the scalar case above.

With the data generated from this d.g.p., we implement the proposed AR test, and present its size by simulation in Table 2. Since the robust t-test of Feir et al. (2016) is not applicable under $d = 2$, we compare the AR test with the conventional t-test only. In addition, we slightly increase the bandwidth $h \in \{1, 2\}$, to further check the sensitivity of the simulation outcome to the choice of bandwidth.

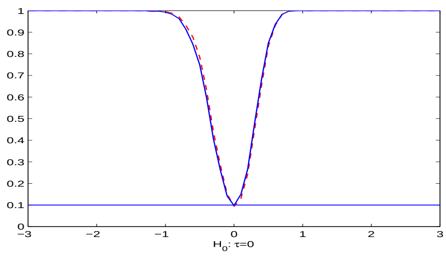
Table 2: Sizes of t-test and AR test in RD with $d = 2$

Panel A: $h = 1$					
		at 5%		at 10%	
ρ	c	t	AR	t	AR
0.5	10	0.042	0.057	0.090	0.101
0.5	1	0.032	0.044	0.064	0.084
0.5	0.1	0.009	0.051	0.023	0.103
0.99	10	0.054	0.051	0.093	0.105
0.99	1	0.053	0.049	0.082	0.094
0.99	0.1	0.125	0.050	0.171	0.103
Panel B: $h = 2$					
		at 5%		at 10%	
ρ	c	t	AR	t	AR
0.5	10	0.038	0.043	0.089	0.098
0.5	1	0.037	0.043	0.083	0.098
0.5	0.1	0.011	0.055	0.026	0.112
0.99	10	0.051	0.051	0.098	0.101
0.99	1	0.056	0.059	0.097	0.105
0.99	0.1	0.117	0.049	0.161	0.092

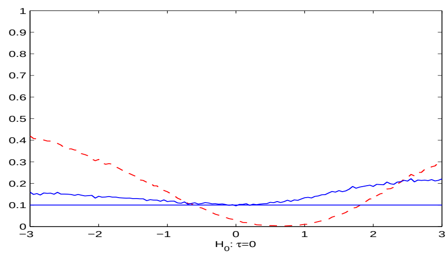
Note: This table reports the rejection frequencies of $H_0 : \tau = 0$ by two different tests (t and AR) in the multi-dimensional RD design, based on the average of 2000 Monte Carlo replications. The d.g.p. is described in the main text. Two choices of the bandwidth h are used in test implementation, with $h = 1$ for Panel A, $h = 2$ for Panel B. The value of c reflects the magnitude of discontinuity and thus the identification strength in the simulation. $\rho \in \{0.5, 0.99\}$ is the correlation coefficient of the used error terms in d.g.p..

Similar to Table 1, Table 2 shows that (i) when identification is not very weak, both the conventional t-test and AR test have actual sizes close to the nominal 5% and 10%; (ii) when identification gets weak as c decreases, the conventional t-test is not size-correct while the AR test remains trustworthy. Furthermore, similar to Figure 1, Figure 2 presents the power curve of the AR test under strong and weak identification with $d = 2$. Like the scalar case in Figure 1, Figure 2 suggests the proposed AR test performs well in size and power in the multi-dimensional case.

Figure 2: Power curves of t-test and AR test in RD with $d = 2$



(a) Strong Identification with $c = 10$



(b) Weak Identification with $c = 0.1$

Note: This figure compares the power plots of the conventional t-test (dashed) and the AR test (solid) at the nominal 10% for $H_0 : \tau = 0$ in the multi-dimensional RD design. The value of c reflects the magnitude of discontinuity and thus the identification strength: $c = 10$ for strong identification in (a), and $c = 0.1$ for weak identification in (b). The d.g.p. is described in the main text, with $h = 2$ and $\rho = 0.5$. The number of replications is 10000.

To summarize, the simulation findings in Tables 1 - 2 and Figures 1 - 2 are consistent with our discussion in Section 2.

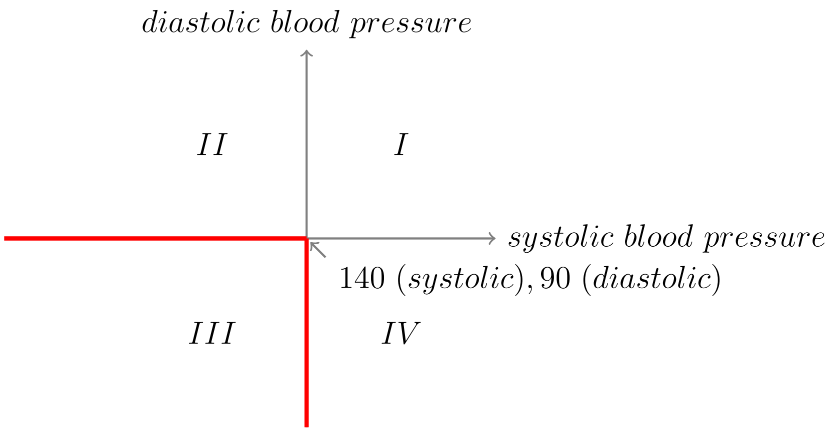
4 Application

In this section, we use hypertension diagnosis as an example for illustration.

4.1 RD Design of Hypertension

Hypertension is one of the most critical risk factors of major chronic diseases, and it affects approximately one third of the world's population (World Health Statistics 2012). According to the American Heart Association, an individual is considered to have hypertension if his/her systolic blood pressure is equal to or above 140 $mmHg$, and/or if the diastolic blood pressure is equal to or above 90 $mmHg$.

Figure 3: Graphic illustration of hypertension diagnosis



Graphically, individuals in the quadrants I, II and IV of Figure 3 are diagnosed with hypertension, while individuals in III are not. The boundary of III thus serves as the boundary of the regression discontinuity design of hypertension with $d = 2$ forcing variables, the systolic and diastolic blood pressures. The treatment region includes I, II and IV (as well as the boundary), while the control region is III.

Once individuals are diagnosed with hypertension, will they subsequently change their health behavior, such as food demand? This question has been studied in the literature: for example, Zhao et al. (2013) find that the provision of hypertension information substantially decreases fat intake (-7.7 g per day). The regression discontinuity analysis in Zhao et al. (2013), however, uses only the systolic blood pressure as the forcing variable for RD. In particular, Zhao et al. (2013) consider a sharp scalar RD design of hypertension diagnosis, i.e., an individual whose systolic blood pressure is equal to or greater than 140 $mmHg$ is considered in the treatment group. In Figure 3, Zhao et al. (2013) thus consider I, IV as the treatment region, and II, III as the control region.

We revisit the question studied in Zhao et al. (2013) by adopting a RD design with two forcing variables, to be consistent with the common practice of hypertension diagnosis. To be specific, $\boldsymbol{X}_i = (\boldsymbol{X}_{1i}, \boldsymbol{X}_{2i})' = (systolic_i, diastolic_i)'$, and the treatment assignment rule is $T_i = \mathbf{1}(systolic_i \geq 140\ mmHg\ or\ diastolic_i \geq 90\ mmHg)$.

Due to the noise in medical practice, not all individuals in the treatment region are aware that they are diagnosed with hypertension; similarly, there are also individuals who have blood pressures below the cutoff point, but mistakenly think they are diagnosed with hypertension. Johnston et al. (2009) show the gap between subjective and objective measures of health is large, and false reporting is prevalent for low-income groups. In the RD design of hypertension, consequently, the treatment variable W_i for the awareness of hypertension does not coincide with T_i. For the reasons above, we face a fuzzy RD design with $d = 2$ forcing variables, and the targeted outcome variable Y_i is fat intake.

4.2 Data

The dataset of $(Y_i, W_i, \boldsymbol{X}_i)$ is from the China Health and Nutrition Survey (CHNS, 1997, 2000 and 2004), which is also used in Zhao et. al. (2013).[5] The description of the survey and the data can be found in Zhao et. al. (2013) as well as the survey website.[6] We use the self-reported hypertension status as the treatment variable, the same-period fat intake as the outcome variable, and the one-period lagged systolic pressure and diastolic pressure as two forcing variables. In total, 14,154 observations are used in our empirical analysis. The summary statistics are provided in Table 3.

[5] We thank the authors for providing the codes for cleaning the data and replicating their results.

[6] $http://www.cpc.unc.edu/projects/china.$

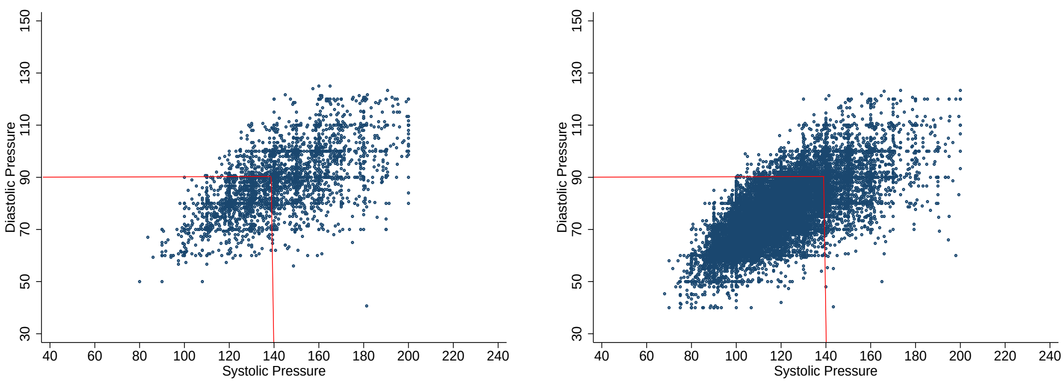
Table 3: Summary statistics

Variable	Mean	SD	Median	Min	Max
Y_i : fat intake	7.26	5.30	6.58	0.08	340.16
W_i : awareness of hypertension	0.10	0.30	0	0	1
$\boldsymbol{X}_{1i}$: systolic blood pressure	121.28	17.72	120	74.67	200
$\boldsymbol{X}_{2i}$: diastolic blood pressure	78.37	10.93	79.33	40.33	125

Note: fat intake, g per day; awareness of hypertension, 1 (Yes) or 0 (No); systolic and diastolic blood pressures, $mmHg$.

To provide a graphic overview, we plot the data of forcing variables in Figure 4 below. The left Panel (a) of Figure 4 plots the systolic and diastolic pressures of those who report they have hypertension ($W_i = 1$), while the right Panel (b) are for those who think they do not have it ($W_i = 0$). Both (a) and (b) in Figure 4 show that the RD design is fuzzy, because noncompliance with the treatment assignment rule commonly exists. The comparison of (a) and (b) also indicates that $W^+(\boldsymbol{x}) - W^-(\boldsymbol{x})$ is non-zero but likely small, because individuals in the control region appear slightly less likely to report the self-awareness of hypertension, compared to those in the treatment region.

Figure 4: Forcing variables in treatment and control groups



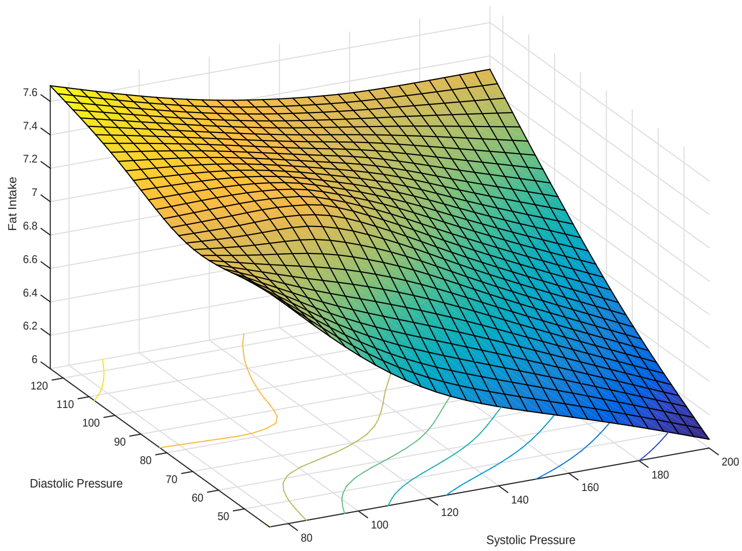
(a) Treatment Group

(b) Control Group

Note: This figure plots the systolic and diastolic blood pressures of the individuals used in RD analysis. Panel (a) uses the individuals whose self-reported hypertension status is 1 (Yes); Panel (b) uses the individuals whose self-reported hypertension status is 0 (No). The boundary of hypertension diagnosis is also plotted.

The outcome variable for fat intake is plotted in Figure 5, given the corresponding systolic and diastolic blood pressures. It is not clearly visible from Figure 5 whether there is a sudden decrease in fat intake near the boundary of hypertension ($systolic$: 140 $mmHg$, $diastolic$: 90 $mmHg$). If provision of the hypertension information does cause fat intake to substantially fall, then we expect to see such a sudden decrease. Unlike in scalar RD designs, however, eyeballing jumps in multi-dimensional RD is known to be difficult. We thus turn to the proposed AR test to investigate the impact of hypertension awareness on fat intake.

Figure 5: Fat intake by systolic and diastolic blood pressures



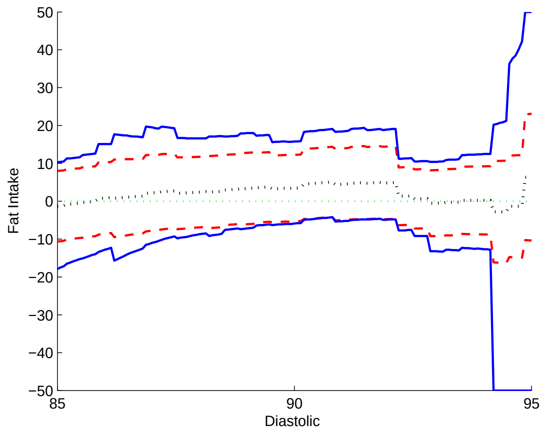
Note: This figure plots the amount of fat intake (g per day) of the individuals used in RD analysis, given their systolic and diastolic blood pressures ($mmHg$).

4.3 Findings by AR

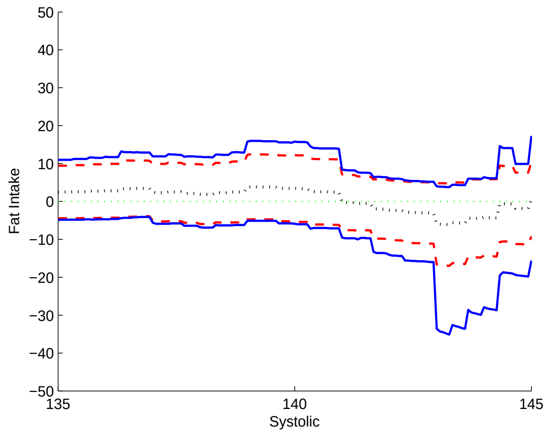
For each point denoted by $\boldsymbol{x}$ on the boundary of hypertension diagnosis, we can conduct the AR test described in Section 2 for $H_0 : \tau_{RD}(\boldsymbol{x}) = \tau_{RD,0}$. Values of $\tau_{RD,0}$ that are not rejected by the AR test thus constitute a confidence interval of $\tau_{RD}(\boldsymbol{x})$.

We note that in Table 3, fat intake lies between 0.08 and 340.16, which provides the guidance of the possible value of $\tau_{RD,0}$, the change in fat intake as a result of hypertension diagnosis. When implementing the AR test in the empirical application, we can thus consider $[-340.16, 340.16]$ as the possible range of $\tau_{RD,0}$. The outcome of our application is presented in Figure 6.

Figure 6: Treatment effects on the boundary of hypertension



(a) Systolic=140 $mmHg$



(b) Diastolic=90 $mmHg$

Note: This figure plots the estimated treatment effects (dotted) and the associated 95% confidence intervals by the conventional t-test (dashed) and the proposed AR test (solid), for points on the boundary of the fuzzy RD design of hypertension diagnosis. The uniform kernel is adopted, and the bandwidth is selected using the method in Imbens and Zajonc (2011), with $h = (22.98, 14.14)'$. The outcome variable is fat intake.

The left Panel (a) of Figure 6 presents the estimated treatment effects on fat intake at the points with $systolic_i = 140$ and $diastolic_i \in [85, 95]$, while the right Panel (b) is for points with $diastolic_i = 90$ and $systolic_i \in [135, 145]$. All of these points lie on the boundary of hypertension, and they are adopted in our application because the number of observations lying close to these points is large, which facilitates empirical analysis. Given a point $\boldsymbol{x}$ on the boundary, we compute the conditional treatment effect $\hat{\tau}_{RD}(\boldsymbol{x})$ (dotted), as well as the associated 95% confidence intervals by the conventional t-test (dashed) and the proposed AR test (solid), both are discussed in Section 2. The robust t-test of Feir et al. (2016) is not adopted, since our application involves two forcing variables.

Figure 6 shows that confidence intervals by t and AR largely overlap but do not coincide. There are also points on the boundary for which confidence intervals by the AR test substantially differ from those by the conventional t-test. This difference signals that the identification strength is not sufficiently strong. In addition, since both confidence intervals cover zero, we can not reject that the treatment effect is zero, i.e., fat intake does not necessarily change, if the information of hypertension is received. Furthermore, since confidence intervals in Panel (a) of Figure 6 contain -7.7, the point estimate of the causal effect of hypertension defined by the systolic pressure in Zhao et al. (2013), we do not reject fat intake decreases by this amount either.

5 Conclusions

The fuzzy regression discontinuity estimator could also be viewed as an instrumental variable estimator, so the Anderson-Rubin test from the instrumental variable literature is applicable to regression discontinuity. This test is particularly appealing, when fuzzy regression discontinuity designs are associated with multiple forcing variables and/or non-substantial magnitude of discontinuity in the probability of receiving treatment. The regression-based test is also attractive from the practical point of view, since it is easy to implement. By applying the test to study the impact of hypertension on fat intake, we can not reject that fat intake remains unaffected by hypertension awareness, and we can not reject fat intake decreases by the amount reported in Zhao et al. (2013) either.

The Anderson-Rubin test procedure we describe in this paper does have limitations. First, we do not discuss the choice of bandwidth, but rely on the existing literature to suggest its value. Second, if the bandwidth or kernel used for each of the four components of the fuzzy regression discontinuity estimator varies (which is in principle feasible), then the coincidence of the regression discontinuity estimator with the instrumental variable estimator breaks down;

consequently, our proposed test is no longer applicable. Nevertheless, with these limitations, the proposed test still helps resolve the inference on causal effects, as shown by our simulation as well as the empirical application in this paper. We leave the limitations for future research.

References

[1] TW Anderson and H. Rubin. Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics*, pages 46–63, 1949.

[2] Joshua D Angrist and Victor Lavy. Using maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*, 114(2):533–575, 1999.

[3] Sandra E Black. Do better schools matter? parental valuation of elementary education. *Quarterly Journal of Economics*, 114(2):577–599, 1999.

[4] John Bound, David A Jaeger, and Regina M Baker. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430):443–450, 1995.

[5] Wen S Chern, Edna T Loehman, and Steven T Yen. Information, health risk beliefs, and the demand for fats and oils. *Review of Economics and Statistics*, pages 555–564, 1995.

[6] Sarah R Cohodes and Joshua S Goodman. Merit aid, college quality, and college completion: Massachusetts' adams scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics*, 6(4):251–285, 2014.

[7] Peter H Egger and Georg Wamser. The impact of controlled foreign company legislation on real investments abroad. a multi-dimensional regression discontinuity design. *Journal of Public Economics*, 129:77–91, 2015.

[8] Karen Eggleston, Ang Sun, and Zhaoguo Zhan. The Impact of Rural Pensions in China on Labor Migration. *World Bank Economic Review*, forthcoming, 2016.

[9] Jianqing Fan and Irene Gijbels. *Local polynomial modelling and its applications*. CRC Press, 1996.

[10] Donna Feir, Thomas Lemieux, and Vadim Marmer. Weak identification in fuzzy regression discontinuity designs. *Journal of Business and Economic Statistics*, 34(2):185–196, 2016.

[11] Stefano Gagliarducci and M Daniele Paserman. Gender interactions within hierarchies: evidence from the political arena. *Review of Economic Studies*, 79(3):1021–1052, 2012.

[12] Jinyong Hahn, Petra Todd, and Wilbert Van der Klaauw. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209, 2001.

[13] Guido Imbens and Karthik Kalyanaraman. Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies*, 79(3):933–959, 2012.

[14] Guido Imbens and Tristan Zajonc. Regression discontinuity design with multiple forcing variables. 2011.

[15] Guido W Imbens and Thomas Lemieux. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635, 2008.

[16] David W. Johnston, Carol Propper, and Michael A. Shields. Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient. *Journal of Health Economics*, 28(3):540 – 552, 2009.

[17] F. Kleibergen. Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, pages 1781–1803, 2002.

[18] David S Lee. Randomized experiments from non-random selection in us house elections. *Journal of Econometrics*, 142(2):675–697, 2008.

[19] David S Lee and Thomas Lemieuxa. Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355, 2010.

[20] Marcelo J Moreira. A conditional likelihood ratio test for structural models. *Econometrica*, 71(4):1027–1048, 2003.

[21] World Health Organization. World health statistics. 2012.

[22] John P Papay, John B Willett, and Richard J Murnane. Extending the regression-discontinuity approach to multiple assignment variables. *Journal of Econometrics*, 161(2):203–207, 2011.

[23] Jack Porter. Estimation in the regression discontinuity model. 2003.

[24] Jutta Roosen, Stéphan Marette, Sandrine Blanchemanche, and Philippe Verger. Does health information matter for modifying consumption? a field experiment measuring the impact of risk information on fish consumption. *Review of Agricultural Economics*, 31(1):2–20, 2009.

[25] Connan Snider and Jonathan W Williams. Barriers to entry in the airline industry: A multidimensional regression-discontinuity analysis of air-21. *Review of Economics and Statistics*, 97(5):1002–1022, 2015.

[26] Douglas Staiger and James H Stock. Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586, 1997.

[27] Wilbert Van der Klaauw. Estimating the effect of financial aid offers on college enrollment: A regression–discontinuity approach. *International Economic Review*, 43(4):1249–1287, 2002.

[28] H. White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838, 1980.

[29] Meng Zhao, Yoshifumi Konishi, and Paul Glewwe. Does information on health status lead to a healthier lifestyle? evidence from china on the effect of hypertension diagnosis on food consumption. *Journal of Health Economics*, 32(2):367–385, 2013.