Regression Discontinuity Design with Endogenous Covariates

Kyoo il Kim*

Abstract   Empirical researchers often include observed covariates in the Wald-type implementation of Regression Discontinuity Design (RDD) estimators. When those included covariates are endogenous, we find that the resulting RDD estimator suffers from a larger asymptotic bias than the estimator with exogenous covariates but it is still consistent. We further show that the order of bias increase due to the endogeneity is the same order of bias reduction due to the inclusion of relevant endogenous covariates.

Keywords   Regression Discontinuity Design, Treatment Effect, Asymptotic Bias, Endogeneity, Kernel estimations

JEL Classification   C2, C21

* Correspondence: Kyoo il Kim, Department of Economics, University of Minnesota, 4-129 Hanson Hall, 1925 Fourth Street South, MN 55455, Tel: +1 612 625 6793, E-mail: kyookim@umn.edu.

1. INTRODUCTION

The Regression Discontinuity (RD) design is a quasi-experimental design that provides a useful source of identifying effects of a treatment, which is often exploited in economic and other applications. The RD design has the key feature of discontinuity in receiving treatment due to one or more underlying variables, named as *forcing variable(s)* because it governs the treatment. The earliest application of the RD design even goes back to Thistlethwaite and Campbell (1960). For more recent applications using the RD design see Chay and Greenstone (2005), Greenstone and Gallagher (2008), and Black (1999) who have exploited either a discontinuity in the application of a regulation or a structural break due to a boundary.

Econometric theories on Regression Discontinuity Design (RDD) estimators typically do not include observed covariates or attributes other than the forcing variable(s) that governs the discontinuity. However, many empirical researchers using the RDD often include such covariates and implement the estimator using a Wald-type regression. This discrepancy between theory and practice arises due to the nonparametric nature of the RDD estimator. In purely theoretical point of view, being conditioned on the forcing variable close to the *cut off* point, including other covariates in the regression should have little effect on the RDD estimator since the indicator of having a treatment or not is independent of other covariates at the cut off. However, in practice, we almost always have to include data points with values of the forcing variable *not being very close to* the cut off. In this case, including other covariates improves the RDD estimator because it reduces some bias caused by the inclusion of such data points and it reduces the variance when the covariates are correlated with the potential outcomes (see Imbens and Lemieux 2008).

Here our main question is about how the positive effects of including covariates will change if the covariates are endogenous due to e.g. omitted covariates. In most applications economic agents have superior information which the agents can utilize for their decisions compared to econometricians. This suggests it is inevitable that some unobserved covariates to researchers but observed by the economic agents exist. Moreover these unobserved covariates tend to be correlated with observed covariates and hence endogeneity is the rule rather than the exception. For example, Chay and Greenstone (2005) exploits quasi-random variation in EPA air quality attainment status to measure the effect of air pollution reflected on housing prices. Data sets on housing, however, typically do not report features such as the curb appeal or home repair status, both of which may be important to home buyers. Here suppose that the unobserved

covariates (to econometricians) include the curb appeal of a home. Exogeneity assumption would imply that the expected value of curb appeal is the same for small homes in low income neighborhoods as it is for very high priced homes in exclusive neighborhoods. However, it makes more sense to expect desirable omitted covariates to be positively correlated with desirable observed covariates, which violates the exogeneity assumption.

Our starting point for analysis is to note that even though the RDD estimator is a nonparametric one by design, empirical researchers typically implement the estimator using a Wald-type regression (see e.g. Lee and Lemieux 2010 for detailed discussion and references therein), which becomes nothing but a 2SLS estimation. This is totally innocuous because two estimators are shown to be numerically identical when we do not include other covariates (see Hahn, Todd, and Van der Klaauw (HTV) 2001 and Imbens and Lemieux 2008). However, it does not mean that two estimators have the same asymptotic properties. A Wald estimator is unbiased by assumption under standard exogeneity conditions while the nonparametric RDD estimator such as kernel estimator or local linear estimator has asymptotic biases. This means that even though we implement the RDD estimation by a Wald regression, inference must be done in the nonparametric context. Interestingly we find that the equivalence result no more holds if we include covariates in the RDD estimation. We then provide explicit relationships between a RDD estimator obtained by a Wald-type regression (we call this estimator as *Wald estimator* hereafter) and the nonparametric RDD estimator such as kernel estimator or local linear estimator.

We then show that the RDD estimator is still consistent as long as some continuity conditions are maintained. We also show that if we include observations with values of the forcing variable not close to the cut off and do not correctly handle the endogeneity, then the RDD estimator suffers from a larger asymptotic bias though it is still consistent. This suggests that unless a chosen bandwidth of the RDD estimator is small enough, the resulting estimator will suffer from a considerable bias under endogeneity. Interestingly we find that when we include other observable covariates, it can potentially reduce the bias of the treatment effect estimator. However, when those observable covariates are endogenous, the order of bias due to the endogeneity is the same order of bias reduction. Moreover, we find that the coefficient estimators on endogenous covariates will be inconsistent although the treatment effect estimator is consistent.

2. RDD WITH ENDOGENOUS COVARIATES

We focus on the binary treatment with a status indicator, w_i of individual $i = 1, \ldots, n$. Let y_i denote the outcome of the treatment. The treatment status w_i depends on a forcing variable z_i. With a sharp RD design, the treatment variable w_i is a deterministic function of the forcing variable z_i such as $w_i = 1\{z_i \geq z_0\}$ with the cut off value of z_i equal to z_0. With a fuzzy RD design, which is our focus, w_i is a random variable given z_i and we have $p(z) = E[w_i|z_i = z]$ such that $p(z)$ becomes discontinuous at $z = z_0$. When other covariates x_i are available, one can obtain the treatment effects (e.g. $E[\beta_i]$ below) from the equation

$$y_i = \alpha + w_i \beta_i + x_i' \pi + u_i$$

where in our situation an unobservable covariate u_i - that affects the outcome - can be potentially correlated with x_i, i.e., x_i is endogenous. Note that the relationship between w_i and u_i does not need to be specified in the RD design.

To simplify our discussion, we assume the treatment effect $\beta_i = \beta$ is constant for all i (when β_i is heterogeneous, we can identify e.g. the average treatment effect instead) and analyze whether β is identified under a similar set of assumptions with Hahn, Todd, and Van der Klaauw (2001). We first assume z_i is a valid forcing variable

Assumption 1. *(RD): (i) The limits $w^+ \equiv \lim_{z \to z_0^+} E[w_i|z_i = z]$ and $w^- \equiv \lim_{z \to z_0^-}$ $E[w_i|x_i = z]$ exist and (ii) $w^+ \neq w^-$.*

This assumption RD means the probability of receiving treatment is discontinuous at z_0. Next we assume two conditional mean functions are continuous at the cut off point.[1]

Assumption 2. *(Continuity) $E[x_i|z_i = z]$ and $E[u_i|z_i = z]$ are continuous in z at z_0.*

We observe that the treatment effect β is nonparametrically identified as long as above two conditions hold. Define $y^+ \equiv \lim_{z \to z_0^+} E[y_i|z_i = z]$ and $y^- \equiv \lim_{z \to z_0^-} E[y_i|z_i = z]$.

Theorem 1. *Suppose β_i is constant as β and suppose Assumptions 1-2 hold. Then, we have $\beta = \frac{y^+ - y^-}{w^+ - w^-}$.*

---

[1] Frölich (2007) also studies RDD estimations with covariates where the continuity assumption on the covariates we make here (Assumption **Continuity**) may not hold but he does not consider endogeneity of covariates as we do in this paper.

This identification result is obvious from the theorem 1 of HTV due to the continuity assumption imposed on $E[x_i|z_i = z]$ and $E[u_i|z_i = z]$ at $z = z_0$. We provide the following proof for completeness.

Proof. Note that the mean difference in outcomes for individuals above and below the discontinuity becomes

$$
\begin{aligned}
&E[y_i|z_i = z_0 + e] - E[y_i|z_i = z_0 - e] \\
=\ &\beta\{E[w_i|z_i = z_0 + e] - E[w_i|z_i = z_0 - e]\} \\
&+E[x_i'\pi + u_i|z_i = z_0 + e] - E[x_i'\pi + u_i|z_i = z_0 - e].
\end{aligned}
$$

Under Assumption **Continuity**, after we take the limit $e \to 0^+$, we obtain

$$
y^+ - y^- = \beta\{w^+ - w^-\}.
$$

Then by Assumption **RD**, the conclusion follows. $\square$

3. ESTIMATION AND ASYMPTOTIC BIAS

The identification result suggests we can consistently estimate the treatment effect by replacing $(y^{+(-)}, w^{+(-)})$ in β with their consistent nonparametric estimators $(\widehat{y}^{+(-)}, \widehat{w}^{+(-)})$ as

$$
\widehat{\beta} = (\widehat{y}^+ - \widehat{y}^-)/(\widehat{w}^+ - \widehat{w}^-).
$$

One can estimate $(y^{+(-)}, w^{+(-)})$ using various kernel methods and HTV and Imbens and Lemieux (2008) show that the local linear estimators have a better bias property on boundaries than a kernel estimator. With a bandwidth parameter h_n tending to zero as the sample size n tends to infinity, the local linear estimator has the asymptotic bias of the order h_n^2 while a kernel estimator has the asymptotic bias of the order h_n.

Instead of this nonparametric RDD estimator $\widehat{\beta}$ from the identification result, empirical researchers often implement the RDD estimator using a Wald regression based on 2SLS intuition from the equations: for observations i such that z_i is close to the cut off z_0,

$$
\begin{aligned}
y_i &= \alpha + w_i\beta + x_i'\pi + u_i & (1) \\
w_i &= \alpha_w + \delta \cdot 1(z_i > z_0) + x_i'\pi_w + \varepsilon_i & (2)
\end{aligned}
$$

where, in addition to the potential endogeneity of w_i, x_i is also endogenous as $E[x_iu_i|z_i = z_0] \neq 0$. Here for transparency we impose that the coefficients on x_i

are constants while allowing for the correlation between w_i and x_i or between z_i and x_i. One can also allow for heterogeneous π by assuming π_i and x_i are independent conditional on z_i (at least around z_0). Below we analyze the properties of the Wald estimators of (1)-(2) that are equivalent to the nonparametric RDD estimators such as kernel estimator and local linear estimator.

3.1. WALD ESTIMATOR EQUIVALENT TO KERNEL RDD ESTIMATOR

Let $\mathscr{Z}$ be a set of indices of i denoting the subsample of z_i such that $z_0 - h_n < z_i < z_0 + h_n$ with the bandwidth $h_n \to 0$ and define $c_i = 1(z_i > z_0)$. We first consider the relationship between the Wald estimator equivalent to a kernel estimator and the kernel RDD estimator. As (1)-(2) the Wald estimator is obtained by two stage least squares (2SLS) or IV estimation of

$$y_i = \alpha + w_i \beta + x_i' \pi + u_i$$

applied to the subsample $\mathscr{Z}$ where c_i is used as an excluded instrumental variable (IV) for w_i.

This Wald estimator is the most common type in the RDD literature. When x_i is not included in the regression, the Wald estimator is given by

$$\widetilde{\beta} = \frac{\widetilde{y}^+ - \widetilde{y}^-}{\widetilde{w}^+ - \widetilde{w}^-}$$

where $\widetilde{y}^{+(-)}$ and $\widetilde{w}^{+(-)}$ are the uniform kernel conditional mean estimators such that for $r = y$ or w,

$$\widetilde{r}^+ = \frac{\sum_{i \in \mathscr{Z}} c_i r_i}{\sum_{i \in \mathscr{Z}} c_i} \text{ and } \widetilde{r}^- = \frac{\sum_{i \in \mathscr{Z}} (1 - c_i) r_i}{\sum_{i \in \mathscr{Z}} (1 - c_i)}. \tag{3}$$

Therefore, the Wald estimator and the kernel RDD estimator are numerically identical in this case (see HTV).

Next we derive the Wald estimator when x_i is included in the estimation (assuming x_i is scalar without loss of generality).

Theorem 2. *1. Let $\widetilde{y}^{+(-)}$ and $\widetilde{w}^{+(-)}$ be the uniform kernel conditional mean estimators defined in (3). Also let $\widetilde{x}^{+(-)}$ be the uniform kernel conditional mean estimators similarly defined as in (3). Then the Wald treatment effect estimator of (1)-(2), denoted by $\widetilde{\beta}^*$, becomes*

$$\widetilde{\beta}^* = \frac{\widetilde{y}^+ - \widetilde{y}^- - \frac{\sum_{i \in \mathscr{Z}} x_i y_i}{\sum_{i \in \mathscr{Z}} x_i^2} (\widetilde{x}^+ - \widetilde{x}^-) - \frac{\sum_{i \in \mathscr{Z}} x_i}{\sum_{i \in \mathscr{Z}} x_i^2} \left((\widetilde{y}^+ - \widetilde{y}^-) \widetilde{x}^- - \widetilde{y}^- (\widetilde{x}^+ - \widetilde{x}^-) \right)}{\widetilde{w}^+ - \widetilde{w}^- - \frac{\sum_{i \in \mathscr{Z}} x_i w_i}{\sum_{i \in \mathscr{Z}} x_i^2} (\widetilde{x}^+ - \widetilde{x}^-) - \frac{\sum_{i \in \mathscr{Z}} x_i}{\sum_{i \in \mathscr{Z}} x_i^2} \left((\widetilde{w}^+ - \widetilde{w}^-) \widetilde{x}^- - \widetilde{w}^- (\widetilde{x}^+ - \widetilde{x}^-) \right)}.$$

2. Let $\widetilde{y}_i = \widetilde{\alpha} + \widetilde{\beta} w_i$ *be the fitted value from the Wald estimation of (1)-(2) without including* x_i *where* $(\widetilde{\alpha}, \widetilde{\beta})$ *are the resulting Wald estimators. Then we further obtain*

$$\widetilde{\beta}^* - \widetilde{\beta}$$

$$= -(\widetilde{x}^+ - \widetilde{x}^-)\left\{\frac{\sum_{i \in \mathscr{Z}} x_i(y_i - \widetilde{y}_i)}{\sum_{i \in \mathscr{Z}} x_i^2}\right\} / \left(\widetilde{w}^+ - \widetilde{w}^- + O_p(h_n)\right) \tag{4}$$

$$= -(\widetilde{x}^+ - \widetilde{x}^-)\left\{\pi + (\alpha - \widetilde{\alpha})\frac{\sum_{i \in \mathscr{Z}} x_i}{\sum_{i \in \mathscr{Z}} x_i^2} + (\beta - \widetilde{\beta})\frac{\sum_{i \in \mathscr{Z}} x_i w_i}{\sum_{i \in \mathscr{Z}} x_i^2} + \frac{\sum_{i \in \mathscr{Z}} x_i u_i}{\sum_{i \in \mathscr{Z}} x_i^2}\right\}$$

$$/ \left(\widetilde{w}^+ - \widetilde{w}^- + O_p(h_n)\right). \tag{5}$$

See Appendix A for the proof.

The formula (4-5) in Theorem 2 implies several interesting points to note. We focus on the numerator in the bias terms of (4) because the denominator will converge to a nonzero term $w^+ - w^- \neq 0$ under the assumption RD. First, we can interpret $\frac{\sum_{i \in \mathscr{Z}} x_i(y_i - \widetilde{y}_i)}{\sum_{i=\mathscr{Z}} x_i^2}$ in (4) as the OLS coefficient from the regression of $y_i - \widetilde{y}_i$ on x_i where $y_i - \widetilde{y}_i$ is the residual of the regression of y_i on w_i only. Therefore, the term $(\widetilde{x}^+ - \widetilde{x}^-)\left\{\frac{\sum_{i \in \mathscr{Z}} x_i(y_i - \widetilde{y}_i)}{\sum_{i=\mathscr{Z}} x_i^2}\right\}$ in (4) captures the bias reduction due to the inclusion of x_i. However, this bias reduction becomes negligible as $h_n \to 0$ as long as the continuity condition, $\lim_{z \to z_0^+} E[x_i|z_i = z] = \lim_{z \to z_0^-} E[x_i|z_i = z]$ holds since $\widetilde{x}^+ - \widetilde{x}^- = O_p(h_n)$. The first term $-(\widetilde{x}^+ - \widetilde{x}^-)\pi$ in (5) dominates the second term because

$$-(\widetilde{x}^+ - \widetilde{x}^-)\left\{(\alpha - \widetilde{\alpha})\frac{\sum_{i \in \mathscr{Z}} x_i}{\sum_{i \in \mathscr{Z}} x_i^2} + (\beta - \widetilde{\beta})\frac{\sum_{i \in \mathscr{Z}} x_i w_i}{\sum_{i \in \mathscr{Z}} x_i^2}\right\} = O_p(h_n^2).$$

Next note that the last term in (5) captures the bias due to the endogeneity of x_i. When x_i is endogenous, we have

$$\left(\widetilde{x}^+ - \widetilde{x}^-\right)\frac{\sum_{i \in \mathscr{Z}} x_i u_i}{\sum_{i \in \mathscr{Z}} x_i^2} = O_p(h_n)\frac{E[x_i u_i|z_0]}{E[x_i^2|z_0]} + o_p(h_n) \text{ and } E[x_i u_i|z_0] \neq 0.$$

This implies that the order of the asymptotic bias due to endogeneity is $O(h_n)$, which is the same order of bias reduction, $O(h_n)$ when x_i is endogenous. This also suggests that when the included covariates are irrelevant (so $\pi = 0$), the bias due to this inclusion is the order of $O(h_n^2)$ while the bias due to the endogeneity is the order of $O(h_n)$. Next note $\widetilde{\beta}^* - \beta = \widetilde{\beta} - \beta + O_p(h_n) = O_p(h_n)$ because $\widetilde{\beta} = \beta + O_p(h_n)$ (see HTV), so the Wald estimator is consistent whether or not x_i

is endogenous. We conclude that whether x_i is endogenous as $E[x_i u_i | z] \neq 0$ for values of z close to the cut off z_0 is irrelevant in terms of consistency of the Wald estimator although it potentially causes a larger asymptotic bias.

3.2. WALD ESTIMATOR EQUIVALENT TO LOCAL LINEAR RDD ESTIMATOR

Next we show a similar result with the local linear RDD estimator. The Wald estimator equivalent to the local linear RDD estimator is obtained by 2SLS or IV estimation of

$$y_i = \alpha + b_u c_i (z_i - z_0) + b_l (1 - c_i)(z_i - z_0) + w_i \beta + x_i' \pi + u_i \tag{6}$$

applied to the subsample $\mathscr{Z}$ where c_i is used as an excluded IV for w_i. When x_i is not included in the regression, the Wald estimator equivalent to the local linear RDD estimator becomes $\widehat{\beta} = \frac{\widehat{y}^+ - \widehat{y}^-}{\widehat{w}^+ - \widehat{w}^-}$ where $\widehat{y}^{+(-)}$ and $\widehat{w}^{+(-)}$ are the local linear conditional mean estimators. Specifically we have

$$\widehat{y}^+ = \frac{\sum_{i \in \mathscr{Z}} c_i y_i (z_i - z_0) \sum_{i \in \mathscr{Z}} c_i (z_i - z_0) - \sum_{i \in \mathscr{Z}} c_i (z_i - z_0)^2 \sum_{i \in \mathscr{Z}} c_i y_i}{\left(\sum_{i \in \mathscr{Z}} c_i (z_i - z_0)\right)^2 - \sum_{i \in \mathscr{Z}} c_i \sum_{i \in \mathscr{Z}} c_i (z_i - z_0)^2} \tag{7}$$

and others are defined similarly. Therefore the Wald estimator and the local linear RDD estimator are numerically identical when we do not include x_i in the regression. This numerical equivalence was noted by Imbens and Lemiuex (2008).

Next we derive the Wald estimator of (6) when x_i is included in the estimation (assuming x_i is scalar without loss of generality). In this case the numerical equivalence does not hold anymore.

Theorem 3. *Let* $\widehat{y}^{+(-)}$, $\widehat{w}^{+(-)}$, *and* $\widehat{x}^{+(-)}$ *be the local linear conditional mean estimators similarly defined as (7), respectively. Also let* $\widehat{y}_i = \widehat{\alpha} + \widehat{\beta} w_i$ *be the fitted value from the Wald estimation of (6) without including* x_i *where* $(\widehat{\alpha}, \widehat{\beta})$ *are the resulting Wald estimators. Then the Wald treatment estimator of (6) equivalent to the local linear estimator, denoted by* $\widehat{\beta}^*$, *can be written as*

$$\widehat{\beta}^* - \widehat{\beta} \tag{8}$$

$$= -\frac{\left(\widehat{x}^+ - \widehat{x}^-\right) \frac{\sum_{i \in \mathscr{Z}} x_i (y_i - \widehat{y}_i)}{\sum_{i \in \mathscr{Z}} x_i^2}}{\left(\widehat{w}^+ - \widehat{w}^-\right) \frac{\widehat{Var}[x_i | z_i = z_0]}{\widehat{E}[x_i^2 | z_i = z_0]} - \left(\widehat{x}^+ - \widehat{x}^-\right) \frac{\widehat{Cov}[x_i, w_i | z_i = z_0]}{\widehat{E}[x_i^2 | z_i = z_0]}}$$

$$= -\frac{\left(\widehat{x}^+ - \widehat{x}^-\right) \left\{ \pi + (\alpha - \widehat{\alpha}) \frac{\sum_{i \in \mathscr{Z}} x_i}{\sum_{i \in \mathscr{Z}} x_i^2} + (\beta - \widehat{\beta}) \frac{\sum_{i \in \mathscr{Z}} x_i w_i}{\sum_{i \in \mathscr{Z}} x_i^2} + \frac{\sum_{i \in \mathscr{Z}} x_i u_i}{\sum_{i \in \mathscr{Z}} x_i^2} \right\}}{\left(\widehat{w}^+ - \widehat{w}^-\right) \frac{\widehat{Var}[x_i | z_i = z_0]}{\widehat{E}[x_i^2 | z_i = z_0]} - \left(\widehat{x}^+ - \widehat{x}^-\right) \frac{\widehat{Cov}[x_i, w_i | z_i = z_0]}{\widehat{E}[x_i^2 | z_i = z_0]}}$$

where $\widehat{Var}[x_i|z_i = z_0]$, $\widehat{Cov}[x_i, w_i|z_i = z_0]$, *and* $\widehat{E}[x_i^2|z_i = z_0]$ *are consistent estimators of* $Var[x_i|z_i = z_0]$, $Cov[x_i, w_i|z_i = z_0]$, *and* $E[x_i^2|z_i = z_0]$, *respectively (see their definitions in Appendix B).*

See Appendix B for the proof.

From the result of Theorem 3 we find that as long as the continuity condition $\lim_{z \to z_0^+} E[x_i|z_i = z] = \lim_{z \to z_0^-} E[x_i|z_i = z]$ holds, endogeneity of x_i does not affect the consistency and the order of asymptotic bias. In the local linear estimation case, the order of asymptotic bias equal to $O\left(h_n^2\right)$ and so we have $\widehat{x}^+ - \widehat{x}^- = O_p\left(h_n^2\right)$ when the continuity condition holds. Note the term $(\widehat{x}^+ - \widehat{x}^-)\frac{\sum_{i \in \mathscr{Z}} x_i u_i}{\sum_{i \in \mathscr{Z}} x_i^2}$ in (9) reflects the possible bias due to the endogeneity of x_i, i.e., $E[x_i u_i|z_i = z_0] \neq 0$. This suggests that when x_i is endogenous, the order of the asymptotic bias $(\widetilde{x}^+ - \widetilde{x}^-)\frac{\sum_{i \in \mathscr{Z}} x_i u_i}{\sum_{i \in \mathscr{Z}} x_i^2} = O_p(h_n^2)$ and it is the same order with bias reduction. It follows that $\widehat{\beta}^* - \beta = \widehat{\beta} - \beta + O_p(h_n^2) = O_p(h_n^2)$ because $\widehat{\beta} = \beta + O_p(h_n^2)$ (see HTV). Therefore like the kernel estimator case the endogeneity of x_i is irrelevant in terms of consistency and the order of the asymptotic bias $O\left(h_n^2\right)$ but it can potentially cause an additional asymptotic bias with the same order of $O\left(h_n^2\right)$.

3.3. COEFFICIENTS ESTIMATES ON ENDOGENOUS COVARIATES

Finally we ask how the estimator of the coefficients on covariates is affected by the endogeneity. We obtain

Theorem 4. *1. Let* $\widetilde{y}^{+(-)}$, $\widetilde{w}^{+(-)}$, *and* $\widetilde{x}^{+(-)}$ *be the uniform kernel conditional mean estimators, respectively. Then for the Wald estimation of (1)-(2), we have*

$$\widehat{\pi} = \frac{(\widetilde{w}^+ - \widetilde{w}^-)\frac{\sum_{i \in \mathscr{Z}} x_i y_i}{\sum_{i \in \mathscr{Z}} x_i^2} - (\widetilde{y}^+ - \widetilde{y}^-)\frac{\sum_{i \in \mathscr{Z}} x_i w_i}{\sum_{i \in \mathscr{Z}} x_i^2} - (\widetilde{y}^- \widetilde{w}^+ - \widetilde{w}^- \widetilde{y}^+)\frac{\sum_{i \in \mathscr{Z}} x_i}{\sum_{i \in \mathscr{Z}} x_i^2}}{(\widetilde{w}^+ - \widetilde{w}^-) - (\widetilde{x}^+ - \widetilde{x}^-)\frac{\sum_{i \in \mathscr{Z}} x_i w_i}{\sum_{i \in \mathscr{Z}} x_i^2} - (\widetilde{w}^+ \widetilde{x}^- - \widetilde{w}^- \widetilde{x}^+)\frac{\sum_{i \in \mathscr{Z}} x_i}{\sum_{i \in \mathscr{Z}} x_i^2}}.$$

2. Further suppose Assumption **RD** *and Assumption* **Continuity** *hold. Then we obtain*

$$\widehat{\pi} - \pi = \frac{E[x_i u_i|z_0]}{E[x_i^2|z_0]} + O_p(h_n).$$

See Appendix A for the proof.

The results in Theorem 4 show that $\widehat{\pi}$ becomes inconsistent when x_i is endogenous as $E[x_i u_i|z_0] \neq 0$. We conjecture a similar result holds for the Wald estimator equivalent to the local linear RDD estimator.

4. CONCLUSION

Empirical researchers often include observable covariates or attributes other than the *forcing* variable in the implementation of RDD as the Wald-type estimation. This helps because it reduces some bias caused by inclusion of data points *not very close to* the cut off and it reduces the variance when the covariates is correlated with the potential outcomes. This note poses a question how the positive effects of including covariates will change if the covariates are endogenous. We show that the Wald estimator of the RD design is still consistent as long as the continuity conditions of the RD design hold. We further show that adding covariates in the Wald estimation can potentially reduce the bias of the RDD treatment effect estimator. However, when those covariates are endogenous, the order of bias due to endogeneity is the same order with the bias reduction. Moreover, the coefficients on covariates will be inconsistent.

APPENDIX

A. WALD ESTIMATOR EQUIVALENT TO THE KERNEL ESTIMATOR (PROOF OF THEOREM 2 AND THEOREM 4)

First consider the Wald estimator without x_i from the regression:

$$y_i = \alpha + \beta w_i + v_i. \tag{9}$$

The Wald estimator is obtained by IV estimation with c_i as the excluded IV for w_i, applied to the subsample $\mathscr{Z}$ such that

$$
\begin{bmatrix} \widetilde{\alpha} \\ \widetilde{\beta} \end{bmatrix} = \begin{bmatrix} \sum_{i\in\mathscr{Z}} 1 & \sum_{i\in\mathscr{Z}} w_i \\ \sum_{i\in\mathscr{Z}} c_i & \sum_{i\in\mathscr{Z}} c_i w_i \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i\in\mathscr{Z}} y_i \\ \sum_{i\in\mathscr{Z}} c_i y_i \end{bmatrix}
$$

$$
= \begin{bmatrix} \frac{\sum_{i\in\mathscr{Z}} c_i w_i \sum_{i\in\mathscr{Z}} y_i - \sum_{i\in\mathscr{Z}} c_i y_i \sum_{i\in\mathscr{Z}} w_i}{\sum_{i\in\mathscr{Z}} 1 \sum_{i\in\mathscr{Z}} c_i w_i - \sum_{i\in\mathscr{Z}} c_i \sum_{i\in\mathscr{Z}} w_i} \\ \frac{\sum_{i\in\mathscr{Z}} 1 \sum_{i\in\mathscr{Z}} c_i y_i - \sum_{i\in\mathscr{Z}} c_i \sum_{i=1}^{n} y_i}{\sum_{i\in\mathscr{Z}} 1 \sum_{i\in\mathscr{Z}} c_i w_i - \sum_{i\in\mathscr{Z}} c_i \sum_{i=1}^{n} w_i} \end{bmatrix}.
$$

Now note

$$
\begin{aligned}
\widetilde{y}^+ - \widetilde{y}^- &= \frac{\sum_{i\in\mathscr{Z}} c_i y_i}{\sum_{i\in\mathscr{Z}} c_i} - \frac{\sum_{i\in\mathscr{Z}} (1-c_i) y_i}{\sum_{i\in\mathscr{Z}} (1-c_i)} \\
&= \frac{\sum_{i\in\mathscr{Z}} (1-c_i) \sum_{i\in\mathscr{Z}} c_i y_i - \sum_{i\in\mathscr{Z}} c_i \sum_{i\in\mathscr{Z}} (1-c_i) y_i}{\sum_{i\in\mathscr{Z}} c_i \sum_{i\in\mathscr{Z}} (1-c_i)} \\
&= \frac{\sum_{i\in\mathscr{Z}} 1 \sum_{i\in\mathscr{Z}} c_i y_i - \sum_{i\in\mathscr{Z}} y_i \sum_{i\in\mathscr{Z}} c_i}{\sum_{i\in\mathscr{Z}} c_i \sum_{i\in\mathscr{Z}} (1-c_i)}.
\end{aligned} \tag{10}
$$

Similarly we obtain

$$\widetilde{w}^+ - \widetilde{w}^- = \frac{\sum_{i\in\mathscr{Z}}1\sum_{i\in\mathscr{Z}}c_iw_i - \sum_{i\in\mathscr{Z}}w_i\sum_{i\in\mathscr{Z}}c_i}{\sum_{i\in\mathscr{Z}}c_i\sum_{i\in\mathscr{Z}}(1-c_i)}.$$

By rearranging terms, we conclude

$$\begin{bmatrix} \widetilde{\alpha} \\ \widetilde{\beta} \end{bmatrix} = \begin{bmatrix} \widetilde{y}^- - \widetilde{w}^-\widetilde{\beta} \\ (\widetilde{y}^+ - \widetilde{y}^-)/(\widetilde{w}^+ - \widetilde{w}^-) \end{bmatrix}.$$

Now the Wald estimator including x_i as an additional regressor is obtained by the IV estimator of the following regression

$$y_i = \alpha + \beta w_i + \pi x_i + u_i$$

with c_i as the excluded IV for w_i for the subsample $\mathscr{Z}$. We have

$$\begin{bmatrix} \widetilde{\alpha}^* \\ \widetilde{\pi}^* \\ \widetilde{\beta}^* \end{bmatrix} = \begin{bmatrix} \sum_{i\in\mathscr{Z}}1 & \sum_{i\in\mathscr{Z}}x_i & \sum_{i\in\mathscr{Z}}w_i \\ \sum_{i\in\mathscr{Z}}x_i & \sum_{i\in\mathscr{Z}}x_i^2 & \sum_{i\in\mathscr{Z}}x_iw_i \\ \sum_{i\in\mathscr{Z}}c_i & \sum_{i\in\mathscr{Z}}c_ix_i & \sum_{i\in\mathscr{Z}}c_iw_i \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i\in\mathscr{Z}}y_i \\ \sum_{i\in\mathscr{Z}}x_iy_i \\ \sum_{i\in\mathscr{Z}}c_iy_i \end{bmatrix}.$$

We obtain

$$\widetilde{\beta}^* = \frac{\left\{ \begin{array}{c} \sum_{i\in\mathscr{Z}}y_ic_i\sum_{i\in\mathscr{Z}}x_i^2 - \sum_{i\in\mathscr{Z}}y_i\sum_{i\in\mathscr{Z}}x_i\sum_{i\in\mathscr{Z}}c_ix_i \\ + \left(n\sum_{i\in\mathscr{Z}}c_ix_i - \sum_{i\in\mathscr{Z}}c_i\sum_{i\in\mathscr{Z}}x_i \right)\sum_{i\in\mathscr{Z}}x_iy_i \\ + \left(\left(\sum_{i\in\mathscr{Z}}x_i\right)^2 - n\sum_{i\in\mathscr{Z}}x_i^2 \right)\sum_{i\in\mathscr{Z}}c_iy_i \end{array} \right\}}{\left\{ \begin{array}{c} \sum_{i\in\mathscr{Z}}w_i\sum_{i\in\mathscr{Z}}c_i\sum_{i\in\mathscr{Z}}x_i^2 - \sum_{i\in\mathscr{Z}}w_i\sum_{i\in\mathscr{Z}}x_i\sum_{i\in\mathscr{Z}}c_ix_i \\ + \left(n\sum_{i\in\mathscr{Z}}c_ix_i - \sum_{i\in\mathscr{Z}}c_i\sum_{i\in\mathscr{Z}}x_i \right)\sum_{i\in\mathscr{Z}}x_iw_i \\ + \left(\left(\sum_{i\in\mathscr{Z}}x_i\right)^2 - n\sum_{i\in\mathscr{Z}}x_i^2 \right)\sum_{i\in\mathscr{Z}}c_iw_i \end{array} \right\}}. \tag{11}$$

We first consider the numerator. By rearranging terms, we obtain the numerator equals to

$$\sum_{i\in\mathscr{Z}}x_i^2\sum_{i\in\mathscr{Z}}c_i\sum_{i\in\mathscr{Z}}(1-c_i)\left\{ \frac{\sum_{i\in\mathscr{Z}}1\sum_{i\in\mathscr{Z}}c_iy_i - \sum_{i\in\mathscr{Z}}y_i\sum_{i\in\mathscr{Z}}c_i}{\sum_{i\in\mathscr{Z}}c_i\sum_{i\in\mathscr{Z}}(1-c_i)} \right\} \tag{12}$$

$$- \sum_{i\in\mathscr{Z}}x_i^2\sum_{i\in\mathscr{Z}}c_i\sum_{i\in\mathscr{Z}}(1-c_i)\left\{ \frac{\sum_{i\in\mathscr{Z}}x_iy_i}{\sum_{i\in\mathscr{Z}}x_i^2}\frac{(\sum_{i\in\mathscr{Z}}1\sum_{i\in\mathscr{Z}}c_ix_i - \sum_{i\in\mathscr{Z}}c_i\sum_{i\in\mathscr{Z}}x_i)}{\sum_{i\in\mathscr{Z}}c_i\sum_{i\in\mathscr{Z}}(1-c_i)} \right\}$$

$$- \sum_{i\in\mathscr{Z}}x_i^2\sum_{i\in\mathscr{Z}}c_i\sum_{i\in\mathscr{Z}}(1-c_i)\left\{ \frac{\sum_{i\in\mathscr{Z}}x_i}{\sum_{i\in\mathscr{Z}}x_i^2}\frac{(\sum_{i\in\mathscr{Z}}c_iy_i\sum_{i\in\mathscr{Z}}x_i - \sum_{i\in\mathscr{Z}}c_ix_i\sum_{i\in\mathscr{Z}}y_i)}{\sum_{i\in\mathscr{Z}}c_i\sum_{i\in\mathscr{Z}}(1-c_i)} \right\}.$$

Similarly with (10) we note

$$\widetilde{x}^+ - \widetilde{x}^- = \frac{\sum_{i \in \mathscr{L}} 1 \sum_{i \in \mathscr{L}} c_i x_i - \sum_{i \in \mathscr{L}} x_i \sum_{i \in \mathscr{L}} c_i}{\sum_{i \in \mathscr{L}} c_i \sum_{i \in \mathscr{L}} (1 - c_i)}.$$

For the third term in (12), we note

$$\frac{\left(\sum_{i \in \mathscr{L}} c_i y_i \sum_{i \in \mathscr{L}} x_i - \sum_{i \in \mathscr{L}} c_i x_i \sum_{i \in \mathscr{L}} y_i \right)}{\sum_{i \in \mathscr{L}} c_i \sum_{i \in \mathscr{L}} (1 - c_i)}$$

$$= \frac{\sum_{i \in \mathscr{L}} c_i y_i \sum_{i \in \mathscr{L}} (1 - c_i) x_i - \sum_{i \in \mathscr{L}} (1 - c_i) y_i \sum_{i \in \mathscr{L}} c_i x_i}{\sum_{i \in \mathscr{L}} c_i \sum_{i \in \mathscr{L}} (1 - c_i)}$$

$$= \widetilde{y}^+ \widetilde{x}^- - \widetilde{y}^- \widetilde{x}^+ = \left(\widetilde{y}^+ - \widetilde{y}^- \right) \widetilde{x}^- - \left(\widetilde{x}^+ - \widetilde{x}^- \right) \widetilde{y}^-$$

Combining the above results, we conclude the numerator in (11) equals to

$$\sum_{i \in \mathscr{L}} x_i^2 \sum_{i \in \mathscr{L}} c_i \sum_{i \in \mathscr{L}} (1 - c_i) \left\{ \widetilde{y}^+ - \widetilde{y}^- - \frac{\sum_{i \in \mathscr{L}} x_i y_i}{\sum_{i \in \mathscr{L}} x_i^2} \left(\widetilde{x}^+ - \widetilde{x}^- \right) - \frac{\sum_{i \in \mathscr{L}} x_i}{\sum_{i \in \mathscr{L}} x_i^2} \left(\widetilde{y}^+ \widetilde{x}^- - \widetilde{x}^+ \widetilde{y}^- \right) \right\}.$$

Similarly we can show that the denominator in (11) equals to

$$\sum_{i \in \mathscr{L}} x_i^2 \sum_{i \in \mathscr{L}} c_i \sum_{i \in \mathscr{L}} (1 - c_i) \left\{ \widetilde{w}^+ - \widetilde{w}^- - \frac{\sum_{i \in \mathscr{L}} x_i w_i}{\sum_{i \in \mathscr{L}} x_i^2} \left(\widetilde{x}^+ - \widetilde{x}^- \right) - \frac{\sum_{i \in \mathscr{L}} x_i}{\sum_{i \in \mathscr{L}} x_i^2} \left(\widetilde{w}^+ \widetilde{x}^- - \widetilde{x}^+ \widetilde{w}^- \right) \right\}.$$

Therefore, we conclude

$$\widetilde{\beta}^* = \frac{\widetilde{y}^+ - \widetilde{y}^- - \frac{\sum_{i \in \mathscr{L}} x_i y_i}{\sum_{i \in \mathscr{L}} x_i^2} \left(\widetilde{x}^+ - \widetilde{x}^- \right) - \frac{\sum_{i \in \mathscr{L}} x_i}{\sum_{i \in \mathscr{L}} x_i^2} \left((\widetilde{y}^+ - \widetilde{y}^-) \widetilde{x}^- - (\widetilde{x}^+ - \widetilde{x}^-) \widetilde{y}^- \right)}{\widetilde{w}^+ - \widetilde{w}^- - \frac{\sum_{i \in \mathscr{L}} x_i w_i}{\sum_{i \in \mathscr{L}} x_i^2} \left(\widetilde{x}^+ - \widetilde{x}^- \right) - \frac{\sum_{i \in \mathscr{L}} x_i}{\sum_{i \in \mathscr{L}} x_i^2} \left((\widetilde{w}^+ - \widetilde{w}^-) \widetilde{x}^- - (\widetilde{x}^+ - \widetilde{x}^-) \widetilde{w}^- \right)}.$$

Now define

$$\widehat{\varphi}_y = \frac{\sum_{i \in \mathscr{L}} x_i y_i}{\sum_{i= \mathscr{L}} x_i^2} \left(\widetilde{x}^+ - \widetilde{x}^- \right) + \frac{\sum_{i \in \mathscr{L}} x_i}{\sum_{i \in \mathscr{L}} x_i^2} \left((\widetilde{y}^+ - \widetilde{y}^-) \widetilde{x}^- - \widetilde{y}^- (\widetilde{x}^+ - \widetilde{x}^-) \right) \text{ and}$$

$$\widehat{\varphi}_w = \frac{\sum_{i \in \mathscr{L}} x_i w_i}{\sum_{i \in \mathscr{L}} x_i^2} \left(\widetilde{x}^+ - \widetilde{x}^- \right) + \frac{\sum_{i \in \mathscr{L}} x_i}{\sum_{i \in \mathscr{L}} x_i^2} \left((\widetilde{w}^+ - \widetilde{w}^-) \widetilde{x}^- - \widetilde{w}^- (\widetilde{x}^+ - \widetilde{x}^-) \right)$$

and obtain

$$\widetilde{\beta}^* = \frac{\widetilde{y}^+ - \widetilde{y}^-}{\widetilde{w}^+ - \widetilde{w}^-} - \frac{(\widetilde{w}^+ - \widetilde{w}^-) \widehat{\varphi}_y - \widehat{\varphi}_w (\widetilde{y}^+ - \widetilde{y}^-)}{(\widetilde{w}^+ - \widetilde{w}^- - \widehat{\varphi}_w)(\widetilde{w}^+ - \widetilde{w}^-)}$$

$$= \widetilde{\beta} - \frac{(\widetilde{w}^+ - \widetilde{w}^-) \widehat{\varphi}_y - \widehat{\varphi}_w (\widetilde{y}^+ - \widetilde{y}^-)}{(\widetilde{w}^+ - \widetilde{w}^- - \widehat{\varphi}_w)(\widetilde{w}^+ - \widetilde{w}^-)} = \widetilde{\beta} - \frac{\widehat{\varphi}_y - \widehat{\varphi}_w \widetilde{\beta}}{(\widetilde{w}^+ - \widetilde{w}^- - \widehat{\varphi}_w)}.$$

Further note that

$$\widehat{\varphi}_y - \widehat{\varphi}_w \widetilde{\beta}$$

$$= \left(\widetilde{x}^+ - \widetilde{x}^- \right) \left\{ \frac{\sum_{i \in \mathscr{L}} x_i y_i}{\sum_{i = \mathscr{L}} x_i^2} - \frac{\sum_{i \in \mathscr{L}} x_i}{\sum_{i \in \mathscr{L}} x_i^2} \widetilde{y}^- - \frac{\sum_{i \in \mathscr{L}} x_i w_i}{\sum_{i \in \mathscr{L}} x_i^2} \widetilde{\beta} + \frac{\sum_{i \in \mathscr{L}} x_i}{\sum_{i \in \mathscr{L}} x_i^2} \widetilde{w}^- \widetilde{\beta} \right\}$$

$$+ \widetilde{x}^- \frac{\sum_{i \in \mathscr{L}} x_i}{\sum_{i \in \mathscr{L}} x_i^2} \left\{ \left(\widetilde{y}^+ - \widetilde{y}^- \right) - \left(\widetilde{w}^+ - \widetilde{w}^- \right) \widetilde{\beta} \right\}$$

$$= \left(\widetilde{x}^+ - \widetilde{x}^- \right) \left\{ \frac{\sum_{i \in \mathscr{L}} x_i y_i}{\sum_{i = \mathscr{L}} x_i^2} - \frac{\sum_{i \in \mathscr{L}} x_i}{\sum_{i \in \mathscr{L}} x_i^2} \widetilde{y}^- - \left(\frac{\sum_{i \in \mathscr{L}} x_i w_i}{\sum_{i \in \mathscr{L}} x_i^2} \widetilde{\beta} - \frac{\sum_{i \in \mathscr{L}} x_i \widetilde{w}^-}{\sum_{i \in \mathscr{L}} x_i^2} \widetilde{\beta} \right) \right\}.$$

To interpret the term $\left\{ \frac{\sum_{i \in \mathscr{L}} x_i y_i}{\sum_{i = \mathscr{L}} x_i^2} - \frac{\sum_{i \in \mathscr{L}} x_i}{\sum_{i \in \mathscr{L}} x_i^2} \widetilde{y}^- - \left(\frac{\sum_{i \in \mathscr{L}} x_i w_i}{\sum_{i \in \mathscr{L}} x_i^2} \widetilde{\beta} - \frac{\sum_{i \in \mathscr{L}} x_i \widetilde{w}^-}{\sum_{i \in \mathscr{L}} x_i^2} \widetilde{\beta} \right) \right\}$, define the fitted value of y_i from the IV regression of (9) as $\widetilde{y}_i \equiv \widetilde{\alpha} + \widetilde{\beta} w_i$ and note

$$\widetilde{y}_i \equiv \widetilde{\alpha} + \widetilde{\beta} w_i = \widetilde{y}^- + \widetilde{\beta} \left(w_i - \widetilde{w}_i^- \right).$$

After plugging in $\widetilde{y}^- = \widetilde{y}_i - \widetilde{\beta} \left(w_i - \widetilde{w}_i^- \right)$ and rearranging terms we obtain

$$\widehat{\varphi}_y - \widehat{\varphi}_w \widetilde{\beta}$$

$$= \left(\widetilde{x}^+ - \widetilde{x}^- \right) \left\{ \frac{\sum_{i \in \mathscr{L}} x_i y_i}{\sum_{i = \mathscr{L}} x_i^2} - \frac{\sum_{i \in \mathscr{L}} x_i}{\sum_{i \in \mathscr{L}} x_i^2} \widetilde{y}^- - \left(\frac{\sum_{i \in \mathscr{L}} x_i w_i}{\sum_{i \in \mathscr{L}} x_i^2} \widetilde{\beta} - \frac{\sum_{i \in \mathscr{L}} x_i \widetilde{w}^-}{\sum_{i \in \mathscr{L}} x_i^2} \widetilde{\beta} \right) \right\}$$

$$= \left(\widetilde{x}^+ - \widetilde{x}^- \right) \frac{\sum_{i \in \mathscr{L}} x_i \left(y_i - \widetilde{y}_i \right)}{\sum_{i = \mathscr{L}} x_i^2}.$$

Combining results above we obtain

$$\widetilde{\beta}^* = \widetilde{\beta} - \left(\widetilde{x}^+ - \widetilde{x}^- \right) \left\{ \frac{\sum_{i \in \mathscr{L}} x_i \left(y_i - \widetilde{y}_i \right)}{\sum_{i = \mathscr{L}} x_i^2} \right\} / \left(\widetilde{w}^+ - \widetilde{w}^- - \widehat{\varphi}_w \right).$$

Without loss of generality, we can let $E[x_i | z_i = z_0] = 0$. Then we obtain

$$\widehat{\varphi}_w = \frac{\sum_{i \in \mathscr{L}} x_i w_i / n}{\sum_{i \in \mathscr{L}} x_i^2 / n} \left(\widetilde{x}^+ - \widetilde{x}^- \right)$$

$$+ \frac{\sum_{i \in \mathscr{L}} x_i / n}{\sum_{i \in \mathscr{L}} x_i^2 / n} \left(\left(\widetilde{w}^+ - \widetilde{w}^- \right) \widetilde{x}^- - \widetilde{w}^- \left(\widetilde{x}^+ - \widetilde{x}^- \right) \right) = O_p(h_n)$$

because each sample means converge to their corresponding population means and because $\widetilde{x}^+ - \widetilde{x}^- = O_p(h_n)$ due to the continuity condition and the order of bias of kernel estimators equal to $O(h_n)$. We therefore conclude $\widetilde{\beta}^* = \widetilde{\beta} -$

$(\widetilde{x}^+ - \widetilde{x}^-) \left\{ \frac{\sum_{i \in \mathscr{L}} x_i (y_i - \widetilde{y}_i)}{\sum_{i = \mathscr{L}} x_i^2} \right\} / (\widetilde{w}^+ - \widetilde{w}^- + O_p(h_n))$. This proves (4). Then (5) follows by plugging in $y_i = \alpha + w_i \beta + x_i' \pi + u_i$ and $\widetilde{y}_i = \widetilde{\alpha} + \widetilde{\beta} w_i$ and after rearranging terms.

Now we turn to the coefficient estimator of the covariate. First, define

$$\widetilde{\Psi}_{w,x} = \widetilde{w}^+ - \widetilde{w}^- - \frac{\sum_{i \in \mathscr{L}} x_i w_i}{\sum_{i \in \mathscr{L}} x_i^2} \left(\widetilde{x}^+ - \widetilde{x}^- \right) - \frac{\sum_{i \in \mathscr{L}} x_i}{\sum_{i \in \mathscr{L}} x_i^2} \left(\widetilde{w}^+ \widetilde{x}^- - \widetilde{x}^+ \widetilde{w}^- \right)$$

and observe that

$$\widetilde{\alpha}^* \;=\; \frac{\left(\widetilde{w}^+ \widetilde{y}^- - \widetilde{y}^+ \widetilde{w}^- \right) - \frac{\sum_{i \in \mathscr{L}} x_i y_i}{\sum_{i \in \mathscr{L}} x_i^2} \left(\widetilde{w}^+ \widetilde{x}^- - \widetilde{w}^- \widetilde{x}^+ \right) - \frac{\sum_{i \in \mathscr{L}} w_i x_i}{\sum_{i \in \mathscr{L}} x_i^2} \left(\widetilde{x}^+ \widetilde{y}^- - \widetilde{x}^- \widetilde{y}^+ \right)}{\widetilde{\Psi}_{w,x}}$$

$$\widetilde{\beta}^* \;=\; \frac{\left(\widetilde{y}^+ - \widetilde{y}^- \right) - \frac{\sum_{i \in \mathscr{L}} x_i y_i}{\sum_{i \in \mathscr{L}} x_i^2} \left(\widehat{x}^+ - \widehat{x}^- \right) - \frac{\sum_{i \in \mathscr{L}} x_i}{\sum_{i \in \mathscr{L}} x_i^2} \left(\widehat{y}^+ \widehat{x}^- - \widehat{x}^+ \widehat{y}^- \right)}{\widetilde{\Psi}_{w,x}}$$

$$\widetilde{\pi}^* \;=\; \frac{\left(\widetilde{w}^+ - \widetilde{w}^- \right) \frac{\sum_{i \in \mathscr{L}} x_i y_i}{\sum_{i \in \mathscr{L}} x_i^2} - \left(\widetilde{y}^+ - \widetilde{y}^- \right) \frac{\sum_{i \in \mathscr{L}} x_i w_i}{\sum_{i \in \mathscr{L}} x_i^2} - \frac{\sum_{i \in \mathscr{L}} x_i}{\sum_{i \in \mathscr{L}} x_i^2} \left(\widetilde{y}^- \widetilde{w}^+ - \widetilde{w}^- \widetilde{y}^+ \right)}{\widetilde{\Psi}_{w,x}}.$$

We obtain

$$\begin{aligned}
\widetilde{\alpha}^* &= \widetilde{\alpha} + O_p(h_n) \\
\widetilde{\beta}^* &= \widetilde{\beta} + O_p(h_n) \\
\widetilde{\pi}^* &= \frac{\sum_{i \in \mathscr{L}} x_i y_i}{\sum_{i \in \mathscr{L}} x_i^2} - \widetilde{\beta} \frac{\sum_{i \in \mathscr{L}} x_i w_i}{\sum_{i \in \mathscr{L}} x_i^2} + O_p(h_n).
\end{aligned}$$

It follows that

$$\begin{aligned}
\widetilde{\pi}^* - \pi &= \frac{\sum_{i \in \mathscr{L}} x_i}{\sum_{i \in \mathscr{L}} x_i^2} \alpha + \left(\beta - \widetilde{\beta} \right) \frac{\sum_{i \in \mathscr{L}} x_i w_i}{\sum_{i \in \mathscr{L}} x_i^2} + \frac{\sum_{i \in \mathscr{L}} x_i u_i}{\sum_{i \in \mathscr{L}} x_i^2} + O_p(h_n) \\
&= \left(\beta - \widetilde{\beta} \right) \frac{\sum_{i \in \mathscr{L}} x_i w_i}{\sum_{i \in \mathscr{L}} x_i^2} + \frac{\sum_{i \in \mathscr{L}} x_i u_i}{\sum_{i \in \mathscr{L}} x_i^2} + O_p(h_n).
\end{aligned}$$

Therefore we conclude

$$plim_{n \to \infty} \left(\widetilde{\pi}^* - \pi \right) = plim_{n \to \infty} \frac{\sum_{i \in \mathscr{L}} x_i u_i}{\sum_{i \in \mathscr{L}} x_i^2}$$

and so $\widetilde{\pi}^*$ is inconsistent when x_i is endogenous, i.e. $E\left[x_i u_i | z_i = z_0 \right] \neq 0$.

B. WALD ESTIMATOR EQUIVALENT TO THE LOCAL LINEAR ESTIMATOR (PROOF OF THEOREM 3)

Without loss of generality, we let $z_0 = 0$. First we note the local linear estimators of the conditional means of $y^{+(-)}$ are given by

$$
\widehat{y}^+ = \frac{\sum_{i \in \mathscr{Z}} c_i y_i z_i \sum_{i \in \mathscr{Z}} c_i z_i - \sum_{i \in \mathscr{Z}} c_i z_i^2 \sum_{i \in \mathscr{Z}} c_i y_i}{\left(\sum_{i \in \mathscr{Z}} c_i z_i\right)^2 - \left(\sum_{i \in \mathscr{Z}} c_i\right) \sum_{i \in \mathscr{Z}} c_i z_i^2}
$$

$$
\widehat{y}^- = \frac{\sum_{i \in \mathscr{Z}} (1 - c_i) y_i z_i \sum_{i \in \mathscr{Z}} (z_i - c_i z_i) - \sum_{i \in \mathscr{Z}} (1 - c_i) y_i \sum_{i \in \mathscr{Z}} \left(z_i^2 - c_i z_i^2\right)}{\left(\sum_{i \in \mathscr{Z}} (z_i - c_i z_i)\right)^2 - \left(\sum_{i \in \mathscr{Z}} \left(z_i^2 - c_i z_i^2\right)\right) \sum_{i \in \mathscr{Z}} (1 - c_i)}
$$

and the local linear estimators of other conditional means are obtained similarly. The Wald estimator equivalent to the local linear estimator is obtained by IV or 2SLS estimation of

$$
y_i = \alpha + b_u c_i z_i + b_l (1 - c_i) z_i + \beta w_i + \pi x_i + u_i
$$

with c_i being the excluded IV for w_i. We let

$$
y_i = Z_i' \theta + \pi x_i + u_i
$$

where $Z_i' = (1 \; c_i z_i \; (1 - c_i) z_i \; w_i)$, $\widetilde{Z}_i = (1 \; c_i z_i \; (1 - c_i) z_i \; c_i)'$, and $\theta = (\alpha \; b_u \; b_l \; \beta)$. We obtain

$$
\begin{bmatrix} \widehat{\theta}^* \\ \widehat{\pi}^* \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}
$$

where $\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} \sum_{i \in \mathscr{Z}} \widetilde{Z}_i Z_i' & \sum_{i \in \mathscr{Z}} \widetilde{Z}_i x_i \\ \sum_{i \in \mathscr{Z}} x_i Z_i' & \sum_{i \in \mathscr{Z}} x_i^2 \end{bmatrix}$ and $\begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} \sum_{i \in \mathscr{Z}} \widetilde{Z}_i y_i \\ \sum_{i \in \mathscr{Z}} x_i y_i \end{bmatrix}$.

Using the inverse of the partitioned matrix, we obtain

$$
\begin{bmatrix} \widehat{\theta}^* \\ \widehat{\pi}^* \end{bmatrix} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1} A_{12} C_{22} A_{21} A_{11}^{-1} & -A_{11}^{-1} A_{12} C_{22} \\ -C_{22} A_{21} A_{11}^{-1} & (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}
$$

with $C_{22} = (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1}$ and that

$$
\widehat{\theta}^* = A_{11}^{-1} q_1 + A_{11}^{-1} A_{12} C_{22} \left(A_{21} A_{11}^{-1} q_1 - q_2\right).
$$

Denote the estimator of θ by $\widehat{\theta}$ when we do not include x_i in the estimation and observe that $\widehat{\theta} = A_{11}^{-1} q_1$ by construction. It follows that

$$
\widehat{\theta}^* = \widehat{\theta} + A_{11}^{-1} A_{12} C_{22} \left(A_{21} \widehat{\theta} - q_2\right).
$$

Now we consider the term $A_{11}^{-1}A_{12}$. Note that

$$A_{11}^{-1}A_{12} = \left(\sum_{i\in\mathscr{L}}\widetilde{Z}_iZ_i'\right)^{-1}\sum_{i\in\mathscr{L}}\widetilde{Z}_ix_i$$

which is the IV regression coefficient of x_i on Z_i with the instrumental variables $\widetilde{Z}_i$. Define $Z_{1i}' = (1 \ c_iz_i \ (1-c_i)z_i)$ and observe

$$A_{11}^{-1}A_{12} = \left[\begin{array}{cc} \sum_{i\in\mathscr{L}}Z_{1i}Z_{1i}' & \sum_{i\in\mathscr{L}}Z_{1i}w_i \\ \sum_{i\in\mathscr{L}}c_iZ_{1i}' & \sum_{i\in\mathscr{L}}c_iw_i \end{array}\right]^{-1}\left[\begin{array}{c} \sum_{i\in\mathscr{L}}Z_{1i}x_i \\ \sum_{i\in\mathscr{L}}c_ix_i \end{array}\right].$$

We let $\left[\begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array}\right] = \left[\begin{array}{cc} \sum_{i\in\mathscr{L}}Z_{1i}Z_{1i}' & \sum_{i\in\mathscr{L}}Z_{1i}w_i \\ \sum_{i\in\mathscr{L}}c_iZ_{1i}' & \sum_{i\in\mathscr{L}}c_iw_i \end{array}\right]$, $\left[\begin{array}{c} d_1 \\ d_2 \end{array}\right] = \left[\begin{array}{c} \sum_{i\in\mathscr{L}}Z_{1i}x_i \\ \sum_{i\in\mathscr{L}}c_ix_i \end{array}\right]$,
and using again the inverse of the partitioned matrix, we obtain that

$$A_{11}^{-1}A_{12} = \left[\begin{array}{c} a_{11}^{-1}d_1 - a_{11}^{-1}a_{12}\left(\frac{\widehat{x}^+ - \widehat{x}^-}{\widehat{w}^+ - \widehat{w}^-}\right) \\ \frac{\widehat{x}^+ - \widehat{x}^-}{\widehat{w}^+ - \widehat{w}^-} \end{array}\right]$$

where $\widehat{x}^{+(-)}$ and $\widehat{w}^{+(-)}$ are the local linear estimators of $x^{+(-)}$ and $w^{+(-)}$, respectively. Now note

$$A_{21}A_{11}^{-1}q_1 - q_2 = A_{21}\widehat{\theta} - q_2$$
$$= \widehat{\alpha}\sum_{i\in\mathscr{L}}x_i + \widehat{b}_u\sum_{i\in\mathscr{L}}c_ix_iz_i + \widehat{b}_l\sum_{i\in\mathscr{L}}(1-c_i)x_iz_i + \widehat{\beta}\sum_{i\in\mathscr{L}}x_iw_i - \sum_{i=1}^{n}x_iy_i$$
$$= -\sum_{i\in\mathscr{L}}x_i(y_i - \widehat{y}_i)$$

where $\widehat{y}_i$ is the fitted value of y_i in the Wald estimation equivalent to the local linear estimation without including x_i.

Next consider

$$C_{22}^{-1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$$
$$= \sum_{i\in\mathscr{L}}x_i^2 - \sum_{i\in\mathscr{L}}x_iZ_{1i}'\left(a_{11}^{-1}d_1 - a_{11}^{-1}a_{12}\left(\frac{\widehat{x}^+ - \widehat{x}^-}{\widehat{w}^+ - \widehat{w}^-}\right)\right) - \left(\frac{\widehat{x}^+ - \widehat{x}^-}{\widehat{w}^+ - \widehat{w}^-}\right)\sum_{i\in\mathscr{L}}x_iw_i$$
$$= \sum_{i\in\mathscr{L}}x_i^2 - \sum_{i\in\mathscr{L}}x_iZ_{1i}'a_{11}^{-1}d_1 - \left(\frac{\widehat{x}^+ - \widehat{x}^-}{\widehat{w}^+ - \widehat{w}^-}\right)\left(\sum_{i\in\mathscr{L}}x_iw_i - \sum_{i\in\mathscr{L}}x_iZ_{1i}'a_{11}^{-1}a_{12}\right).$$

Now observe that

$$\frac{\sum_{i\in\mathscr{X}} x_i^2 - \sum_{i\in\mathscr{X}} x_i Z'_{1i} a_{11}^{-1} d_1}{\sum_{i\in\mathscr{X}} 1}$$

$$= \frac{\sum_{i\in\mathscr{X}} x_i^2}{\sum_{i\in\mathscr{X}} 1} - \frac{\sum_{i\in\mathscr{X}} x_i Z'_{1i}}{\sum_{i\in\mathscr{X}} 1} \left(\frac{\sum_{i\in\mathscr{X}} Z_{1i} Z'_{1i}}{\sum_{i\in\mathscr{X}} 1} \right)^{-1} \frac{\sum_{i\in\mathscr{X}} Z_{1i} x_i}{\sum_{i\in\mathscr{X}} 1}$$

is a consistent estimator of $Var[x_i|z_i = z_0]$ and also observe that

$$\frac{\sum_{i\in\mathscr{X}} x_i w_i - \sum_{i\in\mathscr{X}} x_i Z'_{1i} a_{11}^{-1} a_{12}}{\sum_{i\in\mathscr{X}} 1}$$

$$= \frac{\sum_{i\in\mathscr{X}} x_i w_i}{\sum_{i\in\mathscr{X}} 1} - \frac{\sum_{i\in\mathscr{X}} x_i Z'_{1i}}{\sum_{i\in\mathscr{X}} 1} \left(\frac{\sum_{i\in\mathscr{X}} Z_{1i} Z'_{1i}}{\sum_{i\in\mathscr{X}} 1} \right)^{-1} \frac{\sum_{i\in\mathscr{X}} Z_{1i} w_i}{\sum_{i\in\mathscr{X}} 1}$$

$$= \frac{\sum_{i\in\mathscr{X}} x_i w_i}{\sum_{i\in\mathscr{X}} 1} - \frac{\sum_{i\in\mathscr{X}} x_i Z'_{1i}}{\sum_{i\in\mathscr{X}} 1} \left(\frac{\sum_{i\in\mathscr{X}} Z_{1i} Z'_{1i}}{\sum_{i\in\mathscr{X}} 1} \right)^{-1} \left(\frac{\sum_{i\in\mathscr{X}} Z_{1i} Z'_{1i}}{\sum_{i\in\mathscr{X}} 1} \right) \left(\frac{\sum_{i\in\mathscr{X}} Z_{1i} Z'_{1i}}{\sum_{i\in\mathscr{X}} 1} \right)^{-1} \frac{\sum_{i\in\mathscr{X}} Z_{1i} w_i}{\sum_{i\in\mathscr{X}} 1}$$

is a consistent estimator of $Cov[x_i, w_i|z_i = z_0]$.

We therefore conclude

$$\widehat{\beta}^* = \widehat{\beta} - \frac{\left(\frac{\widehat{x}^+ - \widehat{x}^-}{\widehat{w}^+ - \widehat{w}^-} \right) \sum_{i\in\mathscr{X}} x_i(y_i - \widehat{y}_i)}{\sum_{i\in\mathscr{X}} x_i^2 - \sum_{i\in\mathscr{X}} x_i Z'_{1i} a_{11}^{-1} d_1 - \left(\frac{\widehat{x}^+ - \widehat{x}^-}{\widehat{w}^+ - \widehat{w}^-} \right) \left(\sum_{i\in\mathscr{X}} x_i w_i - \sum_{i\in\mathscr{X}} x_i Z'_{1i} a_{11}^{-1} a_{12} \right)}$$

$$= \widehat{\beta} - \frac{\left(\widehat{x}^+ - \widehat{x}^- \right) \frac{\sum_{i\in\mathscr{X}} x_i(y_i - \widehat{y}_i)}{\sum_{i\in\mathscr{X}} x_i^2}}{\left(\widehat{w}^+ - \widehat{w}^- \right) \frac{\widehat{Var}[x_i|z_i=z_0]}{\widehat{E}[x_i^2|z_i=z_0]} - \left(\widehat{x}^+ - \widehat{x}^- \right) \frac{\widehat{Cov}[x_i,w_i|z_i=z_0]}{\widehat{E}[x_i^2|z_i=z_0]}}$$

where $\widehat{E}[x_i^2|z_i = z_0] = \sum_{i\in\mathscr{X}} x_i^2 / \sum_{i\in\mathscr{X}} 1$. This proves (9).

REFERENCES

Black, S (1999). Do better schools matter? Parental valuation of elementary dducation, The Quarterly Journal of Economics 114, 577–599.

Chay, K., and M. Greenstone (2005). Does air quality matter? Evidence from the housing market, Journal of Political Economy 113, 376–424.

Frölich, M. (2007). Regression discontinuity design with covariates, IZA discussion paper.

Greenstone, M., and J. Gallagher (2008). Does hazardous waste matter? Evidence from the housing market and the superfund program, The Quarterly Journal of Economics 123, 951–1003.

Hahn, J., P. Todd, and W. Van der Klaauw (2001). Identification and estimation of treatment effects with a regression-discontinuity design, Econometrica 69, 201–209.

Imbens, G., and T. Lemieux (2008). Regression discontinuity designs: A guide to practice, Journal of Econometrics 142, 615–635.

Lee, D., and T. Lemieux (2010). Regression discontinuity designs in economics, Journal of Economic Literature 48, 281–355.

Thistlethwaite, D., and D. Campbell (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment, Journal of Educational Psychology 51, 309–317.