Regression Discontinuity Design with Unknown Cutoff: Cutoff Detection & Effect Estimation

by

Tanvir Ahmed Khan Tanu

M.S.S., Economics, East West University, 2018

B.B.A., IBA, University of Dhaka, 2014

A Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF ARTS

In the Department of Economics

University of Victoria

We acknowledge with respect the Lekwungen peoples on whose traditional territory the university stands and the Songhees, Esquimalt, and WSÁNEĆ peoples whose historical relationships with the land continue to this day.

Regression Discontinuity Design with Unknown Cutoff: Cutoff Detection & Effect Estimation

by

Tanvir Ahmed Khan Tanu

M.S.S., Economics, East West University, 2018

B.B.A., IBA, University of Dhaka, 2014

Supervisory Committee

Dr. Felix Pretis, Supervisor

Department of Economics

Dr. Judith A. Clarke, Departmental Member

Department of Economics

Abstract

Regression discontinuity designs are increasingly popular quasi-experimental research designs among applied econometricians desiring to make causal inferences on the local effect of a treatment, intervention, or policy. They are also widely used in social, behavioral, and natural sciences. Much of the existing literature relies on the assumption that the discontinuity point or cutoff is known *a-priori*, which may not always hold. This thesis seeks to extend the applicability of regression discontinuity designs by proposing a new approach towards detection of an unknown discontinuity point using structural-break detection and machine learning methods. The approach is evaluated on both simulated and real data. Estimation and inference based on estimating the cutoff following this approach are compared to the counterfactual scenario where the cutoff is known. Monte Carlo simulations show that the empirical false-detection and true-detection probabilities of the proposed procedure are generally satisfactory. Finally, the approach is further illustrated with an empirical application.

Table of Contents

List of Tables

List of Figures

1 Introduction

This thesis seeks to extend the applicability of regression discontinuity designs by proposing a new approach based on structural-break detection and machine learning towards detection of unknown cutoffs, and estimation of local average treatment effects with unknown cutoffs. Regression discontinuity design (RDD), introduced by Thistlethwaite and Campbell (1960), has become a widely used framework among applied econometricians for measuring treatment effects. There is a vast literature on RDD. Reviews, theoretical developments, and applications of the method can be found in Hahn et al. (2001), Imbens and Lemieux (2008), van der Klaauw (2008), and Lee and Lemieux (2010), while Volume 38 of Advances in Econometrics, edited by Cattaneo and Escanciano (2017), as well as the references cited therein, provides an overview of recent advancements. Lee and Lemieux (2010) give a comprehensive list of applications in the economic literature until 2010, and Hausman and Rapson (2018) provide a list of applications of RDD in a time series setting, which is referred as Regression Discontinuity in Time (RDiT).

While randomized control trials (RCTs) are considered the gold standard for causal inference, such experiments are often not feasible due to time constraints, resource constraints, or ethical boundaries. For example, applying RCT to estimate the causal effect of smaller class sizes in schools on student outcomes can run into ethical concerns. Parent's preference in general, is for a smaller class, see Angrist and Lavy (1999). Parents of children who are assigned to a larger class to serve the purpose of a control group of an RCT are likely to object to the fairness of such assignment, even if the assignment is randomized. Econometricians, therefore, often exploit "natural experiments" – circumstances that lead to quasi-randomization. RDD is a quasi-experimental design that estimates the local causal effect of a treatment/intervention in situations where the treatment assignment is completely determined based on the value of an observable covariate (called the score, assignment, forcing or running variable). If the value of the score/assignment variable for an individual/test-subject is above a threshold or cutoff, the individual/test-subject gets treated, otherwise, treatment is withheld. The identification strategy in RDD is that, under the assumptions that the characteristics of the individuals do not change abruptly at the cutoff, and the individuals, even while having some influence, cannot precisely manipulate the score/assignment variable, the variations in treatment status near the cutoff are as if from a randomized experiment. The average variations in the outcome variable in opposite sides near the cutoff, therefore, can be argued as the causal effect of the treatment. Lee and Lemieux (2010) argue that RDDs require fewer assumptions than most causal inference techniques and are arguably most similar to true-randomized experiments.

Usually, the cutoff is set by the policymaker and is publicly known, however often the cutoff is masked from the public (including econometricians) due to privacy concerns, or concerns about the individuals under observation trying to manipulate their treatment status. For example, in the classical application of RDD by van der Klaauw (2002), who estimated the effect of scholarship and financial aid offers on students' enrollment decisions, the score variable is an underlying index of academic abilities based on various observable characteristics. In the application by van der Klaauw (2002) the cutoffs were known. However, to avoid manipulation by applicants or schools, the construct of such index and the cutoff scores used to differentiate financial offers are not publicly disclosed. Another example is the application by Dell and Querubin (2018) who studied the effectiveness of aerial bombardment, and the 'overwhelming firepower' strategy deployed by the United States to counter insurgency in the Vietnam War. The U.S. Air Force employed a Bayesian algorithm that assigned scores to geographic locations in Vietnam using data from 169 questions, based on which weekly aerial-bombardment allocations were planned. The algorithm generated scores that were continuous in the range from 1 (very insecure) to 5 (very secure). The scores were rounded to the nearest integer for decision-making. The study compared places just below and above the rounding thresholds to isolate the causal effect of bombing. It found that the 'overwhelming firepower' strategy backfired, increasing insurgent activities and recruitment, and worsening attitudes towards the U.S. in regions that barely crossed the threshold to receive a bigger payload of bombs, compared to regions that barely avoided such fate. Dell and Querubin (2018) were able to reconstruct the algorithm and retrieve full information regarding the cutoffs and the assignment rules from declassified documents, whereas the electronic data was preserved by a lucky accident. The data tapes produced by the two IBM 360 computers would have likely been destroyed but were saved as they were subpoenaed during an IBM lawsuit. Dell and Querubin (2018) were able to retrieve the tapes from the U.S. National Archives. However, in situations when full information regarding the cutoffs is not preserved by such happenstance, it is of practical use to have methods at our disposal so that RDDs can be applied even with unknown cutoffs.

A variation of the RDD is Regression Kink Design (RKD), popularized by Card et al. (2008), and Card et al. (2012). In RKD, while the regression function may be continuous, the slope has a discontinuity at a threshold. In Card et al. (2008), the tipping point effect based on the Scheling (1971) model about the dynamic of segregation in city neighborhoods is analyzed. When the minority share in a neighborhood exceeds a tipping point, some whites leave, resulting in the share of whites in the neighborhood to go down further, resulting in more whites leaving in a cascading manner until neighborhoods become segregated by race. The threshold/cutoff value of such a tipping point (kink) is generally unknown. Another example is by Landais (2015) who studied the impact of changes in unemployment insurance benefit level and benefit duration on job-search duration using kinks in the unemployment insurance schedule. Ganong and Jager (2018) provide a comprehensive list of applications of RKD in the economic literature until 2018. Hansen (2017) explored estimation and inference in a regression kink model with an unknown threshold, using methods for inference on non-differentiable functions. RKD detection methods can also have applications beyond economics, such as in the fields of ecology, medicine, epidemiology, and climate studies in the study of tipping-points and critical transitions.

Detection of a kink (discontinuity in the slope of the regression function) in a traditional RDD setup has another useful application. Presence of a kink near the cutoff can indicate presence of a selection effect – indicating that the test subjects under consideration near the cutoff may have been able to manipulate their assignment/score variable to place themselves on the favored side of the cutoff. For example, in the classical application of RDD where students who score above a threshold get a scholarship, if some students just below the cutoff can convince teachers to “mercy pass” them, or if the students are allowed to re-take the exams until they are on their preferred side of the cutoff, this leads to selection bias as the treatment and control groups now differ, and the assignment is no longer random, see Lee and Lemieux (2010). Under this scenario, within a close neighborhood of the cutoff score, students who are more persuasive (in the first case) or have more motivation (in the second case) will be more likely to be just above the threshold than just below the threshold, such that the influence of these factors may cause the regression equation for the outcome variable on the assignment variable to have different slopes on both sides of the cutoff (a kink). It is, therefore, of practical use to be able to not only detect one or more unknown discontinuities, but also to detect whether a particular discontinuity is a discontinuity on the regression function in level (RDD), in slope (RKD), or perhaps both.

Most existing literature on RDDs assumes that the cutoff is known. Porter and Yu (2015) were, to my knowledge, first to propose a two-stage approach for testing and estimation with unknown cutoffs. They attempted to detect the presence of treatment effects in a way that is very close to the nonparametric structural change test inspired by Bierens (1982). The steps involve first estimating the cutoff, using an estimator called the difference kernel estimator (DKE), and secondly estimating the treatment effect, as if the cutoff were known. Their results indicated asymptotic-efficiency of the proposed method. Also, the authors checked for the presence of a kink, or discontinuity in the slope of the regression function, which can be thought of as, depending on the context, a tipping point, or selection effect. Herlands et al. (2018) recently proposed a machine learning approach for automated discovery of localized RDDs and estimation of treatment effect across arbitrarily high dimensional spaces. Their method relies on an iterative search algorithm that imposes localized nearest-neighborhood restrictions, partitions the neighborhood in half, and searches for discontinuity using log-likelihood ratio statistic. The size of the neighborhoods is iterated over and significant discontinuity points are retained.

The proposed approach in my thesis relies on either Andrews' Test by Andrews (1993) or a machine-learning-based structural-break detection method developed by Pretis et al. (2018) for estimation of a `best candidate' for the unknown discontinuity/kink point. Next, the optimal bandwidth-lengths on either side of the estimated cutoff are selected using the MSE-optimal bandwidth selector as suggested by Imbens and Kalyanaraman (2012), Calonico et al. (2014a), Calonico et al. (2018), Calonico et al. (2019), and Calonico et al. (2020). Afterward, the aforementioned machine-learning-based structural-break detection method developed by Pretis et al. (2018) is applied (again) within the bandwidth limit only for estimation of a statistically significant discontinuity/kink as a local average treatment/tipping point effect respectively. Thus, the method involves three sequential steps: estimation of a cutoff (cutoff detection), estimation of a bandwidth (bandwidth selection), and estimation of treatment or tipping effect (effect estimation).

For estimation of a 'best-candidate' for the unknown discontinuity/kink point (cutoff detection), the structural-break detection method developed by Pretis et al. (2018) is more suitable when there is some knowledge on an anticipated cutoff and an associated neighborhood/interval within which the discontinuity/kink point may occur. This method may also be preferred when the data is `messy', with additional structural breaks and outliers apart from the discontinuity/kink point, as it allows for controlling for outliers and additional structural breaks, not in the area of the anticipated cutoff of interest. However, as multiple breakpoints can be detected using the Pretis et al. (2018) approach, some knowledge on the possible cutoff value of interest becomes necessary. I demonstrate this in the considered empirical application.

An advantage of the three-step proposed herein is that, while the approach in Porter and Yu (2015) can test for existence of a selection effect only after excluding the possibility of a nonzero treatment effect (the two-stage testing being sequential), my approach is not bound to such restrictions and can test for either selection effect (discontinuity in the slope of the regression function), or treatment effect (discontinuity in the levels of the regression function), or both.

Throughout this thesis, I focus on the sharp RDD (as opposed to the fuzzy RDD) with a single discontinuity point, as this arrangement simplifies the demonstration of the main idea of the thesis. In sharp RDD the treatment assignment is deterministic whereas in fuzzy RDD it is probabilistic. In other words, the probability of treatment assignment is '0' in one side of the cutoff and 1 on the other side of the cutoff in sharp RDD, whereas fuzzy RDD is a more general case where the probability of treatment jumps discontinuously at the cutoff. It is worth noting that the proposed approach is applicable even when there are multiple unknown discontinuity points, as long as some prior knowledge is available on an interval within which a particular discontinuity/kink point may reside, for example, as in Angrist and Lavy (1999).

My thesis is organized as follows. In section 2, I describe my methodology, in particular the model, the estimation approach, and the expected properties of the procedure. In section 3, I detail the simulation design and present the simulation results. In section 4, I present results of an empirical replication study, ascertaining how my approach compares with the original study. Section 5 concludes the thesis.

2 Methods

2.1 Model

The general model takes the form given in equation (1) where, within an interval $[\gamma - \pi_l, \gamma + \pi_r]$ of the score variable (x),

$$y = z\beta_1 + x\beta_2 + (x - \gamma)_+ \beta_k + \tau\beta_d + \epsilon$$

$$\tau = \begin{cases} 1, & x \in x \geq \gamma \\ 0, & x \in x < \gamma \end{cases} \qquad ...(1)$$

Here, (π) refers to a bandwidth length of the score variable (x), with (π_l) and (π_r) referring to the left and the right bandwidth lengths respectively. This general setup allows for both the presence of a discontinuity or treatment effect (β_d), and/or kink or selection effect (β_k), around the local neighborhood or bandwidth $[\gamma - \pi_l, \gamma + \pi_r]$ of a cutoff value (γ) of the score variable (x). A matrix containing values of all other covariates except the score variable in the columns is denoted by (z). The departure from ordinary RDD estimation in this setup lies in the assumption that the cutoff value (γ) is either not exactly known or unknown, and so needs to be estimated. The standard RDD assumption that all potentially relevant variables other than the treatment variable (τ) and the outcome variable (y) are continuous on either side of the cutoff, is retained. Hence, a violation of this assumption is the presence of a selection effect (the score variable not being continuous at the cutoff), which is captured by the kink coefficient (β_k) not being zero.

This is a non-parametric (local linear regression) set-up. Non-parametric RDD has become increasingly popular as this provides estimates of the local causal effect of a treatment based on data closer to the cut-off and has been shown to have better internal validity over parametric or global RDD. If units are unable to perfectly 'sort' around this cutoff, units with scores barely below the cutoff can be used as a comparison group for units with scores barely above it for estimation of local causal effect according to Cattaneo et al. (2019).

The following illustrates how this general setup can accommodate four possible combinations of Data Generating Processes (DGPs):

- $\beta_d = 0, \beta_k = 0$ – No Treatment or Selection Effect
- $\beta_d \neq 0, \beta_k = 0$ – Discontinuity, presence of a Local Average Treatment Effect (LATE) only.
- $\beta_d = 0, \beta_k \neq 0$ – Kink, presence of a tipping point effect or selection effect only.
- $\beta_d \neq 0, \beta_k \neq 0$ – Both discontinuity and kink, presence of both a local average treatment effect (LATE) and a selection effect.

An alternate specification of the four possible combinations is given in Table 1, where, when the function is of the form: $y = z\beta_1 + x\beta_2 + \epsilon = f + \epsilon$. Porter and Yu (2015) showed a similar table.

Table 1: Behavior of the regression function f and its first derivative f' at Cutoff

	No (Selection/Kink)	(Selection/Kink)
No (Treatment/Discontinuity)	$f_+(\gamma) = f_-(\gamma)$ $f'_+(\gamma) = f'_-(\gamma)$	$f_+(\gamma) = f_-(\gamma)$ $f'_+(\gamma) \neq f'_-(\gamma)$
Treatment/Discontinuity)	$f_+(\gamma) \neq f_-(\gamma)$ $f'_+(\gamma) = f'_-(\gamma)$	$f_+(\gamma) \neq f_-(\gamma)$ $f'_+(\gamma) \neq f'_-(\gamma)$

$f_+(\gamma)$ and $f_-(\gamma)$ refers to the regression function above and below the cutoff (γ) within the bandwidth $f'_+(\gamma)$ and $f'_-(\gamma)$ refers to the slope of the regression function with respect to the score variable (x) above and below the cutoff (γ) within the bandwidth

Gelman and Imbens (2018) suggests using either a local linear or a quadratic polynomial for RDD estimation and argue that to control for higher-order polynomials of the assignment variable in RDD beyond a quadratic polynomial is a flawed approach. In contrast, Pei et al. (2018) challenged the superiority of local linear over higher-order polynomials and called for a computational approach to ascertain the optimal order of the polynomial selection. However, Pei et al. (2018) also noted that reliance on local linear over higher-order estimators is increasing in practice, after surveying leading economics journals from 1999 to 2017. Given this, to apply my estimation approach to the simulated data, I consider a local linear estimator and a quadratic polynomial estimator. It would be interesting to explore estimation using this approach with higher-order polynomial estimators in future research.

2.2 Estimation Approach

The procedure for detecting unknown discontinuity/kink points in my thesis involves three sequential steps. First, a `best candidate' point for the cutoff is estimated (cutoff detection). The second step involves estimating optimum bandwidth-lengths on both sides of the estimated cut-off using a computational approach (bandwidth estimation). The data is then restricted to 'within the bandwidth' for the final step, keeping only observations that fall within the local neighborhood of the estimated cutoff. In the final step, an indicator-saturation based structural-break detection method is applied in the aforementioned local neighborhood of the cutoff towards estimation of the discontinuity/kink (effect estimation).

While the method proposed in Porter and Yu (2015) can test for existence of a selection effect only after excluding the possibility of a nonzero treatment effect (the two-stage testing being sequential), my three-step method is not bound to such restrictions and can test for either selection effect (discontinuity in slope of the regression function), or treatment effect (discontinuity in levels of the regression function), or both. Additionally, my proposed procedure can serve the purpose of a robustness check of RDD when the cutoff point is known a-priori. In the case of known cutoff points, if the estimated cutoff (as if the cutoff were unknown) falls in a nearby interval of the known cutoff, and if the estimated coefficients are similar to those obtained using conventional RDD regression methods, this perhaps suggests stronger evidence of a significant treatment effect/tipping effect. This indicates that the treatment/tipping effect is perhaps sizable enough to cause a structural break and parameter instability at the point of the cutoff, while inconsequential effects are unlikely to be captured using this approach. Details of the steps involved are as follows:

1) The first step (cutoff detection) involves estimating a 'best candidate' point for the unknown discontinuity/kink. This is done using either Andrews' Test by Andrews (1993), or using an indicator-saturation method, as developed by Castle et al. (2015), Hendry et al. (2008), and Pretis et al. (2018). Both of these methods detect structural breaks or outliers. Andrews' Test, which detects a single `best candidate' point for a structural shift, is convenient when the structural break at the discontinuity/kink point is expected to be larger than any other possible structural breaks or outliers in the data. The R package I use to implement Andrews' Test is `strucchange' introduced by Zeileis et al. (2002)[1]. On the other hand, the indicator-saturation method is more applicable when the data may have sizable outliers and structural breaks, apart from the discontinuity/kink point, or an anticipated cutoff value and an associated interval-length is known a-priori and can be approximated, which is often the case

[1] The function I use is `Fstats' (from the `strucchange' package) with a least squares model specification.

with real-world applications even when the exact cutoff may be unknown. The advantage of the indicator-saturation method over Andrews' Test in such a scenario is that it allows for controlling for other outliers and breaks, which may otherwise get picked as the candidate cutoff point. The R package I employ for the indicator-saturation based approach is 'gets', introduced by Pretis et al. (2018)[2]. As both methods apply to indexed/time-series data, a pseudo-index is created by ordering the data based on ascending values of the score variable. The simulation results I present here (Section 3) apply Andrews' Test for this step (as there is a single potential breakpoint by construction), while I apply both methods to real-world data in the empirical results section for comparison (Section 4).

When using Andrews' Test we have, for a parametric model indexed by parameters (β_1,δ_0) for $t = t_1,t_2,\ldots$

- H_0: $\beta_1 = \beta_0$ for all $t \geq 1$ for some $\beta_0 \in B \subset R^P$ (parameter stability)
- H_1: $\beta_1 = \beta_1(c)$ for $t = t_1,t_2, \ldots, t_c$

$$= \beta_2(c) \text{ for } t = t_{c+1},t_{c+2}\ldots \text{ for some } \beta_1(c)\,,\, \beta_2(c) \in B \subset R^P \qquad \ldots(2)$$

The best candidate for the change point index (c), which is unknown, is estimated.

Indictor-saturation method tackles the challenge of detecting outliers and structural breaks in econometric models by starting from a general model, allowing for an outlier or shift at every point (hence, `indicator-saturation') and removing all but significant ones using general-to-specific model selection criteria by applying an automated multi-path search algorithm, see Pretis et al. (2018). The types of indicators I use are: impulse-indicators or IIS for controlling distortionary influence of outliers, see Hendry et al. (2008), and Johansen and Nielsen (2016); step-indicators or SIS for detection of discontinuities, see Castle et al. (2015); and user-designed indicators or UIS for detection of kinks, see Pretis et al. (2016), and Schneider et al. (2017). I start with a general model allowing for a shift or outlier at any observation as follows:

$$y_t = \mu + \sum_{j=1}^{n} l_j 1_{(t=j)} + \sum_{j=2}^{n} d_j 1_{(t\geq j)} + \sum_{j=2}^{n} k_j m_{(t\geq j)} + u_t \qquad \ldots(3)$$

Here, (n) denotes the number of observations inside a selected interval or a local neighborhood within which the discontinuity/kink point, if it exists, is known to occur; (n) can potentially be the number of observations in the whole sample when any prior knowledge of such an interval is absent. The

[2] The function I use is `isat' (from the `gets' package).

parameter (μ) represents the expected value of (y_t). The first summation term provides the impulse-indicators (for detecting outliers and controlling for their distortionary influence in cutoff detection). The second summation term gives the step-indicators (for detecting the discontinuity-cutoff, and also controlling for the distortionary influence of other structural breaks apart from the discontinuity-cutoff). The third summation term represents user-designed indicators, which are used to detect the kink-cutoff. The nominal false-detection rate of the procedure can be set manually via the `isat' function before running the multi-path search algorithm, see Pretis et al. (2018). The nominal false-detection rate in this context refers to the estimated probability of retaining a spurious indicator, in other words, the estimated probability of detecting a step-shift/trend-shift/outlier when there is none.

It is worth mentioning that, without restricting the dataset to within a bandwidth limit around the cutoff, coefficient estimates of step indicators and user-defined indicators detected inside the known interval within which a discontinuity/kink (if exists) is anticipated (based on prior knowledge) are parametric/global estimates of treatment effects and tipping effects respectively. In other words, the magnitude of a break that is returned corresponds to a global effect and not a local effect, as it applies to all data points following the cutoff, not just to those points near the cutoff. However, as RDD has less external validity (treatment effect or tipping effect estimates based on data points far from the cutoff are not valid for causal inference), an optimum bandwidth length needs to be selected on both sides of the estimated cutoff, and the indicator-saturation methods need to be applied (again) in that restricted neighborhood only for a local average treatment effect interpretation. This brings us to step two.

2) The second step (bandwidth selection) involves estimating an optimal bandwidth length on both sides of the estimated cutoff. This is estimated using an MSE-optimal bandwidth selector following the procedure developed in Imbens and Kalyanaraman (2012), Calonico et al. (2014a), Calonico et al. (2018), Calonico et al. (2019), and Calonico et al. (2020). The R package I employ is `rdrobust'[3] introduced in Calonico et al. (2015b). I adopted most of the default argument-options of the `rdrobust' function as detailed in the footnote. The function takes a cutoff value as input along with the model, and outputs both the RD regression (local average causal effect) estimates and the numerically derived

[3] The function I use is `rdrobust' from the `rdrobust' package. The estimated cutoff value from step one enters the function as the cutoff point argument. Default arguments used are strict RD design (as opposed to fuzzy RD design), local linear regression for point estimator, local quadratic regression for bias correction, and MSE optimal bandwidth selector. A uniform kernel, which does not assign more weight to datapoints near the cutoff when constructing the local polynomial estimators but assigns uniform weight, was adopted, as knowledge of the cutoff is not exact. The `deriv' argument specifies the order of the derivative of the regression function to be estimated. The default is `0' which returns the RDD estimates, while setting deriv=1 yields the RKD estimates.

optimal bandwidth lengths on both sides of the cutoff. The MSE-optimal bandwidth selector employed seeks to minimize the Mean Squared Error (MSE) of the local polynomial RD point estimator, given a choice of polynomial order (the default is a linear polynomial). The kernel function is user-specified with the default option being a triangular kernel, which gives more weight to observations closer to the cutoff. As the cutoff is not exact, rather estimated, and is very unlikely to coincide with the true-cutoff if there is a true cutoff, the idea of weighting data points near the estimated cutoff is unsound in the context of this application. Therefore, I adopt a uniform kernel, which assigns equal weight to all observations within the bandwidth that are used in local estimation. The choice of the kernel is trivial. Cattaneo et al. (2019) mention that the estimation and inference results are typically not very sensitive to the particular choice of the kernel used when 'rdrobust' package is employed, as the bandwidth-choice is optimized taking into consideration the particular choice of the kernel. The optimal bandwidth choice equation is shown in Cattaneo et al. (2019) to be:

$$h_{MSE} = \left(\frac{V}{2(p+1)B^2}\right)^{\frac{1}{2p+3}} \quad n^{-1/(2p+3)} \qquad \ldots(4)$$

Here, $(\mathrm{h_{MSE}})$ is the computationally derived optimal bandwidth length that is obtained by minimizing the mean squared error (MSE) of the local polynomial RD point estimator given a choice of polynomial order and kernel function. In the equation, (V) represents the variance of the local polynomial RD point estimator, (B) represents its bias, (p) represents the polynomial order of estimation, and (n) denotes sample size. The optimal bandwidth length decreases with sample size (n) and bias (B) and increases with variance (V). Intuitively, a larger sample allows reducing the error in the approximation by reducing the bandwidth without paying a penalty in added variability. A large asymptotic variance leads to a larger MSE optimal bandwidth as it needs to include more observations for estimation. In contrast, a larger asymptotic bias leads to a smaller MSE-optimal bandwidth as a smaller bandwidth reduces approximation errors and bias of the estimator.

Both the RDD and RKD ecoefficient estimates are also obtained by applying the 'rdrobust' function, following the procedure developed by Calonico et al. (2015b) and Cattaneo et al. (2017), based on the estimated cutoff. This is an alternate approach that is considered for local treatment/tipping effect estimation with unknown cutoff (the third step), in addition to the indicator-saturation approach. I refer to this as the **'RD-regression approach'** in subsequent sections of the thesis. This approach, however, does not produce promising results when an estimated cutoff is used instead of a cutoff that is exact and known a-priori, as this method is found to be sensitive to even small errors in cutoff estimation (see Section 3, simulation results and section 4, empirical results).

3) The third step (effect estimation) involves restricting the data to within the bandwidth and effect estimation. The bandwidth estimates are as detailed in step two, so that this third step only works with data within this interval:

[(estimated cutoff – estimated left bandwidth), (estimated cutoff + estimated right bandwidth)]

The aforementioned automated, machine-learning based, general-to-specific indicator-saturation method (described in details in the first step), is applied to search for statistically significant step-shifts (as discontinuity/local average treatment effect) and shifts in the slope of the regression function (as kink/tipping effect or selection effect). As the application of indicator-saturation in this step (as opposed to in step 1, cutoff detection) is restricted within a local neighborhood, the estimated effect, therefore, has a local average treatment/tipping effect interpretation. I refer to this as the **'Indicator-saturation approach'.**

The aggregate of coefficients of shifts detected within the local neighborhood of the estimated cutoff is taken as an estimate of the treatment effect (for step-indicators) or tipping effect (for user-defined indicators). Monte Carlo simulation of detection and estimation results using this approach show promising properties (section 3). The underlying assumption in aggregating the shifts is that within the local neighborhood of the cutoff there is no other noteworthy discontinuity or kink except at the point of the cutoff (continuity assumption). This is not an additional restrictive assumption required in this method, rather a general assumption in RDD, see Cattaneo et al. (2019).

2.3 Desired Properties

The desired properties of the procedure are that the resulting estimated probability of detecting a false discontinuity/kink (false-detection) is close to the nominal false-detection rate chosen, and the estimated probability of detecting a true discontinuity/kink (true-detection) is higher as the `signal-to-noise' ratio and the sample size increases. Following describes these properties further.

- **False-detection** refers to, detection of discontinuity or kink where they (respectively) do not exist. I simulate this for different scenarios:
  - False-detection of a discontinuity:
    - When neither discontinuity nor kink exists.
    - When a kink exists.
  - False-detection of a kink:
    - When neither discontinuity nor kink exists.
    - When a discontinuity exists.

- **True-detection** refers to, detection of discontinuity or kink where they (respectively) exist. I simulate this property for the following scenarios:
  - Detection of a discontinuity
    - When only a discontinuity exists.
    - When both a discontinuity and a kink exist.
  - Detection of a kink
    - When only a kink exists.
    - When both a discontinuity and a kink exist.

- I use the term **signal-to-noise ratio** to refer to the ratio of the magnitude of discontinuity and the standard deviation of the disturbance term. When this ratio is above 2.58 (so that, the discontinuity is significant at the 1% level or lower when simulating a normally distributed disturbance term), a desired property would be a false-detection rate close to 1%, and a high true-detection rate. As this ratio falls below 2.58, meaning that the discontinuity (treatment effect) or kink (tipping effect) becomes less and less discernable from random noise in the data (so that, the discontinuity is not significant at the nominal 1% level when simulating a normally distributed disturbance term), false-detection rate should increase, and true-detection rate should fall.

Obtaining a closed-form solution to the overall size and power of the testing strategy is complicated by the fact that the procedure relies on three sequential steps, each of which has its own size and power. Hence, I use Monte Carlo simulations to report the estimated probabilities of false-detection and true-detection with synthetic data. More specifically, I generate two classes of synthetic data, where the outcome variable is a linear function of the score variable, and where it is a quadratic function of the score variable. For each of these two classes, I simulate four variants of Data Generating Processes (DGPs) – where there is no discontinuity or kink, where there is a discontinuity only, where there is a kink only, and where there are both. The design of my simulation experiment and results from these experiments are presented in the next section.

3 Simulation

3.1 Simulation Design

Simulation of false-detection rates and true-detection rates of the procedure over a theoretical derivation is motivated by the fact that the method relies on a three-step procedure each of which has its own size and power. In simulating these estimated probabilities, given a DGP, I focus on coefficient estimation (accurate detection of effects and their magnitudes) and not on the cutoff estimation. The reason is two-fold. First, in RDD/RKD, attention is usually on estimating the local average treatment effect/tipping effect, with knowledge of the cutoff value only serving as a means of obtaining coefficient estimates. However, inaccuracy in estimating the cutoff yields bias in the estimator of the effect. Hence, simulating false-detection and true-detection probabilities of the coefficient estimation (the terminal step), already encapsulates the effect of the cutoff estimation, a previous step. Second, the cutoff location estimation approach relies on structural-break detection, properties of which have been studied extensively in the literature, see Andrews (1993) for Andrews' Test and Castle et al. (2015), Hendry et al. (2008), and Pretis et al. (2018) for indicator-saturation.

The testing and estimation procedure are studied using four variants of DGPs (see Table 1) likely to be faced when dealing with real-world data (no effect, discontinuity only, kink only, both discontinuity and kink). For the local linear estimator, the DGPs chosen to represent each case are linearized adaptations of those used by Porter and Yu (2015), who used quadratic DGPs. The DGPs are:

- DGP1.1: No effect

$$y = x + e; \qquad \text{for } (-2 \leq x \leq 3)$$

- DGP1.2: Selection/kink only

$$\begin{aligned} y &= x + e; && \text{for } (-2 \leq x < 1) \\ &= -x + 2 + e && \text{for } (1 \leq x \leq 3) \end{aligned}$$

- DGP1.3: Treatment effect/discontinuity only

$$\begin{aligned} y &= x + e; && \text{for } (-2 \leq x < 1) \\ &= x + 1 + e; && \text{for } (1 \leq x \leq 3) \end{aligned}$$

- DGP1.4: Both selection/kink and treatment effect/discontinuity

$$\begin{aligned} y &= x + e; && \text{for } (-2 \leq x < 1) \\ &= -x + 3 + e && \text{for } (1 \leq x \leq 3) \end{aligned}$$

For the local quadratic estimator, the DGPs are the same as those Porter and Yu (2015) used to study their testing and estimation procedure:

- DGP2.1: No effect

$$y = x^2 + e; \qquad \text{for } (-2 \leq x \leq 3)$$

- DGP2.2: Selection/kink only

$$\begin{aligned} y &= x^2 + e; && \text{for } (-2 \leq x < 1) \\ &= ((x-3)^2 - 3) + e = x^2 - 6x + 6 + e; && \text{for } (1 \leq x \leq 3) \end{aligned}$$

- DGP2.3: Treatment effect/discontinuity only

$$\begin{aligned} y &= x^2 + e; && \text{for } (-2 \leq x < 1) \\ &= (x^2 + 1) + e; && \text{for } (1 \leq x \leq 3) \end{aligned}$$

- DGP2.4: Both selection/kink and treatment effect/discontinuity

$$\begin{aligned} y &= x^2 + e; && \text{for } (-2 \leq x < 1) \\ &= ((x-3)^2 - 2) + e = x^2 - 6x + 7 + e; && \text{for } (1 \leq x \leq 3) \end{aligned}$$

In each case, 1000 replications are used in each simulation experiment for each DGP, for three different sample sizes (200, 500, and 1000) and varying relative magnitudes of discontinuity (as in DGP 1.3, 2.3, & DGP 1.4, 2.4) in the range of 5, 4, 3, and 2 standard deviations of the disturbance term: [e ~ N (0, v^2), v = {0.2, 0.25, 0.33, 0.5}]. Increasing the standard deviation of the disturbance term while keeping the magnitude of the discontinuity/kink fixed is akin to reducing the signal-to-noise ratio of the treatment/tipping effect. The sample size and the number of iterations were restricted to 1000 because the indicator-saturation method, which relies on a machine-learning based multi-path search algorithm, is computation-intensive, and search-time is increasing in sample size.

I show a single draw of each of the DGPs for different relative magnitudes of discontinuity in Figure 1 & Figure 2 for linear and quadratic estimators respectively. The score variable is on the X-axis and the outcome variable is on the Y-axis. Going from left to right are the four variants of DGPs (no effect, a discontinuity cutoff only, a kink cutoff only, a cutoff that is both a discontinuity and a kink). Going from top to bottom, the standard deviation of the disturbance term increases so that the relative magnitude of discontinuity falls. In other words, the signal-to-noise ratio falls going from top to bottom.

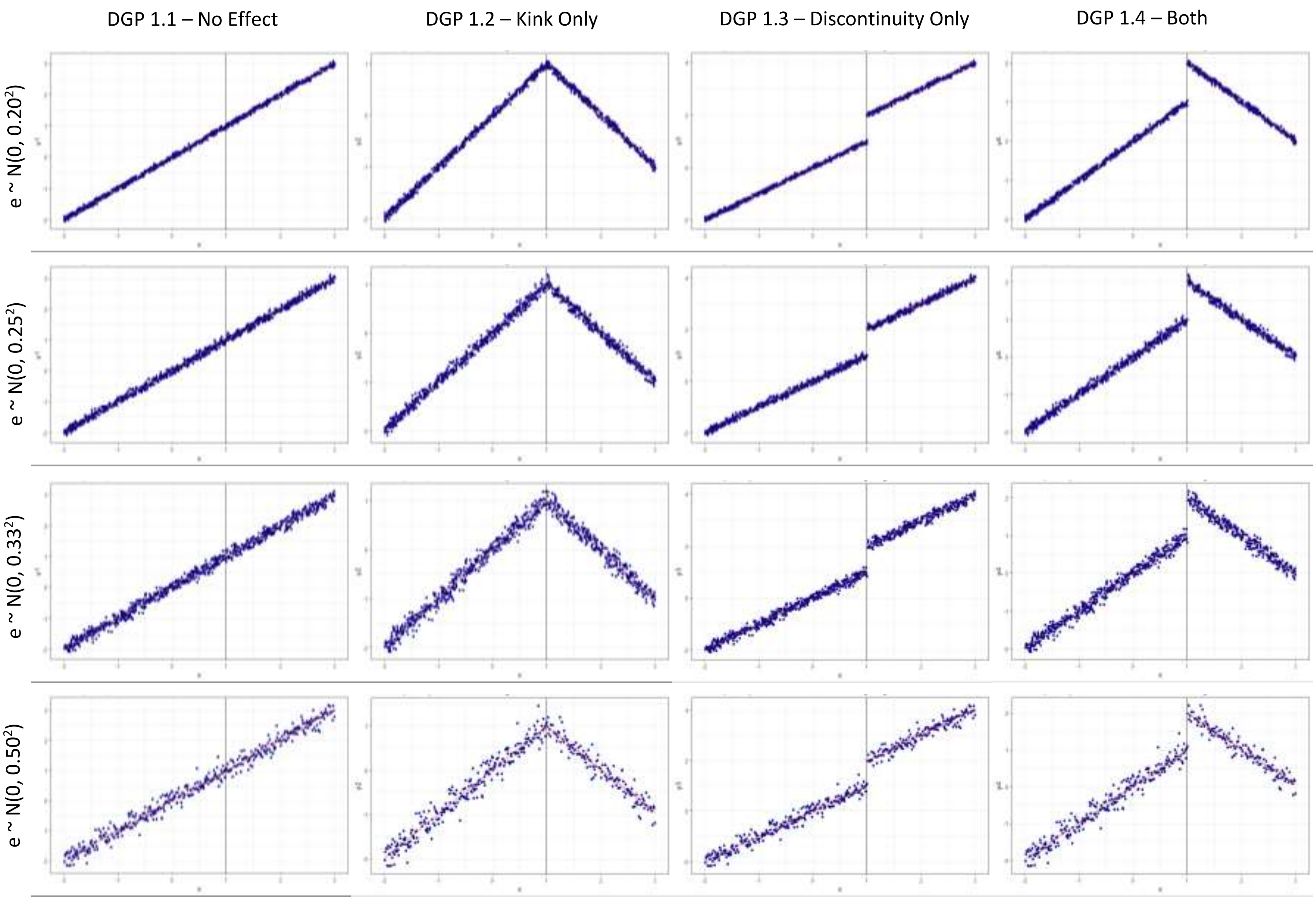


Figure 1: DGPs for Linear estimator with or without selection (kink) or treatment effect (discontinuity).
Standard Deviation of disturbance term calibrated as required to simulate detection with magnitude of discontinuity in the range of 5,4,3, and 2 S.D. of disturbance term.

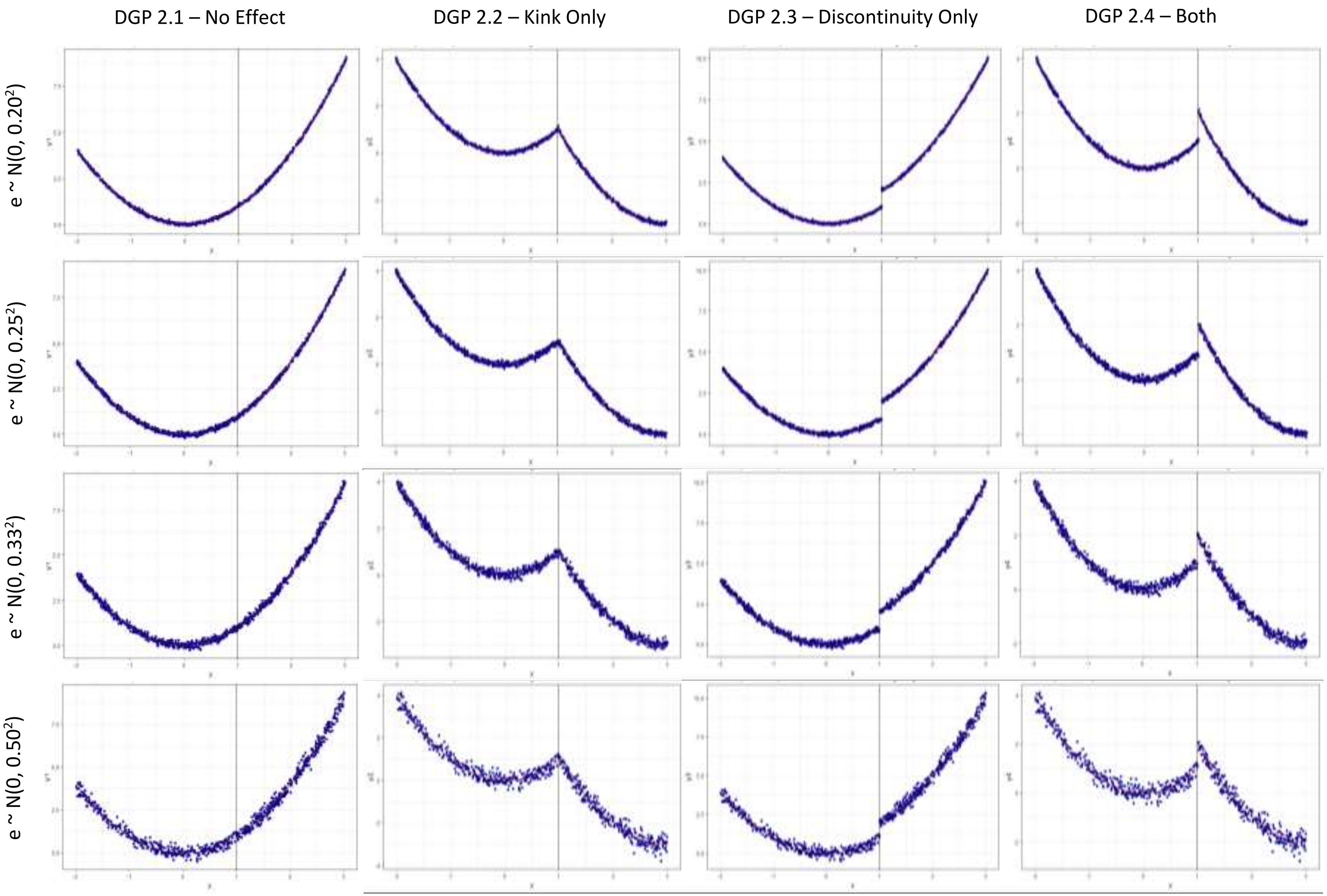


Figure 2: DGPs for Quadratic estimator with or without selection (kink) or treatment effect (discontinuity).
Standard Deviation of disturbance term calibrated as needed to simulate detection with magnitude of discontinuity in the range of 5,4,3, and 2 S.D. of disturbance term.

3.2 Simulation Results

I consider 32 combinations of DGPs in total in my experiment; two classes based on the functional form of the outcome variable with respect to the score variable (linear or quadratic), four variants based on the existence of treatment/tipping effect (no effect, discontinuity only, kink only, or both), and four variants based on the signal-to-noise ratio (relative magnitude of discontinuity being 5, 4, 3, or 2 standard deviations of the disturbance term). First, I run a 'single simulation' – generating one synthetic dataset for each of the 32 combinations of DGPs and applying the cutoff detection and effect estimation procedure once on each dataset. I compare the two approaches for effect estimation when the cutoff is unknown: RD-regression and indicator-saturation. I compare the estimated effects obtained using these two approaches with both the true-effects (designed by construct in the synthetic data) and the estimated effects of conventional RD-regressions when the cutoff is known. Tables 2, 3, 4, & 5 in the following pages show the results of the single simulation.

Results from the single simulation reveal that, while the RD-regression based method shows promise, more often than not detecting 'the presence' of a discontinuity/kink, it performs poorly in estimation. More specifically, there does appear to be significant biases in terms of under-estimation. A bias-correction method may potentially be developed, which is beyond the scope of this thesis. This finding is perhaps not surprising because there are, of course, inaccuracies in the estimated cutoff value. The magnitude of the effect estimates is maximized in RD-regression when the estimated cutoff coincides with the true-cutoff. When this happens, effect estimation with unknown cutoff reduces down to effect estimation using conventional RD-regression where the cutoff is known. Hence, small departure in either direction from the true-cutoff results in an underestimated magnitude of effect estimates.

In contrast, the indicator-saturation approach overcomes this problem of underestimation of effect-coefficients because estimation is not reliant on the exact location of the estimated cutoff within the selected bandwidth. This showed up in the simulated outcomes, with this method performing better than the RD-regression method in both detecting presences of the discontinuity/kink and in generating effect estimates. Furthermore, false-detection of discontinuity/kink is also rare in this single-simulation results. There is only one instance of a false-detection (last row of table 3) out of 24 possible false-detections in the single-simulation result. The false-discontinuity was detected at a low signal-to-noise ratio of 2. At a nominal false detection rate of 1%, this behavior is not unexpected (see section 2.3).

Table 2: Results of Unknown Cutoff Detection (Single Simulation) – (Case: No Discontinuity/Kink in the DGPs)

Magnitude of Discontinuity (if exists)	True Discontinuity Coef	Discontinuity Detection: Est. Discontinuity Coef. — Known Cutoff — RD Regression Approach	Est. Discontinuity Coef. (1) — Estimated Cutoff — RD Regression Approach	Est. Discontinuity Coef. (2) — Estimated Cutoff — Indicator Saturation Approach	True Kink Coef	Kink Detection: Est. Kink Coef. — Known Cutoff — RD Regression Approach	Est. Kink Coef. (1) — Estimated Cutoff — RD Regression Approach	Est. Kink Coef. (2) — Estimated Cutoff — Indicator Saturation Approach
Linear DGP (DGP 1.1) - Discontinuity = FALSE, Kink = FALSE								
5 s.d.	0	-0.001 (0.033) -0.004 [0.039]	-0.049* (0.026) -0.05* [0.03]	0 †††	0	-0.01 (0.13) 0.019 [0.196]	-0.133 (0.138) -0.212 [0.208]	0 †††
4 s.d.	0	-0.004 (0.04) -0.009 [0.047]	-0.076* (0.041) -0.079 [0.047]	0 †††	0	-0.035 (0.164) -0.002 [0.255]	-0.206 (0.217) -0.33 [0.325]	0 †††
3 s.d.	0	-0.011 (0.054) -0.019 [0.064]	-0.135* (0.073) -0.14 [0.085]	0 †††	0	-0.116 (0.249) -0.083 [0.401]	-0.366 (0.391) -0.586 [0.582]	0 †††
2 s.d.	0	-0.024 (0.093) -0.038 [0.111]	-0.304* (0.167) -0.315 [0.195]	0 †††	0	-0.309 (0.478) -0.276 [0.806]	-0.826 (0.894) -1.32 [1.325]	0 †††
Quadratic DGP (DGP 2.1) - Discontinuity = FALSE, Kink = FALSE								
5 s.d.	0	0.003 (0.053) -0.002 [0.063]	-0.048 (0.033) -0.052 [0.038]	0 †††	0	0.115 (0.638) -0.388 [0.678]	0.537 (0.444) 0.115 [0.478]	0 †††
4 s.d.	0	0 (0.062) -0.007 [0.072]	-0.075* (0.044) -0.081 [0.051]	0 †††	0	0.001 (0.606) -0.518 [0.663]	0.56 (0.438) 0.044 [0.494]	0 †††
3 s.d.	0	-0.007 (0.075) -0.018 [0.087]	-0.134* (0.073) -0.141* [0.084]	0 †††	0	-0.052 (0.576) -0.611 [0.672]	0.461 (0.529) -0.221 [0.664]	0 †††
2 s.d.	0	-0.021 (0.112) -0.041 [0.133]	-0.304 (0.165) -0.316 [0.191]	0 †††	0	-0.207 (0.7) -0.844 [0.968]	0.026 (0.976) -1.018 [1.354]	0 †††

Green: Estimated coefficient within 2 s.d. of disturbance term | Red: Estimated coefficient NOT within 2 s.d. of disturbance term

(1) Cutoff point estimated using Andrew's test ('Fstats' function of 'strucchange' package in R, with default parameters).
Then, RD & Kink Regression run based on estimated cutoff ('rdrobust' function of 'rdrobust' package in R, with default parameters).
Both conventional & robust (bias-corrected) coefs. & std. errors reported. *, **, *** represent statisticial significance at 10%, 5%, and 1% levels.

(2) Cutoff point estimated using Andrew's test ('Fstats' function of 'strucchange' package in R, with default parameters).
Then, optimal bandwidth around estimated cutoff estimated ('rdbwselect' function of 'rdrobust package in R, with default parameters).
Then, Indicator saturation applied within estimated bandwidth ('isat' function of 'gets' pacakge in R at 1% nominal false detection rate - denoted †††).

As observable from the results presented in Table 2, when there was no discontinuity and no kink in the true DGP, the RD-regression approach, based on the estimated cutoff point, detected no statistically significant discontinuity or kink when using both the linear and quadratic estimators. The indicator-saturation procedure also detected no discontinuity or kink at the 1% nominal false-detection rate, suggesting that we would conclude no discontinuity or kink.

Table 3: Results of Unknown Cutoff Detection (Single Simulation) – (Case: Kink only in the DGPs)

		Discontinuity Detection				Kink Detection		
Magnitude of Discontinuity (if exists)	True Discontinuity Coef	Est. Discontinuity Coef. **Known** Cutoff **RD Regression** Approach	Est. Discontinuity Coef. (1) **Estimated** Cutoff **RD Regression** Approach	Est. Discontinuity Coef. (2) **Estimated** Cutoff **Indicator Saturation** Approach	True Kink Coef	Est. Kink Coef. **Known** Cutoff **RD Regression** Approach	Est. Kink Coef. (1) **Estimated** Cutoff **RD Regression** Approach	Est. Kink Coef. (2) **Estimated** Cutoff **Indicator Saturation** Approach
Linear DGP (DGP 1.2) - Discontinuity = FALSE, Kink = TRUE								
5 s.d.	0	-0.001 (0.035) -0.004 [0.041]	-0.038 (0.036) -0.041 [0.043]	0.339 †††	-2	-2.008*** (0.132) -1.979*** [0.199]	-2.003*** (0.134) -1.979*** [0.202]	-2.006 †††
4 s.d.	0	-0.003 (0.042) -0.009 [0.049]	-0.043 (0.044) -0.048 [0.052]	0.267 †††	-2	-2.031*** (0.167) -1.996*** [0.258]	-2.022*** (0.175) -1.988*** [0.267]	-2.391 †††
3 s.d.	0	-0.011 (0.056) -0.019 [0.067]	-0.126 (0.107) -0.137 [0.123]	-0.619 †††	-2	-2.111*** (0.255) -2.075*** [0.407]	-2.168*** (0.566) -2.106*** [0.787]	-0.431 †††
2 s.d.	0	-0.024 (0.096) -0.038 [0.115]	-0.191 (0.186) -0.223 [0.214]	-0.433 †††	-2	-2.298*** (0.484) -2.259** [0.812]	-2.606*** (0.99) -2.661** [1.448]	-0.433 †††
Quadratic DGP (DGP 2.2) - Discontinuity = FALSE, Kink = TRUE								
5 s.d.	0	0.003 (0.061) 0.001 [0.073]	0.111 (0.084) 0.089 [0.097]	-0.019 †††	-6	-5.804 (0.634) -6.343 [0.68]	-5.579*** (0.791) -6.083*** [0.883]	-5.701 †††
4 s.d.	0	0.003 (0.07) -0.004 [0.083]	0.12 (0.092) 0.097 [0.107]	-0.02 †††	-6	-5.92*** (0.621) -6.467*** [0.687]	-5.635*** (0.774) -6.164*** [0.886]	-5.684 †††
3 s.d.	0	-0.004 (0.087) -0.015 [0.102]	0.134 (0.11) 0.112 [0.129]	-0.085 †††	-6	-5.955*** (0.6) -6.524*** [0.71]	-5.622*** (0.748) -6.189*** [0.908]	-5.455 †††
2 s.d.	0	-0.021 (0.128) -0.043 [0.151]	0.509*** (0.128) 0.524*** [0.158]	-2.304 †††	-6	-6.154*** (0.742) -6.785*** [1.013]	-4.728*** (0.685) -4.992*** [1.112]	-0.554 †††

Green: Estimated coefficient within 2 s.d. of disturbance term | Red: Estimated coefficient NOT within 2 s.d. of disturbance term

(1) Cutoff point estimated using Andrew's test ('Fstats' function of 'strucchange' package in R, with default parameters).
Then, RD & Kink Regression run based on estimated cutoff ('rdrobust' function of 'rdrobust' package in R, with default parameters).
Both conventional & robust (bias-corrected) coefs. & std. errors reported. *, **, *** represent statisticial significance at 10%, 5%, and 1% levels.

(2) Cutoff point estimated using Andrew's test ('Fstats' function of 'strucchange' package in R, with default parameters).
Then, optimal bandwidth around estimated cutoff estimated ('rdbwselect' function of 'rdrobust package in R, with default parameters).
Then, Indicator saturation applied within estimated bandwidth ('isat' function of 'gets' pacakge in R at 1% nominal false detection rate - denoted †††).

From the outcomes reported in Table 3, when there was no discontinuity but a kink in the true DGP, both procedures detected no statistically significant discontinuity except when using the quadratic estimator with a high dispersion for the disturbance term. Presence of a kink was correctly detected when employing the RD-regression approach, although bias of the estimator increased as the data became noisier (higher dispersion for the disturbance term). The indicator-saturation method failed to detect the presence of a kink as the data became noisier and the 'signal to noise ratio' fell.

Table 4: Results of Unknown Cutoff Detection (Single Simulation) – (Case: Discontinuity only in the DGPs)

Magnitude of Discontinuity (if exists)	True Discontinuity Coef	Discontinuity Detection: Est. Discontinuity Coef. — Known Cutoff — RD Regression Approach	Discontinuity Detection: Est. Discontinuity Coef. (1) — Estimated Cutoff — RD Regression Approach	Discontinuity Detection: Est. Discontinuity Coef. (2) — Estimated Cutoff — Indicator Saturation Approach	True Kink Coef	Kink Detection: Est. Kink Coef. — Known Cutoff — RD Regression Approach	Kink Detection: Est. Kink Coef. (1) — Estimated Cutoff — RD Regression Approach	Kink Detection: Est. Kink Coef. (2) — Estimated Cutoff — Indicator Saturation Approach
Linear DGP (DGP 1.3) - Discontinuity = TRUE, Kink = FALSE								
5 s.d.	1	0.999*** (0.033) 0.996*** [0.039]	0.693** (0.318) 0.665* [0.346]	1.004 †††	0	-0.01 (0.13) 0.019 [0.196]	1.246 (1.342) 1.907 [2.006]	0 †††
4 s.d.	1	0.996*** (0.04) 0.991*** [0.047]	0.703** (0.32) 0.675* [0.349]	1.003 †††	0	-0.035 (0.164) -0.002 [0.255]	1.169 (1.356) 1.817 [2.032]	0 †††
3 s.d.	1	0.989*** (0.054) 0.981*** [0.064]	0.725** (0.325) 0.697* [0.357]	1.006 †††	0	-0.116 (0.249) -0.083 [0.401]	1.003 (1.397) 1.616 [2.103]	0 †††
2 s.d.	1	0.976*** (0.093) 0.962*** [0.111]	0.792** (0.351) 0.764* [0.393]	1.024 †††	0	-0.309 (0.478) -0.276 [0.806]	0.558 (1.565) 1.063 [2.39]	0 †††
Quadratic DGP (DGP 2.3) - Discontinuity = TRUE, Kink = FALSE								
5 s.d.	1	1.003*** (0.053) 0.998*** [0.063]	0.709** (0.339) 0.666* [0.37]	0.997 †††	0	0.115 (0.638) -0.388 [0.678]	1.925 (1.371) 1.866 [2.073]	0 †††
4 s.d.	1	1*** (0.062) 0.993*** [0.072]	0.719** (0.342) 0.676* [0.374]	0.995 †††	0	0.001 (0.606) -0.518 [0.663]	1.862 (1.39) 1.793 [2.104]	0 †††
3 s.d.	1	0.993*** (0.075) 0.982*** [0.087]	0.742** (0.349) 0.699** [0.384]	1.009 †††	0	-0.052 (0.576) -0.611 [0.672]	1.721 (1.44) 1.626 [2.187]	0 †††
2 s.d.	1	0.979*** (0.112) 0.959*** [0.133]	0.808** (0.379) 0.766** [0.424]	0.978 †††	0	-0.207 (0.7) -0.844 [0.968]	1.317 (1.639) 1.131 [2.511]	0 †††

Green: Estimated coefficient within 2 s.d. of disturbance term | Red: Estimated coefficient NOT within 2 s.d. of disturbance term

(1) Cutoff point estimated using Andrew's test ('Fstats' function of 'strucchange' package in R, with default parameters).
Then, RD & Kink Regression run based on estimated cutoff ('rdrobust' function of 'rdrobust' package in R, with default parameters).
Both conventional & robust (bias-corrected) coefs. & std. errors reported. *, **, *** represent statisticial significance at 10%, 5%, and 1% levels.

(2) Cutoff point estimated using Andrew's test ('Fstats' function of 'strucchange' package in R, with default parameters).
Then, optimal bandwidth around estimated cutoff estimated ('rdbwselect' function of 'rdrobust package in R, with default parameters).
Then, Indicator saturation applied within estimated bandwidth ('isat' function of 'gets' pacakge in R at 1% nominal false detection rate - denoted †††).

From the outcomes presented in Table 4, when there was no kink but an unknown discontinuity in the true DGP, it seems that indicator-saturation procedure strictly dominated the RD-regression approach in both the linear and quadratic regressor DGPs. Both the presence and the magnitude of the discontinuity were accurately detected when using the indicator-saturation method, while no case of false-kink detection was found, even for small signal-to-noise ratio (magnitude of discontinuity being as low as 2 standard deviations of the disturbance term). In contrast, the RD-regression procedure successfully detected the presence of a discontinuity, and the detected (nonexistent) kinks were also not reported as statistically significant, but the coefficient values were under-estimated for the discontinuity and over-estimated for the kink compared to the case when the cutoff point is known.

Table 5: Results of Unknown Cutoff Detection (Single Simulation) – (Case: Discontinuity & Kink in the DGPs)

Magnitude of Discontinuity (if exists)	True Discontinuity Coef	Discontinuity Detection: Est. Discontinuity Coef. **Known** Cutoff **RD Regression** Approach	Discontinuity Detection: Est. Discontinuity Coef. (1) **Estimated** Cutoff **RD Regression** Approach	Discontinuity Detection: Est. Discontinuity Coef. (2) **Estimated** Cutoff **Indicator Saturation** Approach	True Kink Coef	Kink Detection: Est. Kink Coef. **Known** Cutoff **RD Regression** Approach	Kink Detection: Est. Kink Coef. (1) **Estimated** Cutoff **RD Regression** Approach	Kink Detection: Est. Kink Coef. (2) **Estimated** Cutoff **Indicator Saturation** Approach
Linear DGP (DGP 1.4) - Discontinuity = TRUE, Kink = TRUE								
5 s.d.	1	0.999*** (0.035) 0.996*** [0.041]	0.713** (0.303) 0.682** [0.332]	1.005 †††	-2	-2.008*** (0.132) -1.979*** [0.199]	-0.674 (1.309) 0.042 [1.976]	-2.016 †††
4 s.d.	1	0.997*** (0.042) 0.991*** [0.049]	0.723** (0.306) 0.691** [0.335]	1.008 †††	-2	-2.031*** (0.167) -1.996*** [0.258]	-0.76 (1.323) -0.068 [2.003]	-2.016 †††
3 s.d.	1	0.989*** (0.056) 0.981*** [0.067]	0.745** (0.312) 0.712** [0.344]	1.015 †††	-2	-2.111*** (0.255) -2.075*** [0.407]	-0.944 (1.362) -0.306 [2.073]	-2.044 †††
2 s.d.	1	0.976*** (0.096) 0.962*** [0.115]	0.811** (0.34) 0.779** [0.382]	1.033 †††	-2	-2.298*** (0.484) -2.259*** [0.812]	-1.415 (1.539) -0.917 [2.37]	-2.098 †††
Quadratic DGP (DGP 2.4) - Discontinuity = TRUE, Kink = TRUE								
5 s.d.	1	1.003*** (0.061) 1.001*** [0.073]	0.767*** (0.298) 0.719** [0.328]	1.005 †††	-6	-5.804*** (0.634) -6.343*** [0.68]	-3.923*** (1.235) -3.801** [1.926]	-6.014 †††
4 s.d.	1	1.003*** (0.07) 0.996*** [0.083]	0.776*** (0.302) 0.728** [0.333]	1.007 †††	-6	-5.92*** (0.621) -6.467*** [0.687]	-4*** (1.261) -3.913** [1.967]	-6.022 †††
3 s.d.	1	0.996*** (0.087) 0.985*** [0.102]	0.797*** (0.312) 0.748** [0.347]	1.013 †††	-6	-5.955*** (0.6) -6.524*** [0.71]	-4.168*** (1.329) -4.158** [2.072]	-6.038 †††
2 s.d.	1	0.979*** (0.128) 0.957*** [0.151]	0.864*** (0.349) 0.813** [0.395]	1.03 †††	-6	-6.154*** (0.742) -6.785*** [1.013]	-4.643*** (1.581) -4.804** [2.459]	-6.086 †††

Green: Estimated coefficient within 2 s.d. of disturbance term | Red: Estimated coefficient NOT within 2 s.d. of disturbance term

(1) Cutoff point estimated using Andrew's test ('Fstats' function of 'strucchange' package in R, with default parameters).
Then, RD & Kink Regression run based on estimated cutoff ('rdrobust' function of 'rdrobust' package in R, with default parameters).
Both conventional & robust (bias-corrected) coefs. & std. errors reported. *, **, *** represent statisticial significance at 10%, 5%, and 1% levels.

(2) Cutoff point estimated using Andrew's test ('Fstats' function of 'strucchange' package in R, with default parameters).
Then, optimal bandwidth around estimated cutoff estimated ('rdbwselect' function of 'rdrobust package in R, with default parameters).
Then, Indicator saturation applied within estimated bandwidth ('isat' function of 'gets' pacakge in R at 1% nominal false detection rate - denoted †††).

Outcomes reported in Table 5, for when there are both a discontinuity and a kink in the true DGP, show that the indicator-saturation based detection approach strictly dominated the RD-regression method for both the linear and quadratic regressor DGPs, at least for the cases I explored. The indicator-saturation detection procedure accurately detected both the presence and the magnitude of the discontinuity and the kink, even for DGPs with a small signal-to-noise ratio. The RD-regression approach was more often than not able to detect the presence of both the discontinuity and the kink, but this method systematically underestimated their magnitude compared to the case when the change point is known a-priori.

As the indicator-saturation approach seems more promising based on the single-simulation result reported in the above tables, I proceed with the indicator-saturation method only in the Monte-Carlo simulations to explore its large-sample behavior. Tables 6 & 7 presents the results of 1000 replications in each simulation experiment for three sample sizes (200, 500, 1000) using the linear and quadratic regressor DGPs, respectively, when I aim to explore the large-sample behavior of the indicator-saturation approach. I concentrate on reporting the estimated false-detection rates and the estimated true detection rates in terms of detecting the 'presence' of discontinuity/kink. For estimating false-detection rates, in simulated DGPs where the true discontinuity/kink coefficient is '0', coefficient estimate falling beyond 1.96 standard deviations of the disturbance term from '0' is taken as a false-detection. For estimating true-detection rates, in DGPs with a discontinuity/kink, coefficient estimate falling beyond 1.96 standard deviations of the disturbance term from '0' in the direction of the true-coefficient is taken as a true-detection. This is motivated by the fact that the disturbance terms are, by the design of the DGPs, normally distributed. The DGPs are generated as linear/quadratic functions of the score variable and an additive normally distributed disturbance term, with a possible kink/discontinuity/or both. All departures from the expected values once the DGPs have been modeled as linear/quadratic functions, therefore, come from either the stochastic disturbance term or the discontinuity/kink. Therefore, for estimating false-detection probabilities (in DGPs where there is no effect), coefficient estimates falling within 1.96 standard deviations of the disturbance term from '0' are taken as 'random noise', whereas estimates falling beyond that range are taken as false-signals (5% nominal significance level). Similarly, for estimating probabilities of true-detection (in DGPs where there is an effect), discontinuity/kink coefficient estimates falling beyond 1.96 standard deviations in the direction of the true discontinuity/kink coefficients is taken as a true-detection of effect-presence.

While the indicator-saturation algorithm is run in this experiment at a nominal false-detection rate of 1%, the choice of an interval of 1.96 standard deviations of the disturbance term requires further justification, as it corresponds to a 5% nominal significance level. This is because, while the indicator-saturation is run at a 1% nominal false detection rate, it corresponds to the nominal false detection rate of the indicator-saturation algorithm only, whereas it is one of the three-steps in my procedure. I, therefore, expect a slightly higher false detection rate. Nevertheless, I also generate estimates of false-detection and true detection rates using an interval of 2.58 standard deviation of the disturbance term, which corresponds to a 1% nominal significance level, as a sensitivity test (see Appendix A). Overall, As expected, when the interval corresponding to 1% nominal significance is used, there is a small decrease in both rates. The exception is when the sample size is smaller (200 or 500) and signal-to-noise ratio is smaller (3 or 2), for which there is a significant drop in true detection rates for a small decrease in false-detection rates in some cases, which is also as expected.

Table 6: Simulated False-Detection Rates & True-Detection Rates for Linear Funcitional Form DGPs

Magnitude of Discontinuity (if exists)	DGPs (Linear)	Parameter	signal detection interval: 1.96 s.d. of disturbance n = 200	n = 500	n = 1000
5 s.d	DGP1.1 Discontinuity/treatment effect = FALSE Kink/selection effect = FALSE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	0.0%	0.9%	0.7%
		False Detection - Kink (Detecting selection effect/kink when there is none)	0.0%	0.8%	0.5%
	DGP1.2 Discontinuity/treatment effect = FALSE Kink/selection effect = TRUE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	0.2%	2.2%	1.2%
		True Detection - Kink (Detecting selection effect/kink when there is one)	99.9%	98.9%	99.9%
	DGP1.3 Discontinuity/treatment effect = TRUE Kink/selection effect = FALSE	True Detection - Discontinuity (Detecting treatment effect when there is one)	100.0%	100.0%	100.0%
		False Detection - Kink (Detecting selection effect/kink when there is none)	0.0%	0.1%	0.0%
	DGP1.4 Discontinuity/treatment effect = TRUE Kink/selection effect = TRUE	True Detection - Discontinuity (Detecting treatment effect when there is one)	99.9%	99.8%	100.0%
		True Detection - Kink (Detecting selection effect/kink when there is one)	99.9%	99.9%	100.0%
4 s.d	DGP1.1 Discontinuity/treatment effect = FALSE Kink/selection effect = FALSE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	0.0%	1.1%	1.4%
		False Detection - Kink (Detecting selection effect/kink when there is none)	0.0%	0.8%	0.9%
	DGP1.2 Discontinuity/treatment effect = FALSE Kink/selection effect = TRUE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	0.5%	4.4%	0.9%
		True Detection - Kink (Detecting selection effect/kink when there is one)	99.6%	97.2%	99.9%
	DGP1.3 Discontinuity/treatment effect = TRUE Kink/selection effect = FALSE	True Detection - Discontinuity (Detecting treatment effect when there is one)	100.0%	100.0%	100.0%
		False Detection - Kink (Detecting selection effect/kink when there is none)	0.3%	0.2%	0.0%
	DGP1.4 Discontinuity/treatment effect = TRUE Kink/selection effect = TRUE	True Detection - Discontinuity (Detecting treatment effect when there is one)	99.9%	99.7%	99.9%
		True Detection - Kink (Detecting selection effect/kink when there is one)	99.9%	99.7%	99.9%
3 s.d	DGP1.1 Discontinuity/treatment effect = FALSE Kink/selection effect = FALSE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	0.1%	2.4%	2.0%
		False Detection - Kink (Detecting selection effect/kink when there is none)	0.0%	1.2%	1.5%
	DGP1.2 Discontinuity/treatment effect = FALSE Kink/selection effect = TRUE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	0.9%	12.8%	4.6%
		True Detection - Kink (Detecting selection effect/kink when there is one)	98.8%	82.6%	94.4%
	DGP1.3 Discontinuity/treatment effect = TRUE Kink/selection effect = FALSE	True Detection - Discontinuity (Detecting treatment effect when there is one)	99.9%	99.3%	100.0%
		False Detection - Kink (Detecting selection effect/kink when there is none)	0.5%	3.8%	1.1%
	DGP1.4 Discontinuity/treatment effect = TRUE Kink/selection effect = TRUE	True Detection - Discontinuity (Detecting treatment effect when there is one)	99.2%	96.1%	99.1%
		True Detection - Kink (Detecting selection effect/kink when there is one)	99.2%	95.5%	98.8%
2 s.d	DGP1.1 Discontinuity/treatment effect = FALSE Kink/selection effect = FALSE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	0.1%	4.5%	4.1%
		False Detection - Kink (Detecting selection effect/kink when there is none)	0.1%	2.9%	2.8%
	DGP1.2 Discontinuity/treatment effect = FALSE Kink/selection effect = TRUE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	0.9%	13.4%	9.5%
		True Detection - Kink (Detecting selection effect/kink when there is one)	0.2%	24.5%	36.7%
	DGP1.3 Discontinuity/treatment effect = TRUE Kink/selection effect = FALSE	True Detection - Discontinuity (Detecting treatment effect when there is one)	0.5%	54.0%	60.7%
		False Detection - Kink (Detecting selection effect/kink when there is none)	0.2%	0.4%	0.0%
	DGP1.4 Discontinuity/treatment effect = TRUE Kink/selection effect = TRUE	True Detection - Discontinuity (Detecting treatment effect when there is one)	0.8%	63.1%	82.0%
		True Detection - Kink (Detecting selection effect/kink when there is one)	0.5%	34.5%	46.5%

Green: False-Detection < 1% / True-Detection > 99% | **Yellow - False-Detection < 5% / True-Detection > 95%** | **Red: False-Detection > 5% / True-Detection < 95%**

(Indicator saturation method applied at 1% nominal false-detection rate)

Table 6 reports results for simulated false-detection and true-detection rates for the indicator-saturation approach based on 1,000 replications for each simulation experiment when the DGPs are linear functions of the score at signal-to-noise ratios of 2,3,4, and 5. I expect the false-detection rate of the three-step procedure to be slightly higher than 1% when I set a nominal false-detection rate of 1% for the indicator-saturation algorithm (see page 24). In the table, I color-code estimated false-detection rates below 1% in green, and above 1% but not exceeding 5% in yellow. Small false detection rates/large true detection rates are not expected at signal-to-noise ratios below 2.58 (see section 2.3).

False-detection of a discontinuity or a kink when neither exists and of a kink when a discontinuity exists, stay less than 5% for all simulated experiments as is expected. However, false-detection of a discontinuity when a kink exists exceeds 5% in the particular case when the sample size is 500 and the signal-to-noise ratio is 3, and estimated true-detection rate, in this case, is also small. Overall, however, false-detection of a discontinuity when there is a kink, stays below 5% in the larger sample size of 1000 for all considered signal-to-noise ratios above 2, and close to 1% for all considered signal-to-noise ratios above 3. Therefore, it seems a sample size of 1000 or more is required for the case when there is a kink but no discontinuity, to have a false-detection rate that is close to the nominal false-detection rate. Overall, the estimated false detection rates are mostly close to 1%, and generally falling with larger samples, albeit nonlinearly at times (increasing as the sample size goes to 500 from 200 but decreasing as it goes to 1000 from 500). A possible explanation for the very small estimated false-detection rates that is not close to the nominal false-detection rate at the small sample size of 200 is perhaps because false-detections can be driven by outliers. For this small sample size, the possibility of an outlier in the normally distributed disturbance term is significantly smaller.

True-detection of a discontinuity, irrespective of whether a kink exists or not, remains above 95% at this nominal false detection rate of 1% for all signal-to-noise ratios above 2, and above 99% for all signal-to-noise ratios above 3, for all three sample sizes. This estimated rate seems to increase as the sample size increases and as the signal-to-noise ratio increases. True detection of a kink irrespective of whether a discontinuity exists or not shows similar high true-detection rates, except in the specific case when the sample size is 500 and the signal-to-noise ratio is 3. As most econometric applications of RDD are interested in estimating the treatment effect and most real-world data seems to exhibit linear behavior at least in the local neighborhood of the treatment, and as the linear functional form for RDD is increasingly becoming popular in empirical econometric applications in place of higher-order polynomials of the score variable, see Pei et al. (2018), the simulation results presented here are promising for treatment effect estimation with unknown cutoff. Kink (tipping effect) detection results are also encouraging, for which larger sample size (1000 or more) seems warranted.

Table 7: Simulated False-Detection Rates & True-Detection Rates for Quadratic Functional Form DGPs

Magnitude of Discontinuity (if exists)	DGPs (Quadratic)	Parameter	signal detection interval: 1.96 s.d. of disturbance n = 200	n = 500	n = 1000
5 s.d	DGP2.1 Discontinuity/treatment effect = FALSE Kink/selection effect = FALSE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	0.9%	0.7%	0.3%
		False Detection - Kink (Detecting selection effect/kink when there is none)	0.8%	0.7%	0.3%
	DGP2.2 Discontinuity/treatment effect = FALSE Kink/selection effect = TRUE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	15.6%	34.5%	23.6%
		True Detection - Kink (Detecting selection effect/kink when there is one)	92.5%	84.7%	89.9%
	DGP2.3 Discontinuity/treatment effect = TRUE Kink/selection effect = FALSE	True Detection - Discontinuity (Detecting treatment effect when there is one)	100.0%	100.0%	100.0%
		False Detection - Kink (Detecting selection effect/kink when there is none)	7.5%	0.1%	0.0%
	DGP2.4 Discontinuity/treatment effect = TRUE Kink/selection effect = TRUE	True Detection - Discontinuity (Detecting treatment effect when there is one)	94.9%	98.9%	100.0%
		True Detection - Kink (Detecting selection effect/kink when there is one)	97.2%	99.8%	100.0%
4 s.d	DGP2.1 Discontinuity/treatment effect = FALSE Kink/selection effect = FALSE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	1.6%	0.5%	0.9%
		False Detection - Kink (Detecting selection effect/kink when there is none)	1.4%	0.2%	0.7%
	DGP2.2 Discontinuity/treatment effect = FALSE Kink/selection effect = TRUE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	19.1%	31.8%	28.2%
		True Detection - Kink (Detecting selection effect/kink when there is one)	90.6%	86.3%	86.1%
	DGP2.3 Discontinuity/treatment effect = TRUE Kink/selection effect = FALSE	True Detection - Discontinuity (Detecting treatment effect when there is one)	99.8%	100.0%	100.0%
		False Detection - Kink (Detecting selection effect/kink when there is none)	12.8%	1.0%	0.2%
	DGP2.4 Discontinuity/treatment effect = TRUE Kink/selection effect = TRUE	True Detection - Discontinuity (Detecting treatment effect when there is one)	94.3%	98.4%	99.9%
		True Detection - Kink (Detecting selection effect/kink when there is one)	97.1%	99.7%	99.8%
3 s.d	DGP2.1 Discontinuity/treatment effect = FALSE Kink/selection effect = FALSE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	2.0%	2.4%	2.3%
		False Detection - Kink (Detecting selection effect/kink when there is none)	1.7%	1.6%	0.9%
	DGP2.2 Discontinuity/treatment effect = FALSE Kink/selection effect = TRUE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	27.7%	35.0%	30.1%
		True Detection - Kink (Detecting selection effect/kink when there is one)	82.4%	78.3%	80.8%
	DGP2.3 Discontinuity/treatment effect = TRUE Kink/selection effect = FALSE	True Detection - Discontinuity (Detecting treatment effect when there is one)	94.2%	99.3%	99.5%
		False Detection - Kink (Detecting selection effect/kink when there is none)	18.8%	4.6%	1.7%
	DGP2.4 Discontinuity/treatment effect = TRUE Kink/selection effect = TRUE	True Detection - Discontinuity (Detecting treatment effect when there is one)	85.2%	96.9%	99.4%
		True Detection - Kink (Detecting selection effect/kink when there is one)	94.4%	98.8%	99.1%
2 s.d	DGP2.1 Discontinuity/treatment effect = FALSE Kink/selection effect = FALSE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	4.6%	3.9%	4.2%
		False Detection - Kink (Detecting selection effect/kink when there is none)	2.8%	2.5%	2.3%
	DGP2.2 Discontinuity/treatment effect = FALSE Kink/selection effect = TRUE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	56.3%	44.4%	43.8%
		True Detection - Kink (Detecting selection effect/kink when there is one)	40.2%	41.9%	49.4%
	DGP2.3 Discontinuity/treatment effect = TRUE Kink/selection effect = FALSE	True Detection - Discontinuity (Detecting treatment effect when there is one)	39.6%	47.6%	52.5%
		False Detection - Kink (Detecting selection effect/kink when there is none)	11.2%	2.6%	1.0%
	DGP2.4 Discontinuity/treatment effect = TRUE Kink/selection effect = TRUE	True Detection - Discontinuity (Detecting treatment effect when there is one)	39.6%	48.7%	55.9%
		True Detection - Kink (Detecting selection effect/kink when there is one)	64.2%	82.0%	89.5%

Green: False-Detection < 1% / True-Detection > 99% | Yellow - False-Detection < 5% / True-Detection > 95% | Red: False-Detection > 5% / True-Detection < 95%

(Indicator saturation method applied at 1% nominal false-detection rate)

Table 7 presents outcomes for simulated false-detection and true-detection rates based on 1,000 replications for each experiment when the DGPs are of quadratic functional form. As in Table 6, three sample sizes and four integer-valued signal-to-noise ratios are considered. We see that false detection of a discontinuity or a kink, when neither exists, is estimated at less than 5% for all cases, as is the case for the linear functional forms of DGPs. However, estimated false-detection rates of a discontinuity when only a kink exists, are very high, for example, 23.6% even when the sample size is 1,000 and signal-to-noise ratio is 5, which is the maximum studied in this experiment. We see that estimated true-detection rates of a kink also are small whenever there is only a kink but no discontinuity (for example 89.9% at the sample size of 1,000 and signal-to-noise ratio of 5). The procedure, therefore, does not seem to perform well in detecting kinks only when there is no discontinuity (pure tipping effects). Nevertheless, estimated false-detection rates in this case seem to be falling in larger sample sizes and larger signal-to-noise ratios, while true-detection rates seem to be rising.

For all cases considered other than the kink-only case, I find estimated false-detection rates below 5% and estimated true-detection rates above 95% for sample sizes of 500 and 1,000 at signal-to-noise ratios of 3 and above.

Overall, in both linear and quadratic functional forms of the DGPs, estimated false-detection rates are small and as expected for large sample sizes for treatment effect estimation (irrespective of whether there is a kink or not), and true-detection rates are also large (above 95%). Thus, the simulation results are satisfactory. The simulated results on false-detection rates and true-detection rates for this procedure when the sample size is large seem slightly more encouraging in the case of the linear functional form of the score variable over the quadratic functional form, considering that the estimation with quadratic functional form performed poorly (large false detection rates, small true-detection rates) when there is only a kink in the DGPs, which is a pure tipping effect with no treatment effect. Also, for the linear functional form, estimated false-detection rates are small and true-detection rates are large even when the sample size is only 200.

As the use of the linear estimator for RDD dominates applications and seems to be becoming more popular, over higher-order polynomial estimations, (the regression function in the local neighborhood is linear or can be approximated using a linear function even though it may be nonlinear from a global point of view), the simulation results for the linear functional forms are particularly promising.

It is noteworthy that, when simulating the experiment 1000 times, it ran into an issue and stopped after completing arbitrary number of iterations (about 100). In closer look, the issue was being caused by the bandwidth selector function in case of kink-only DGPs. The computational procedure in obtaining optimum bandwidth length around a given cutoff makes use of an intermediate matrix which needs to be positive definite. When this requirement is violated, the bandwidth selector function generates an error. Changing the kernel function used in solving for the optimum bandwidth length resolves the issue (the options supported in the 'rdrobust' function are uniform, triangular, and epanechnikov kernels). As mentioned in the methodology section, the choice of the kernel is trivial when 'rdrobust' function is used to generate optimum bandwidth length, see Cattaneo et al. (2019), as the estimation and inference results are not very sensitive and almost invariant. This is because the bandwidth-choice is optimized taking into consideration the particular choice of the kernel. However, to maintain consistency in the method used to generate 1000 replications of the experiment, choice of kernel function was kept the same. Instead, data obtained up to the number of successfully terminated iteration were stored, and the experiment was reinitiated using a seed equals to the serial number of the iteration at which it ran into the error, plus one. The result of 1000 replications, therefore, are stitched together.

It is also worth noting that there are alternate approaches to evaluating properties of a procedure, for example, a Mean Squared Error (MSE) measure of the estimator, that is not included in this thesis. However, distributions of the estimated coefficients are presented graphically which indicates, overall, bias and variance of the estimators are falling in larger sample sizes and larger signal to noise ratios.

Figures 3, 4, 5, 6, 7, 8, 9, and 10 depict the estimated bias-behavior based on 1000 replications in each simulation experiment. More specifically, I present the distributions of the estimated discontinuity and kink coefficients for the 32 combinations of DGPs. From these figures, we can see that the estimated bias diminishes, and the estimated coefficients become more and more centered on the true coefficient value as the sample size increases and as the signal-to-noise ratio increases, which is a desired property. In other words, larger effects that are more discernible from random noise in the data are estimated more accurately, smaller variance in the disturbance term results in a more accurate estimation of coefficients, and larger sample size yields more accurate estimates.

In each of these figure panels, the black vertical line represents the true-coefficient. Red, green, and blue lines represent the distributions of the estimated effect coefficients at sample sizes of 200, 500, and 1000. Signal-to-noise ratio (relative magnitude of discontinuity if it exists as a multiple of the standard deviation of the disturbance term) increases going from top to bottom along the panels. The signal-to-noise ratios of 5,4,3, and 2 corresponds to standard deviation values of 0.2, 0.25, 0.33, and 0.5 for the disturbance terms respectively.

Linear DGPs – (Discontinuity = FALSE, Kink = FALSE)

Sample size: 200 500 1,000

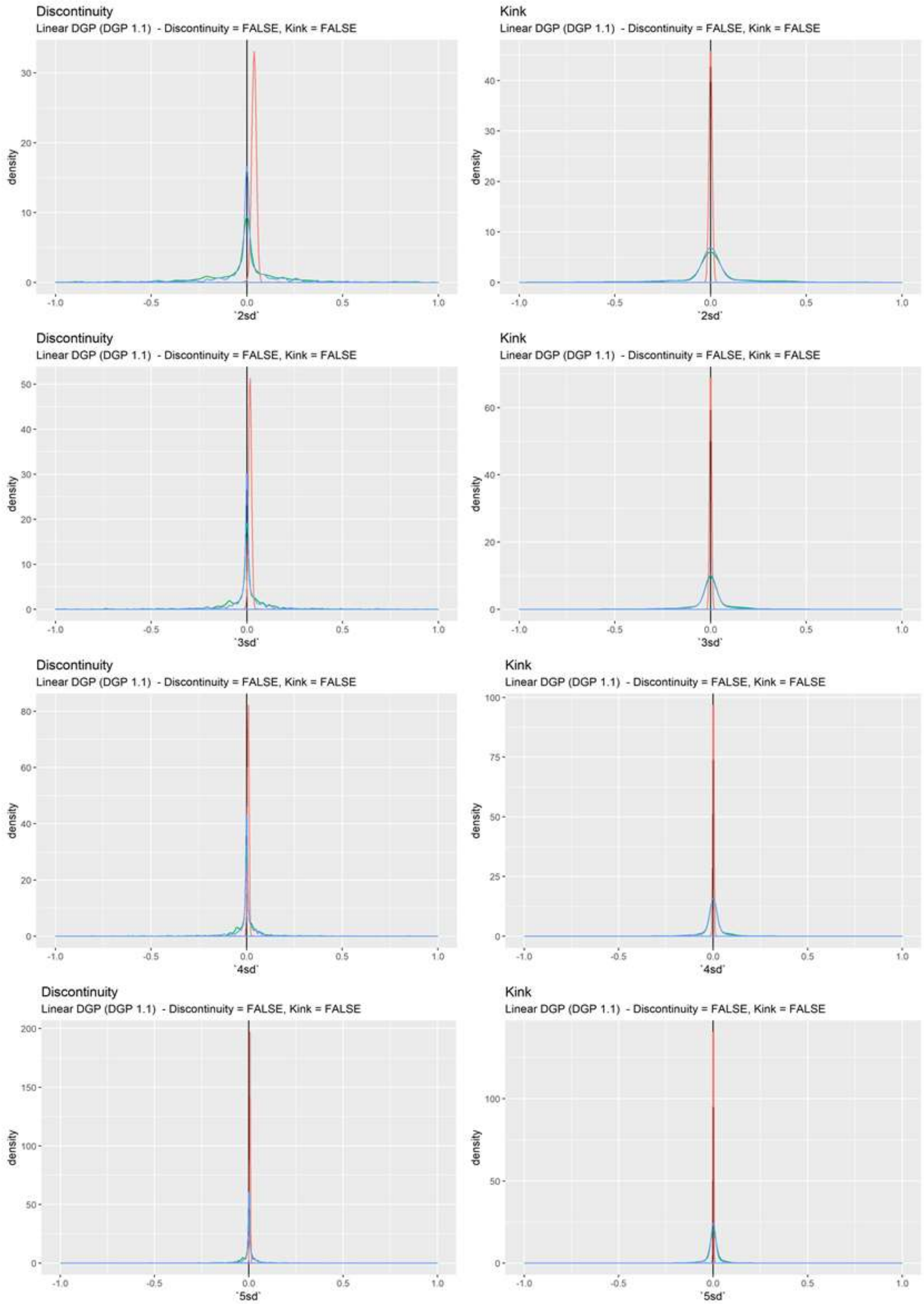


Figure 3: Simulated Bias when neither Discontinuity nor Kink exists – Linear Functional Form DGPs
(Standard Deviation of the disturbance term decreases going from top to bottom)

Quadratic DGPs – (Discontinuity = FALSE, Kink = FALSE)

Sample size: 200 500 1,000

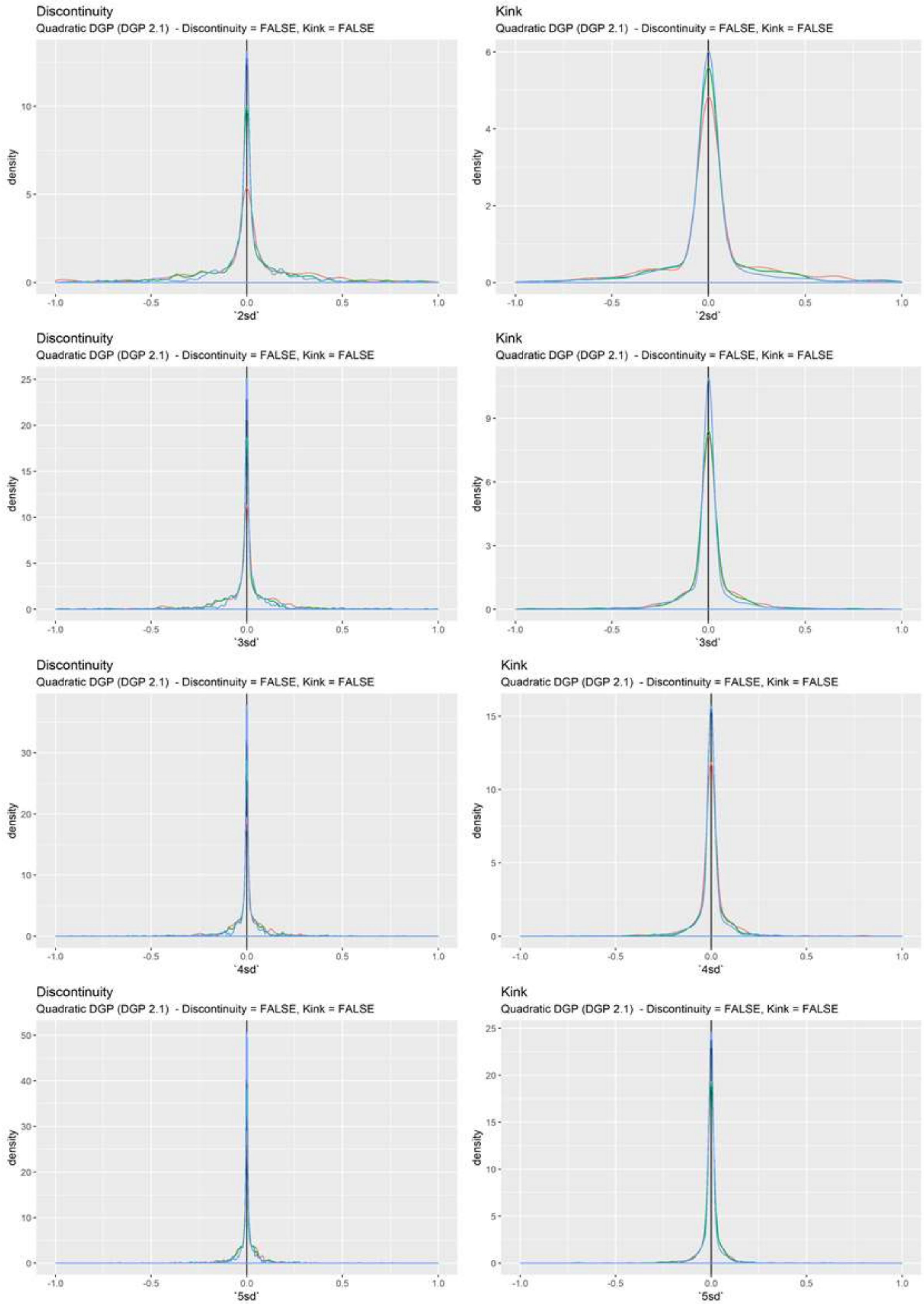


Figure 4: Simulated Bias when neither Discontinuity nor Kink exists – Quadratic Functional Form DGPs
(Standard Deviation of the disturbance term decreases going from top to bottom)

Linear DGPs – (Discontinuity = FALSE, Kink = TRUE)

Sample size: 200 500 1,000

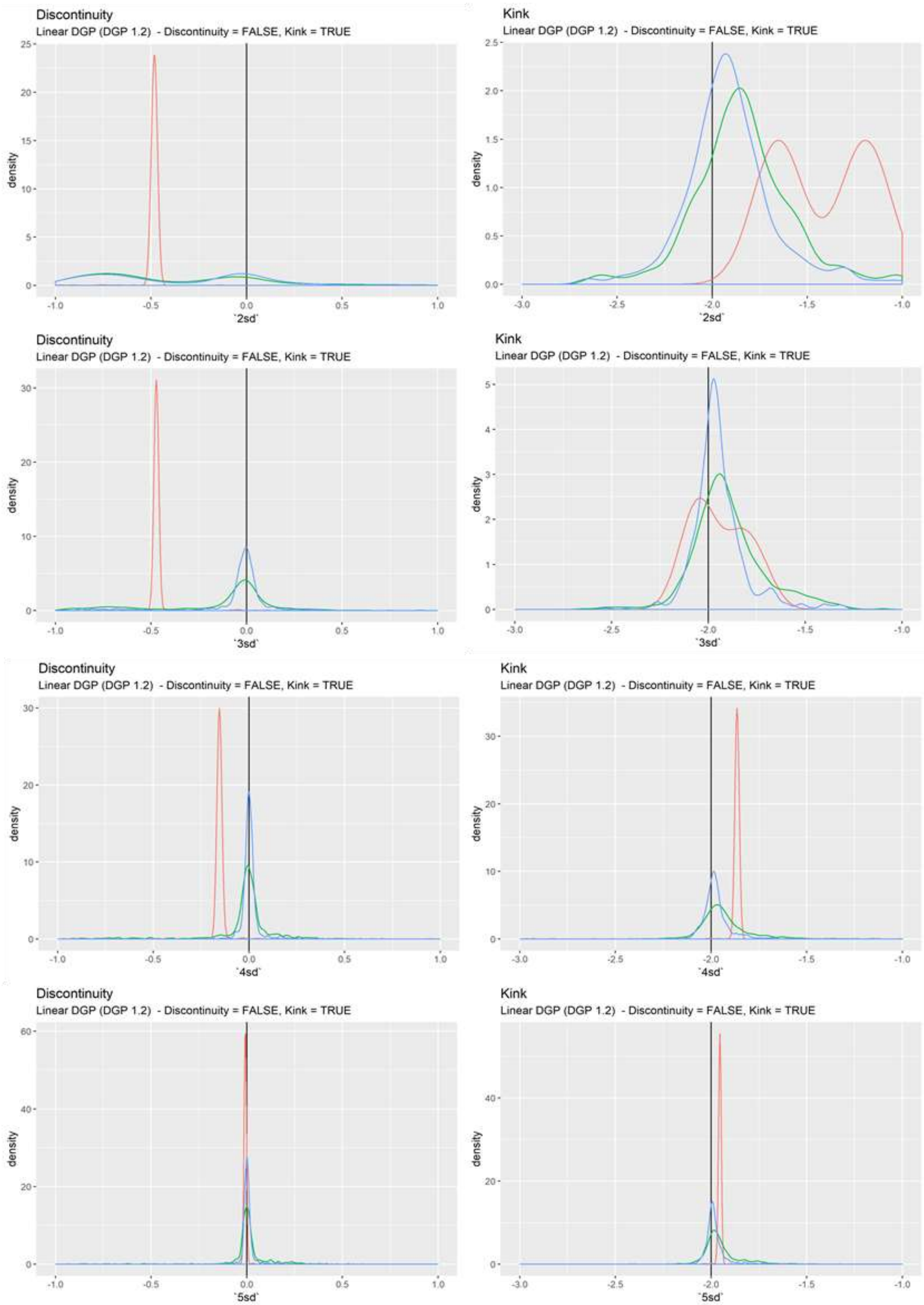


Figure 5: Simulated Bias when only a Kink exists – Linear Functional Form DGPs
(Standard Deviation of the disturbance term decreases going from top to bottom)

Quadratic DGPs – (Discontinuity = FALSE, Kink = TRUE)

Sample size: 200 500 1,000

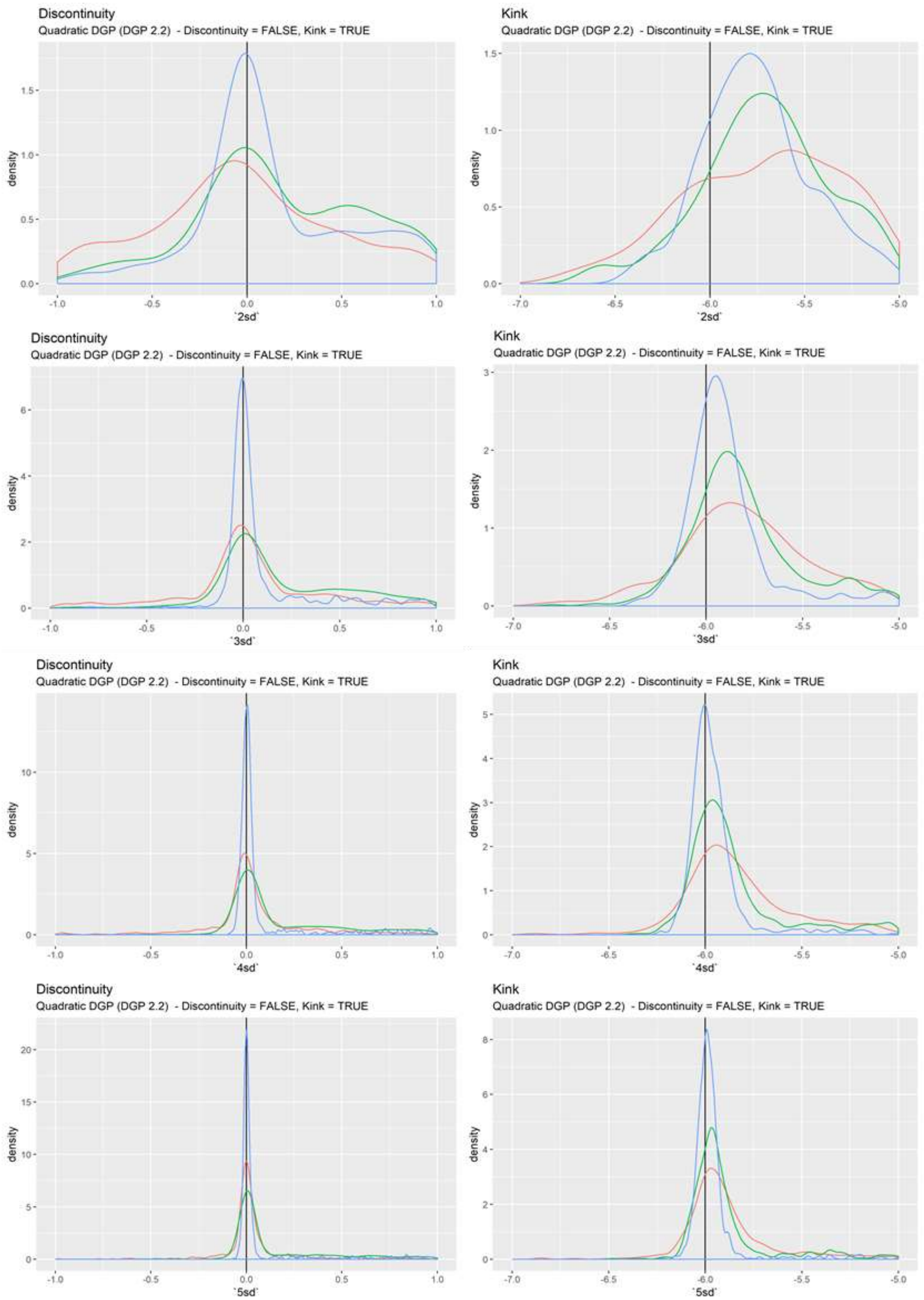


Figure 6: Simulated Bias when only a Kink exists– Quadratic Functional Form DGPs
(Standard Deviation of the disturbance term decreases going from top to bottom)

Linear DGPs – (Discontinuity = TRUE, Kink = FALSE)

Sample size: 200 500 1,000

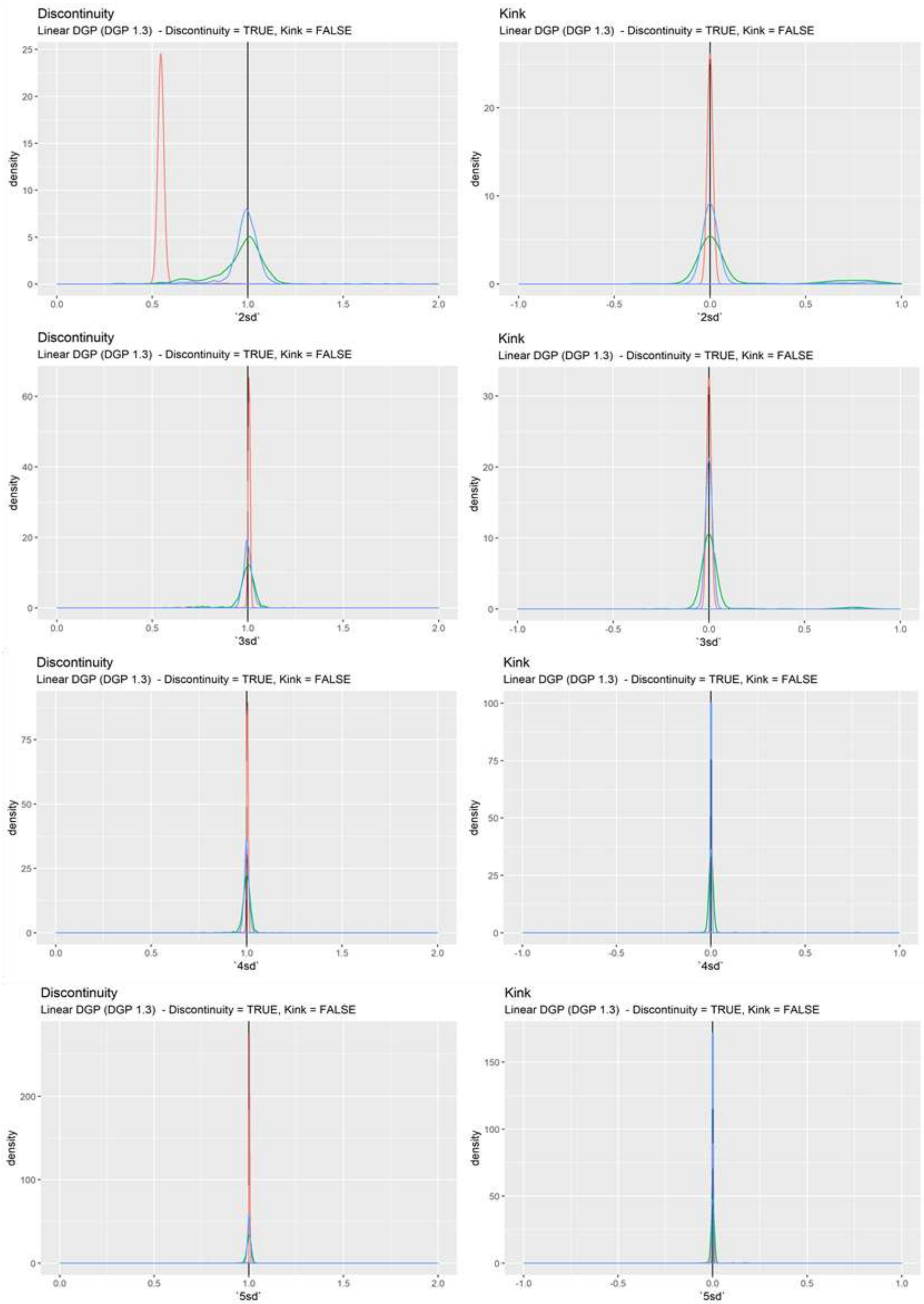


Figure 7: Simulated Bias when only a Discontinuity exists– Linear Functional Form DGPs
(Standard Deviation of the disturbance term decreases going from top to bottom)

Quadratic DGPs – (Discontinuity = TRUE, Kink = FALSE)

Sample size: 200 500 1,000

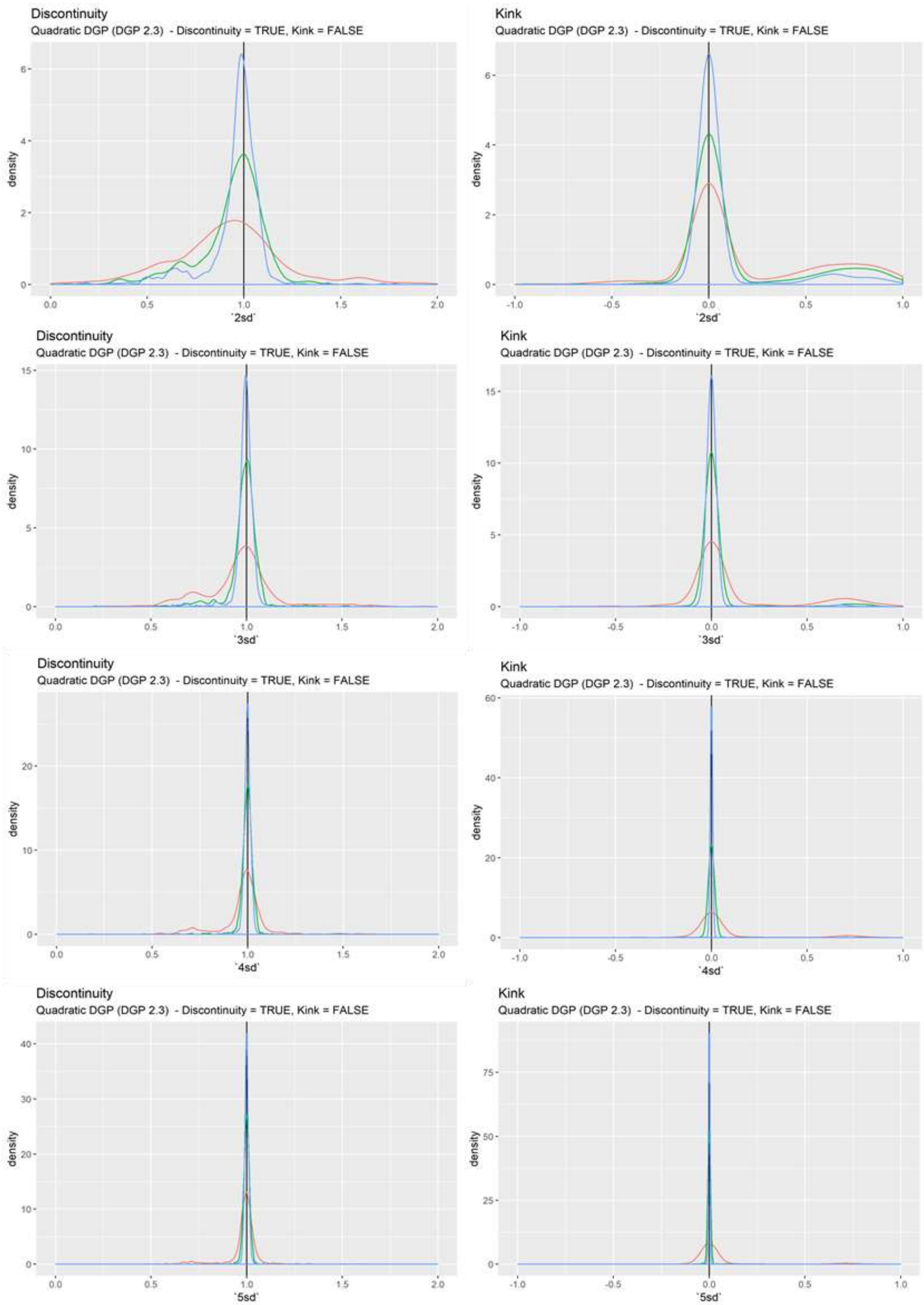


Figure 8: Simulated Bias when only a Discontinuity exists – Quadratic Functional Form DGPs
(Standard Deviation of the disturbance term decreases going from top to bottom)

Linear DGPs – (Discontinuity = TRUE, Kink = TRUE)

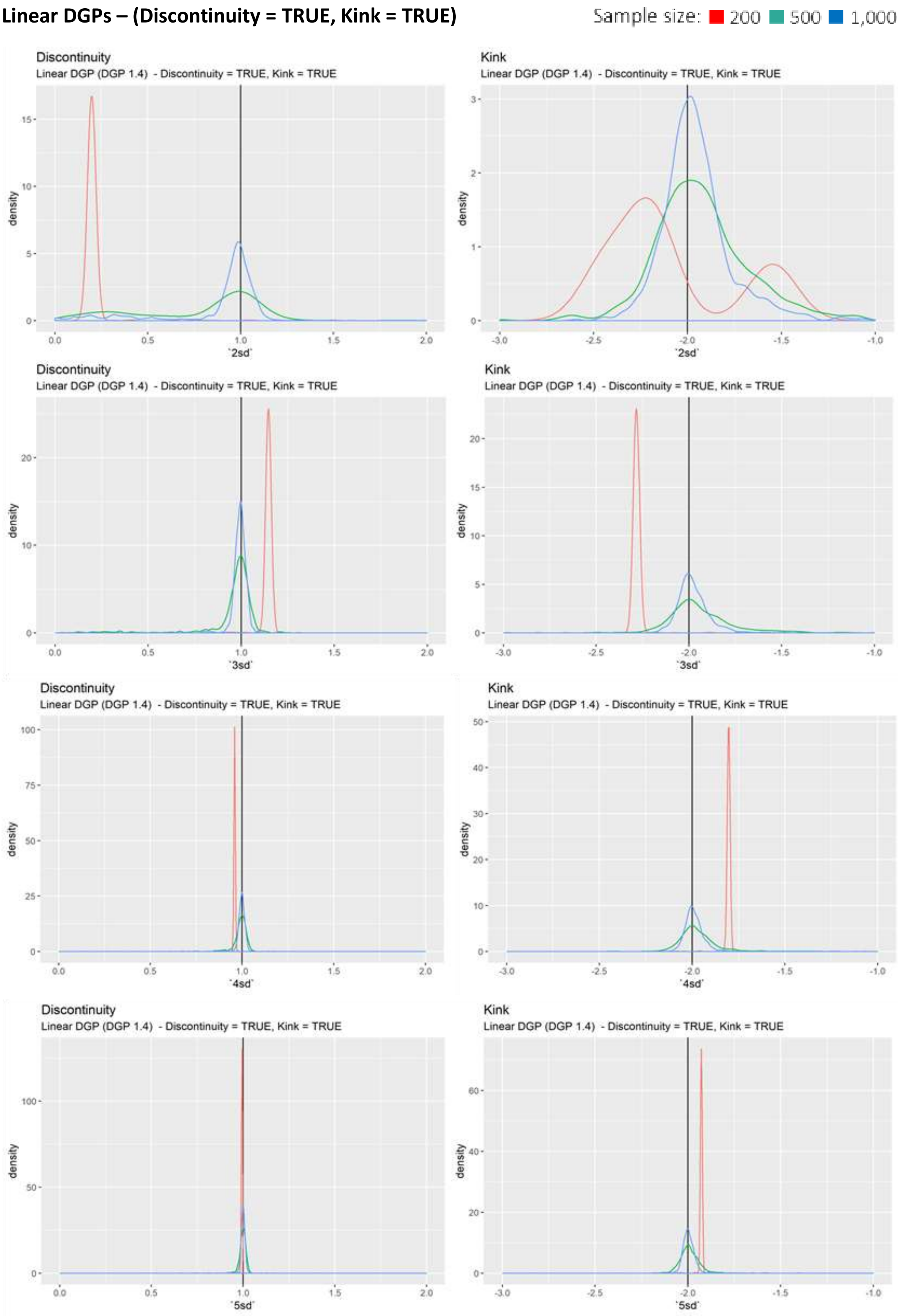


Figure 9: Simulated Bias when both a Discontinuity and a Kink exist – Linear Functional Form DGPs
(Standard Deviation of the disturbance term decreases going from top to bottom)

Quadratic DGPs – (Discontinuity = TRUE, Kink = TRUE)

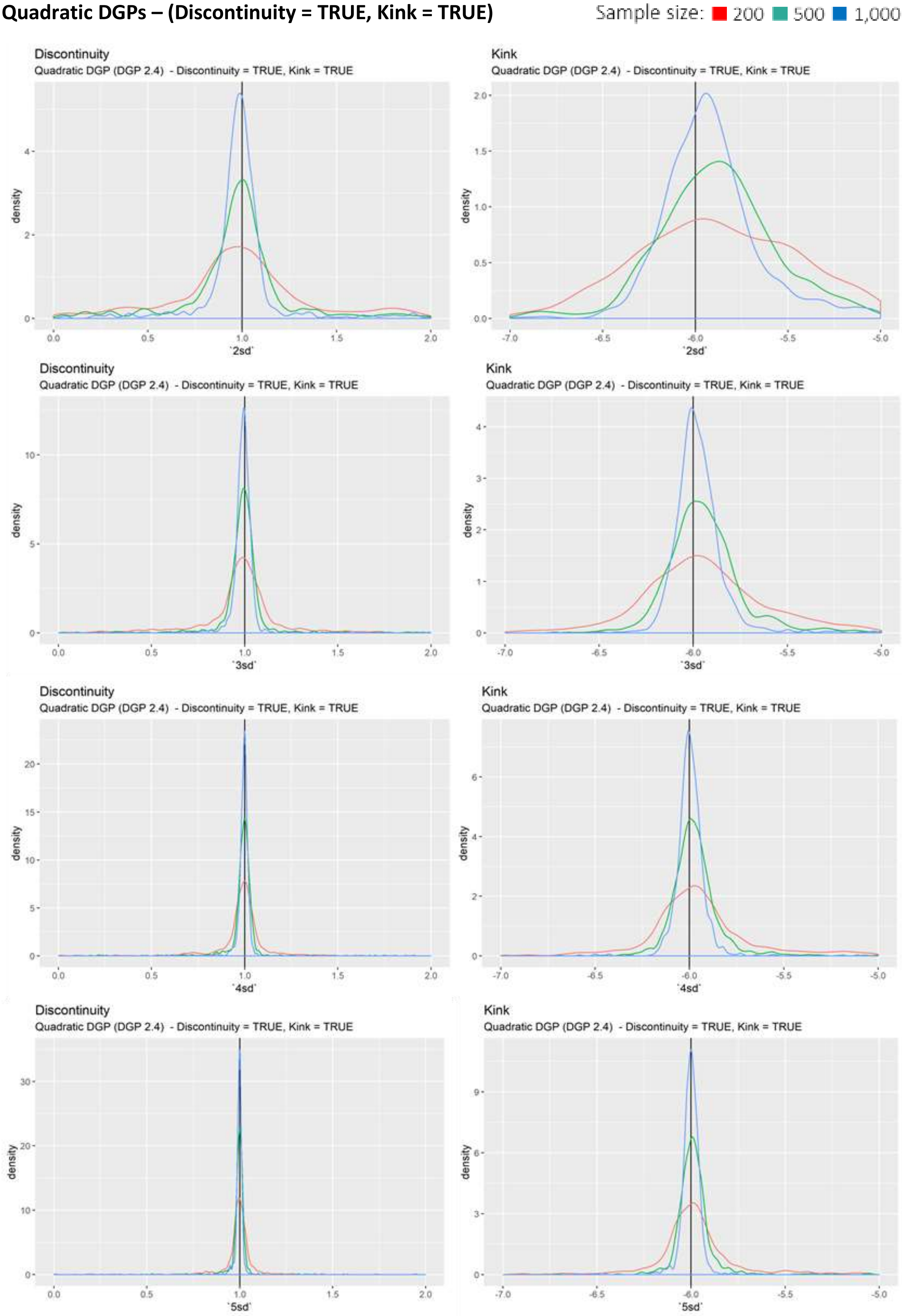


Figure 10: Simulated Bias when both a Discontinuity and a Kink exist – Quadratic Functional Form DGPs

(Standard Deviation of the disturbance term decreases going from top to bottom)

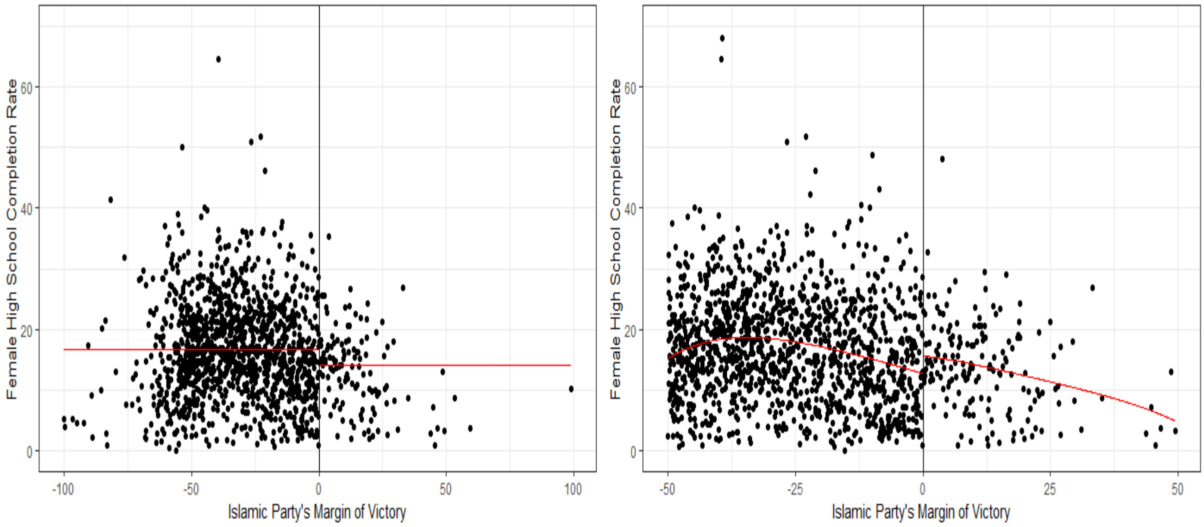
4 Empirical Results

In this section, I apply my proposed approach to regression discontinuity detection with an unknown cutoff on a data set from a published paper. Specifically, I chose to replicate the work of Meyersson (2014). Cattaneo et al. (2019) also use this dataset in their book, which is written as a practitioners' guide, to illustrate the application of RDDs using the `rdrobust' package in R. Overall, my proposed method produces treatment effect estimates that are congruent with those of Meyersson (2014) and Cattaneo et al. (2019), even when the exact cutoff is assumed to be 'not known a-priori.

Meyersson (2014) applied RDD to explore whether Islamic political controls affect women's empowerment and whether a poor woman's rights situation, often exhibited by constituencies with Islamic political control, reflect a causal relationship or a spurious one. More specifically, using RDD, he compared municipalities in Turkey where an Islamic-party affiliated mayor barely won or lost elections. In 1994, Turkey experienced a political shift in local elections when the pro-Islamic Refah Party became the second-largest party in terms of votes, winning 12% of municipalities. Meyersson found that, despite overall negative raw correlations, over a period of six years, municipalities run by Islamic-party affiliated mayors increased female secular high-school education (positive causal effect of about 3% which is a 20% relative increase) and decreased adolescent marriages, with this effect remaining persistent up to 17 years after. Corresponding effects for men were found to be systematically smaller and less precise. He concluded this was consistent with the explanation that the Islamic party in question was effective in overcoming barriers to entry for females among the poor and pious (e.g., refusal of the party to uphold the headscarf bans in educational institutions).

In the context of my work, the vote-margin of the Islamic party, which ranges from -100 to 99.051, is the score or assignment variable based on which the treatment (incumbency of an Islamic-party affiliated Mayor) is assigned. The known cutoff is at '0', as above '0', the Islamic-party wins the majority of votes, thereby ruling in that municipality. The outcome variable for my replication study is the high-school completion rate of women aged 15-20 by the year 2000. The dataset has 2629 municipalities. This is arguably sufficiently large based on the outcomes of my simulation experiments.

Figure 11 presents two scatter plots, the raw comparison of means (left panel) versus local comparison of means in the neighborhood of the cutoff (right panel) in female high school completion rate, with the margin of victory of the Islamic party reported on the X-axis.



Raw Comparison of Means

Local Comparison of Means

Figure 11: Female High School Completion Rate & Islamic Party's Margin of Victory[4]

The scatter plots show that, despite the overall slightly negative slope on either side of the cutoff (left panel), in the neighborhood of the cutoff, the effect of election victory of an Islamic party affiliated mayor on high school completion rate of women seems positive (right panel). In the right panel, an estimated 4th-order global polynomial is used to approximate the population conditional mean functions for control and treated units. Although a global polynomial approach does not deliver good point estimators and inference procedures with good properties for the RD treatment effect estimation, Cattaneo et al. (2019) argue that, for a visual inspection of potential treatment effect, the global approach with higher-order polynomials of the score variable is appropriate.

In the context of this data, it is not conceivable to have an unknown cutoff, as a positive margin of victory results in the party winning that municipality, hence the known cutoff is at '0'. Therefore, at first, I apply only the 2nd and 3rd step of the three-step procedure (bandwidth selection and effect estimation), omitting the cutoff-detection step and taking '0' as the known cutoff instead. This allows us to examine the treatment effect estimation component of the procedure in isolation from the cutoff detection component. Table 8 presents comparisons of the resulting causal effect estimates with those of Meyersson (2014) and those obtained using RD regression along with MSE optimal bandwidth described in Cattaneo et al. (2019).

[4] Plot generated using codes of Cattaneo et al. (2019).

Table 8: Estimated Local Treatment Effect with Known Cutoff – Comparison with Meyersson (2014)

Bandwidth	MSE Optimal (Local)			
Covariates (Control)	no	yes		
Polynomial Order	linear (1)	linear (2)	quadratic (3)	cubic (4)
Est. Treatment Effect (Meyersson, 2014)	0.032*** (0.010)	0.028*** (0.007)	0.028*** (0.011)	0.043*** (0.016)
Est. Treatment Effect (RD Regression with Robust s.e.)[1]	0.032** (0.016)	0.03** (0.015)	0.033* (0.017)	0.031 (0.021)
Est. Treatment Effect (Indicator Saturation - 1% nominal false detection)[2,4]	0.024 †††	0.027 †††	0.019 †††	0.028 †††
Est. Treatment Effect (Indicator Saturation - 5% nominal false detection)[2,5]	0.021 ††	0.033 ††	0.030 ††	0.023 ††
Est. Bandwidth Interval (MSE Optimal) (Score variable - Margin of Victory)	[-15.45, 15.45]	[-12.87, 12.87]	[-19.56, 19.56]	[-21.84, 21.84]
Bandwidth Coverage (% of Data Covered)	27.6%	23.6%	33.3%	36.3%
Effective No. of Observations (Below Cutoff + Above Cutoff)	725 (474 + 251)	619 (390 + 229)	875 (597 + 278)	952 (666 + 286)
Known Cutoff Location - Local (within Bandwidth) (No. of Observation)	475	391	598	667
Detected Cutoff Location - Local (within Bandwidth) (Step-shift nearest to Known Cutoff)[2,3]	369 ††	399 ††	606 ††	675 ††
Total Interval Length (Score variable - Margin of Victory)	[-100, 99.05]	[-100, 99.05]	[-100, 99.05]	[-100, 99.05]
Total Number of Observations - Global (below cutoff + Above Cutoff)	2629 (2314 + 315)	2629 (2314 + 315)	2629 (2314 + 315)	2629 (2314 + 315)
Known Cutoff Location - Global (No. of Observation)	2315	2315	2315	2315
Detected Cutoff Location - Global (Step-shift nearest to Known Cutoff)[2,3]	2209 ††	2323 ††	2323 ††	2323 ††
Known Cutoff Value	0	0	0	0
Detected Cutoff Value (step-shift nearest to known cutoff)	-10.00	0.36	0.36	0.36
Known Cutoff Location - Global	2315	2315	2315	2315

1) *, **, *** represent statisticial significance at 10%, 5%, and 1% levels for RD regression (robust s.e.).
2) †, ††, ††† represent statisticial significance at nominal false-detection rate of 10%, 5%, and 1% levels for Indicator saturation method.
3) For cutoff location detection, Indicator saturation method applied at 1% false-detection rate first, if no effect is found near the known cutoff, false-detection rate relaxed to 5% to increase power. Step indicator (to detect structural shifts) has been used.Cutoff location refers to the no. of observation where cutoff is found after the data has been reordered by ascending values of the Score variable. This creates a pseudo-time index, making the data amenable to time series methods (i.e. indicator saturation, Andrew's test).
4) For effect estimation, At 1% nominal false detection rate, power is relatively small (compared to 5%), Probability of detection of break in the immediate neighbourhood of cutoff diminishes. Hence, within the bandwidth, Coefficients of step indicators above the known cutoff were aggregated and subtracted from aggregate of coefficients of step indicators below the known cutoff to produce an estimate of the local avg. treatment effect.
5) At 5% nominal false detection rate, size is relatively large (compared to 1%). risk of spurious break detection increases. Hence, coefficients of step indicator cluster nearest to the known cutoff were aggregated to produce an estimate of the local avg. treatment effect.

We can see from the outcomes reported in Table 8 that the estimated local average treatment effects using the proposed indicator-saturation method when the cutoff is known are similar, under the different model specifications, to those obtained by Meyersson (2014) and those that can be obtained by applying robust RD regression with computationally derived optimal bandwidth. The effect estimate from using the indicator-saturation approach is particularly close to that obtained by Meyersson (2014) in the case of the linear estimator with covariates as controls, which is the main model explored by Meyersson (2014). This finding suggests that the procedure developed in this thesis may serve as an additional robustness-check to RD regression, even when the cutoff is known.

It is to be noted that, while the indicator-saturation method detects a structural shift near the cutoff at a 5% nominal false-detection rate, it does not detect any such shift at a lower, 1% false-detection rate. This is in congruence with the results obtained from the simulation experiments, where we saw that the indicator-saturation approach, at a 1% false-detection rate, only detects local treatment effects that are sufficiently large and sufficiently indistinguishable from random noise in the data. This highlights that some knowledge of an expected location of the cutoff is necessary for this approach to distinguish a small treatment effect from other structural shifts that may be present in the data.

In applying indicator-saturation method, step-indicators (to detect structural shifts) have been used only for discontinuity detection since plotting the data (see Figure 11) did not reveal any symptom of a kink to warrant testing for a kink. For the detected cutoffs reported in Table 8, I first applied the indicator-saturation method at a 1% nominal false-detection rate, then, if no effect was found near the known cutoff, the false-detection rate was relaxed to 5% to increase the power of the procedure. The 'cutoff location' reported in the table refers to the index (serial number) of the observation where the cutoff is found, after the data has been reordered by ascending values of the score variable. This reordering creates a pseudo-time index, making the data amenable to time series methods.

For effect magnitude estimation, at a 1% nominal false-detection rate for the indicator-saturation algorithm, the true-detection rate is relatively small (compared to at a 5% nominal false-detection rate), the estimated probability of detection of a break in the immediate neighborhood of the cutoff diminishes. Hence, within the bandwidth (restricted data), estimated coefficients of step-indicators above the known cutoff were aggregated, and subtracted from the aggregate of estimated coefficients of step-indicators below the known cutoff, to produce an estimate of the local average treatment effect. At a 5% nominal false-detection rate, however, the risk of a spurious break detection increases. Hence, estimated coefficients of the step-indicator cluster that is nearest to the known cutoff only were aggregated to produce an estimate of the local avg. treatment effect.

As the point estimate of treatment effect are generated in this method through aggregation of multiple step-shifts, this complicates obtaining estimates for associated standard error, even though each detected step-shift coefficient has its own standard error reported by the indicator saturation algorithm. Applying propagation of error method, which is summing squared standard errors of the aggregated step-shifts and taking square root of the sum as standard error estimate, would assign larger standard error estimate to treatment effect point estimates that are broken down into many intermediate step-shifts within the bandwidth over those that are broken down into few, hence would be a biased approach. Therefore, multiple point estimates of treatment effect are generated under different model specifications, such as under additional covariates and different polynomial orders of the score variable used, to see whether they concur or not.

Next, the procedure is applied as if the cutoff were unknown. Detection of an estimated cutoff point and the resulting causal effect estimation is reported under different assumptions on a-priori knowledge on a score interval within which a cutoff may reside (namely, full dataset, 50%, and 25% of the score range centered around the true cutoff). To further illustrate, when the range is the full dataset, the assumption is that there is no a-priori knowledge at all on an interval within which the cutoff does not reside so that no interval-restriction can be applied. This assumption is too restrictive and often unrealistic. To provide a hypothetical example, if students are allocated to remedial classes (a treatment) based on exam-scores which range from 1-100, and if the exact cutoff is unknown, one would reasonably search for a cutoff in the lower score ranges and not on the entire score-range. If the cutoff value is expected to be in the range of 25 to 75, this range is 50% of the total score range.

Figure 12 provides the distribution of the score variable and the distribution within a 50% and 25% score range centered around the true cutoff.

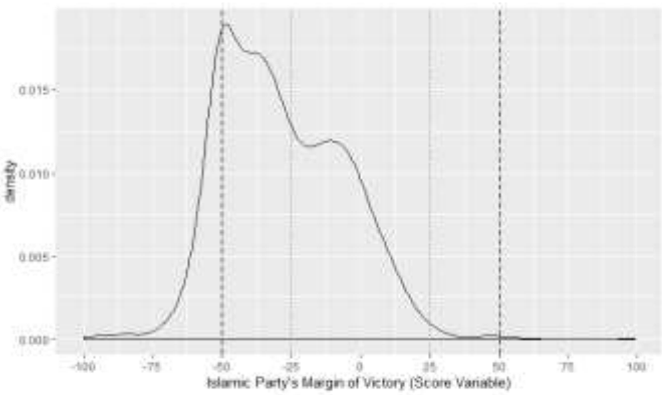


Figure 12: Distribution of the Score Variable

We see from Figure 12 that the distribution is negatively skewed (the party lost more municipalities than it won). Thus, retaining data within a 50% interval centered on the true cutoff only for application of the three-step procedure, in this case, results in about 83% of the data being retained.

In Table 9, results are shown for cutoff detection and effect estimation, as if the cutoff was unknown, under different assumptions on the a-priori knowledge on a score interval within which a cutoff may reside (full dataset, 50%, and 25% of the score range centered around the true cutoff). I compare the results of cutoff detection using both Andrews' Test and indicator-saturation (step 1). I also compare application of conventional RD regression and indicator-saturation for effect estimation (step 2 & 3).

Table 9: Estimated Local Treatment Effect with Unknown Cutoff – Comparison with Meyersson (2014)

A-priori Knowledge on Potential Cutoff (Knowledge of Interval within which the Cutoff Score may reside, centered on True Cutoff)	100% interval length	50% interval length	25% interval length
Covariates (Control)	yes	yes	yes
Polynomial Order	linear	linear	linear
True Cutoff Score	0	0	0
Detected Cutoff Score - Andrew's Test[1]	-8.46***	-0.84***	-0.52***
Detected Cutoff Score - Indicator Saturation at 1% False Detection Rate[2,4]	-39.38 †††	9.85 †††	9.85 †††
Detected Cutoff Score - Indicator Saturation at 5% False Detection Rate[2,4]	-28.76 ††	9.85 ††	0.36 ††
Detected Cutoff Score - Indicator Saturation at 10% False Detection Rate[2,4]	3.82 †	0.36 †	0.36 †
True Cutoff Location (No of Observation)	2314	2314	2314
Detected Cutoff Location - Andrew's Test[1]	2065***	2297***	2301***
Detected Cutoff Location - Indicator Saturation at 1% False Detection Rate[2,4]	948 †††	2505 †††	2505 †††
Detected Cutoff Location - Indicator Saturation at 5% False Detection Rate[2,4]	1420 ††	2505 ††	2325 ††
Detected Cutoff Location - Indicator Saturation at 10% False Detection Rate[2,4]	2400 †	2325 †	2325 †
Retained Observations (% of Observations)	2629 (100%)	2177 (82.81%)	1049 (39.9%)
Est. Treatment Effect (Meyersson, 2014)		**0.028*** (0.007)**	
Est. Treatment Effect - (RD Regression with Robust s.e.)[3] based on known Cutoff		**0.030** (0.015)**	
Est. Treatment Effect - (RD Regression with Robust s.e.)[3] based on estimated Cutoff from Andrew's Test	-0.031** (0.013)	0.018 (0.013)	0.024 (0.02)
Est. Treatment Effect - (RD Regression with Robust s.e.)[3] based on estimated Cutoff from Indicator Saturation at 10% False Detection Rate	0.003 (0.015)	0.033* (0.013)	0.033* (0.013)
Est. Treatment Effect - (Indicator Saturation - 1% nominal false detection)[2,5] based on known Cutoff		**0.027*****	
Est. Treatment Effect - (Indicator Saturation - 1% nominal false detection)[2,5] based on estimated Cutoff from Andrew's Test	0 †††	0.016 †††	0.016 †††
Est. Treatment Effect - (Indicator Saturation - 1% nominal false detection)[2,5] based on estimated Cutoff from Indicator Saturation at 10% False Detection Rate	-0.028 †††	0.012 †††	0.012 †††
Est. Treatment Effect - (Indicator Saturation - 5% nominal false detection)[2,6] based on known Cutoff		**0.033****	
Est. Treatment Effect - (Indicator Saturation - 5% nominal false detection)[2,6] based on estimated Cutoff from Andrew's Test	-0.018 ††	0.02 ††	0.02 ††
Est. Treatment Effect - (Indicator Saturation - 5% nominal false detection)[2,6] based on estimated Cutoff from Indicator Saturation at 10% False Detection Rate	-0.02 ††	0.031 ††	0.031 ††

Green: Est. cutoff score rounds up or down to true cutoff / Est. treatment effect has the same sign as that found in (Meyersson, 2014)
Red: Est. cutoff score does NOT round up or down to true cutoff / Est. treatment effect has different sign than that found in (Meyersson, 2014)

1) *, **, *** represent statisticial significance at 10%, 5%, and 1% levels for Andrew's test.
2) †, ††, ††† represent statisticial significance at nominal false-detection rate of 10%, 5%, and 1% levels for Indicator saturation method.
3) *, **, *** represent statisticial significance at 10%, 5%, and 1% levels for RD regression (robust s.e.).
4) For cutoff location detection, Indicator saturation method applied at 1% false-detection rate first, if no effect is found near the expected cutoff, false-detection rate relaxed to 5% and 10% to increase power. Step indicator (to detect structural shifts) has been used. Cutoff location refers to the no. of observation where cutoff is found after the data has been reordered by ascending values of the Score variable. This creates a pseudo-time index, making the data amenable to time series methods (i.e. indicator saturation, Andrew's test).
5) For effect estimation, At 1% nominal false detection rate, power is relatively small (compared to 5%), Probability of detection of break in the immediate neighbourhood of cutoff diminishes. Hence, within the bandwidth, Coefficients of step indicators above the known cutoff were aggregated and subtracted from aggregate of coefficients of step indicators below the known cutoff to produce an estimate of the local avg. treatment effect.
6) At 5% nominal false detection rate, size is relatively large (compared to 1%). risk of spurious break detection increases. Hence, coefficients of step indicator cluster nearest to the known cutoff were aggregated to produce an estimate of the local avg. treatment effect.

Results using both Andrews' Test and indicator-saturation are compared for the case of unknown cutoff detection. It is to be noted that the required assumptions regarding a-priori knowledge of the cutoffs are necessarily different between these two approaches. When using Andrews' Test for cutoff detection, a-priori knowledge on a suspected cutoff value/location is not required, only a potential score-range is needed. I have chosen to center the score range on the true cutoff in this application (as reported in Table 9). However, this is without loss of generality. As long as the restricted interval contains the true-cutoff, and the true-cutoff is not near the two boundaries of the interval, the detection results are reasonably expected to be invariant to whatever score value it is centered on. On the other hand, when using the indicator-saturation method for unknown cutoff detection, both a-priori knowledge of an anticipated cutoff value/location and a potential score-range is required. This is because, without any knowledge of an expected location/value of the cutoff, a structural shift induced by the treatment is indistinguishable from other structural shifts that may be present in the data away from the cutoff.

The downside to using Andrews' Test for cutoff detection is that it may detect other structural breaks apart from the treatment cutoff, with this risk increasing when there are many outliers and step-shifts, other than the one occurring near the cutoff. However, this risk is reduced with more precise a-priori knowledge on a possible range within which the cutoff may reside. The results provided in Table 9 show, without imposing an assumption of any a-priori knowledge on a potential interval (using 100% of the data), the detected cutoff value is -8.461, which is far off from the true cutoff. Looking at the scatter plots given in Figure 11, there does appear to be possible outliers and potential structural breaks away from the cutoff, which may explain this result. Several municipalities falling to the left of the cutoff have very high female high-school completion rates compared to other municipalities with similar levels of Islamic party's vote shares. However, not surprisingly, the accuracy increases as an assumption is imposed on the knowledge of a potential interval. When I adopt a 50% interval length, the detected cutoff value is -0.838, whereas, with a 25% interval length, it is – 0.52, both of which are reasonably accurate as they round up to the true cutoff.

The advantage of the indicator-saturation based approach to cutoff detection is that it can handle the presence of sizable additional possible structural breaks and outliers, apart from the discontinuity/kink point, as long as an expected cutoff value is also known a-priori with some degree of confidence, in addition to an expected interval within which the cutoff may reside. This assumption is arguably not too restrictive, as applied researchers, in most circumstances, most likely have at least some knowledge of an expected cutoff score, even when the precise cutoff score is not known. This knowledge helps distinguish a structural break near the expected cutoff from other structural breaks.

Results of cutoff detection using indicator-saturation as provided in Table 9, are obtained by first applying the indicator-saturation method with a 1% nominal false-detection rate. Score variable value at the retained step-indicator that is nearest to the expected cutoff is taken as the best approximation of the cutoff. Then, increasing the nominal false-detection rate to 5% and 10%, I see whether the approximated cutoff becomes nearer to the expected cutoff. The nearest of the approximated cutoffs is then taken forward as the estimated cutoff for steps 2 and 3. The cutoff detection was accurate (rounds up to the true cutoff) in this approach at a 5% false-detection rate with a 25% interval length, and at a 10% false-detection rate with a 50% interval length. Detection accuracy is found to be increasing in the nominal false-detection rate, and increasing as the interval-length is reduced within which a cutoff is expected to reside.

Based on the estimated cutoffs from using both the Andrews' Test and the indicator-saturation approach, I report the local treatment effect estimates. We see that these effect estimates are of the same sign as those reported by Meyersson (2014), whenever the cutoff detection was reasonably accurate, either rounding up or down to the true cutoff. The cutoff detection was reasonably accurate whenever some knowledge is assumed of an interval-length of the score variable within which the cutoff may reside. It is to be noted that, 'exact' detection of cutoff using this approach is an impossibility because there is no data point with a value of the score variable that is exactly '0' – which is the true cutoff in this application. The cutoff detection in this application is reasonably accurate in all cases except in the particular case when it is assumed that even knowledge of a score-interval within which the cutoff does not reside is unavailable. Such an assumption is unrealistic and unnecessarily restrictive for most real-world situations where an RDD might be used, as some interval-range of the score variable can be readily ruled out (see page 42). In this application, the cutoff detection was reasonable even at a high 50% interval length, or even when we know a 50% score-range of the total score-range within which the cutoff may reside, which is a wide range.

Overall, compared to the 2.8% positive local average treatment effect estimate reported by Meyersson (2014), my procedure, developed in this thesis, found a 2.7%-3.3% effect when the cutoff was taken as 'known', and a 1.2% – 3.1% effect when the cutoff was taken as 'unknown' and estimated under different parameter choices. The slight loss of precision when the exact cutoff value/location is taken as 'unknown' is expected, as precise knowledge of the cutoff is additional information that is expected to add value in both estimation and causal inference.

5 Conclusion

This thesis introduced a three-step procedure based on structural-break detection and machine learning towards discerning unknown discontinuity points in regression discontinuity designs (RDDs) and regression kink designs (RKDs). I also illustrated using this approach to enable effect estimation. The main idea is that, for sizable treatment effects/tipping effects, the cutoff score, if unknown, might be detectable, as it may coincide with a structural break – once the data has been ordered in ascending values of the score variable, appropriate controls are added, and an indicator-saturation based break detection method is applied, provided some prior knowledge exists as to a potential/expected location/range of the cutoff. Local estimation of the treatment effect, once a cutoff has been estimated, then relies on restricting the data within computationally derived optimal bandwidth lengths on both sides of the cutoff and applying the indicator-saturation method again within the bandwidth only. Retained step-indicator coefficients on either side of the estimated cutoff are then aggregated. Subtracting the aggregate of step-indicator coefficients below the estimated cutoff from the aggregate of step-indicator coefficients above the estimated cutoff yields the effect estimates. For tipping effect detection, instead of a step-indicator, I propose a user-defined indicator that can capture a step-shift in the slope of the regression function, with the other steps remaining the same.

Large and small sample properties (false-detection rates and true-detection rates) are simulated suggesting that the approach may work well. I also illustrated my proposed method using an empirical application. Nicely for researchers, my proposed procedure can serve as an additional robustness check, even when the cutoff is known. In such a case, the treatment effect (or tipping effect) being picked up, without imposing its location based on a-priori knowledge, can indicate that the effect is sizable enough to be detected as a structural break in parameter stability, and is therefore of practical significance. This can further bolster evidence in favor of a treatment effect (or tipping effect), as the effect is sufficiently large to be detected using machine learning methods without imposing it on the data based on a-priori knowledge.

An advantage of my proposed three-step procedure is that it can enable testing for either an unknown selection effect/tipping effect, which is a discontinuity in the slope of the regression function, or an unknown treatment effect, which is a discontinuity in the levels of the regression function, independently as well as simultaneously. Also, both the cutoff detection step and the effect estimation step offer convenient alternatives. The cutoff detection step can make use of either the indicator-saturation method or Andrews' Test for parameter instability. Taken together I showed that the approach robustly handles the possible presence of outliers and other step-shifts away from the true cutoff, as well as different ranges of certainty as to the location of an expected cutoff. The estimation step of my procedure can conveniently accommodate different nominal false-detection rates when using the indicator-saturation approach. Finally, this step can also make use of a conventional RD regression method (e.g., the `rdrobust' package in R) taking the estimated cutoff as an input, which serves as an additional layer of robustness checks. However, when undertaking this, it is important to keep in mind that the simulation results suggested a downward bias in the estimated effect coefficients when estimated cutoffs are used in conventional RD regression instead of the exact cutoffs. Together, these flexible options allow for a wide range of robustness checks for RDD and RKD with unknown cutoffs, thus extending the applicability of regression discontinuity designs.

There are some interesting areas unexplored in this thesis. First, due to the long computation time required which is increasing in sample size and number of replications, Monte Carlo simulation was restricted to 1000 replications, and a maximum sample size of 1,000 was studied. A sample size of 1,000 was found to be sufficient in this simulation to produce estimated false-detection rates that are close to the nominal false-detection rate chosen, and true-detection rates that are high, except in the particular case when quadratic functional form of DGPs was studied that had a kink but no discontinuity; perhaps a larger sample size will produce false-detection rates close to the nominal false-detection rate chosen and high true-detection rates in this case as well. This can be studied in future research. Second, linear and quadratic functional forms of DGPs were studied that were 'well-behaved' and simple under five integer-valued signal-to-noise ratios. Future research can study false-detection and true-detection properties under additional simulated adverse conditions such as unobserved variables, model misspecification, and substantial noise in the form of presence of many outliers and other structural breaks not occurring at the point of the cutoff. Also, Monte-Carlo simulation in this thesis was confined to studying the linear and the quadratic functional form of DGPs as they represent the majority of RDD applications in the literature. However, higher-order polynomial functional forms are also used, albeit rarely. Besides, in many applications of RDDs where the linear and quadratic functional forms are applied as the main model, higher-order polynomial functional forms are also applied as robustness-checks. Small and large sample properties of this method when higher-order polynomials of the score variable are used therefore, is another area of future research. Finally, while the method produced satisfactory results when applied to the particular empirical data that is examined in this thesis, additional exploration with more real-world datasets is necessary, which can be taken up in future research.

References

Andrews, D. W. K. (1993). Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica*, *61*(4), 821–856.

Angrist, J. D., & Lavy, V. (1999). Using Maimonides ' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *The Quarterly Journal of Economics*, *114*(2), 533–575.

Bierens, H. J. (1982). Consistent Model Specification Tests. *Journal of Econometrics*, *20*(1), 105–134.

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2020). Optimal Bandwidth Choice for Robust Bias-corrected Inference in Regression Discontinuity Designs. *The Econometrics Journal*, *23*(2), 192–210.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2019). Regression Discontinuity Designs Using Covariates. *Review of Economics and Statistics*, *101*(3), 442–451.

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2018). On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. *Journal of the American Statistical Association*, *113*(522), 767–779.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2015). rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs. *R Journal*, *7*(1), 38–51.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, *82*(6), 2295–2326.

Card, D., Mas, A., Moretti, E., & Saez, E. (2012). Inequality at Work: The Effect of Peer Salaries on Job Satisfaction. *American Economic Review*, *102*(6), 2981–3003.

Card, D., Mas, A., & Rothstein, J. (2008). Tipping and the Dynamics of Segregation, *The Quarterly Journal of Economics*, *123*(1), 177–218.

Castle, J. L., Doornik, J. a, Hendry, D. F., & Pretis, F. (2015). Detecting Location Shifts by Step-indicator-saturation. *Econometrics*, *3*(2), 240–264.

Cattaneo, M.D., Idrobo, N., & Titiunik, R. (2019). A Practical Introduction to Regression Discontinuity Designs: Foundations (*Elements in Quantitative and Computational Methods for the Social Sciences*). Cambridge: Cambridge University Press.

Cattaneo, M. D., & Escanciano, J. C. (2017). Regression Discontinuity Designs: Theory and Applications (*Advances in Econometrics, volume 38*). Emerald Group Publishing.

Dell, M., & Querubin, P. (2018). Nation Building Through Foreign Intervention: Evidence from Discontinuities in Military Strategies. *The Quarterly Journal of Economics*, *133*(2), 701–764.

Ganong, P., & Jäger, S. (2018). A Permutation Test for the Regression Kink Design. *Journal of the American Statistical Association*, *113*(522), 494–504.

Gelman, A., & Imbens, G. (2018). Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. *Journal of Business and Economic Statistics*, *37*(3), 447–456.

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

Hansen, B. E. (2017). Regression Kink with an Unknown Threshold. *Journal of Business and Economic Statistics*, *35*(2), 228–240.

Hausman, C., & Rapson, D. S. (2018). Regression Discontinuity in Time: Considerations for Empirical Applications. *Annual Review of Resource Economics*, *10*(1), 533–552.

Hendry, D. F., Johansen, S., & Santos, C. (2008). Automatic Selection of Indicators in a Fully Saturated Regression*. Computational Statistics 23,* 317–335*.*

Herlands, W., Wilson, A. G., McFowland, E., & Neill, D. B. (2018). Automated Local Regression Discontinuity Design Discovery. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1512–1520.

Imbens, G., & Kalyanaraman, K. (2012). Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Review of Economic Studies*, *79*(3), 933–959.

Imbens, G. W., & Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, *142*(2), 615–635.

Johansen, S., & Nielsen, B. (2016). Asymptotic Theory of Outlier Detection Algorithms for Linear Time Series Regression Models. *Scandinavian Journal of Statistics*, *43*(2), 321–348.

Landais, C. (2015). Assessing the Welfare Effects of Unemployment Benefits Using the Regression Kink Design. *American Economic Journal: Economic Policy*, *7*(4), 243–278.

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

Meyersson, E. (2014), Islamic Rule and the Empowerment of the Poor and Pious. *Econometrica*, *82*(1), 229-269.

Pei, Z., Card, D., Lee, D. S., & Weber, A. (2018). Local Polynomial Order in Regression Discontinuity Designs. *Working Paper No. 622*, Princeton University, Department of Economics.

Porter, J., & Yu, P. (2015). Regression Discontinuity Designs with Unknown Discontinuity Points: Testing and Estimation. *Journal of Econometrics*, *189*(1), 132–147.

Pretis, F., Reade, J. J., & Sucarrat, G. (2018). Automated General-to-specific (GETS) Regression Modeling and Indicator-saturation for Outliers and Structural Breaks. *Journal of Statistical Software*, *86*(3).

Pretis F, Schneider L, Smerdon JE, Hendry DF (2016). Detecting Volcanic Eruptions in Temperature Reconstructions by Designed Break-Indicator-saturation. *Journal of Economic Surveys*, *30*(3), 403–429.

Schelling, T.C. (1971). Dynamic Models of Segregation, *The Journal of Mathematical Sociology*, *1*(2), 143-186.

Schneider L, Smerdon J, Pretis F, Hartl-Meier C, Esper J (2017). A New Archive of Large Volcanic Events over the Past Millennium Derived from Reconstructed Summer Temperatures. *Environmental Research Letters*, *12*(9).

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-Discontinuity Analysis: An Alternative to the Ex post facto Experiment. *Journal of Educational Psychology, 51*(6), 309–317.

van der Klaauw, W. (2008). Regression-Discontinuity Analysis: A Survey of Recent Developments in Economics. *Labour*, *22*(2), 219–245.

van der Klaauw, W. (2002). Estimating the Effect of Financial Aid Offers on College Enrollment. *International Journal of Economic Review*, *43*(4), 1249–1287.

Zeileis, A., Leisch, F., Homik, K., & Kleiber, C. (2002). strucchange: An R Package for Testing for Structural Change. *Journal of Statistical Software*, *7*(2), 1–38.

Appendix A

Extended Table 6 - Simulated False-Detection Rates & True-Detection Rates for Linear DGPs

(Results when signal detection Interval of 2.58 standard deviation of disturbance term used are included)

Magnitude of Discontinuity (if exists)	DGPs (Linear)	Parameter	signal detection interval: 1.96 s.d. of disturbance			signal detection interval 2.58 s.d. of disturbance		
			n = 200	n = 500	n = 1000	n = 200	n = 500	n = 1000
5 s.d	DGP1.1 Discontinuity/treatment effect = FALSE Kink/selection effect = FALSE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	0.0%	0.9%	0.7%	0.0%	0.7%	0.4%
		False Detection - Kink (Detecting selection effect/kink when there is none)	0.0%	0.8%	0.5%	0.0%	0.6%	0.2%
	DGP1.2 Discontinuity/treatment effect = FALSE Kink/selection effect = TRUE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	0.2%	2.2%	1.2%	0.1%	1.7%	0.6%
		True Detection - Kink (Detecting selection effect/kink when there is one)	99.9%	98.9%	99.9%	99.9%	98.3%	99.9%
	DGP1.3 Discontinuity/treatment effect = TRUE Kink/selection effect = FALSE	True Detection - Discontinuity (Detecting treatment effect when there is one)	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
		False Detection - Kink (Detecting selection effect/kink when there is none)	0.0%	0.1%	0.0%	0.0%	0.1%	0.0%
	DGP1.4 Discontinuity/treatment effect = TRUE Kink/selection effect = TRUE	True Detection - Discontinuity (Detecting treatment effect when there is one)	99.9%	99.8%	100.0%	99.9%	99.8%	100.0%
		True Detection - Kink (Detecting selection effect/kink when there is one)	99.9%	99.9%	100.0%	99.9%	99.8%	100.0%
4 s.d	DGP1.1 Discontinuity/treatment effect = FALSE Kink/selection effect = FALSE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	0.0%	1.1%	1.4%	0.0%	0.4%	0.7%
		False Detection - Kink (Detecting selection effect/kink when there is none)	0.0%	0.8%	0.9%	0.0%	0.5%	0.5%
	DGP1.2 Discontinuity/treatment effect = FALSE Kink/selection effect = TRUE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	0.5%	4.4%	0.9%	0.5%	2.5%	0.4%
		True Detection - Kink (Detecting selection effect/kink when there is one)	99.6%	97.2%	99.9%	99.6%	95.0%	99.6%
	DGP1.3 Discontinuity/treatment effect = TRUE Kink/selection effect = FALSE	True Detection - Discontinuity (Detecting treatment effect when there is one)	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
		False Detection - Kink (Detecting selection effect/kink when there is none)	0.3%	0.2%	0.0%	0.3%	0.2%	0.0%
	DGP1.4 Discontinuity/treatment effect = TRUE Kink/selection effect = TRUE	True Detection - Discontinuity (Detecting treatment effect when there is one)	99.9%	99.7%	99.9%	99.9%	99.7%	99.9%
		True Detection - Kink (Detecting selection effect/kink when there is one)	99.9%	99.7%	99.9%	99.8%	99.6%	99.6%
3 s.d	DGP1.1 Discontinuity/treatment effect = FALSE Kink/selection effect = FALSE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	0.1%	2.4%	2.0%	0.0%	1.2%	1.4%
		False Detection - Kink (Detecting selection effect/kink when there is none)	0.0%	1.2%	1.5%	0.0%	0.9%	0.9%
	DGP1.2 Discontinuity/treatment effect = FALSE Kink/selection effect = TRUE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	0.9%	12.8%	4.6%	0.4%	4.2%	1.0%
		True Detection - Kink (Detecting selection effect/kink when there is one)	98.8%	82.6%	94.4%	0.9%	78.0%	92.4%
	DGP1.3 Discontinuity/treatment effect = TRUE Kink/selection effect = FALSE	True Detection - Discontinuity (Detecting treatment effect when there is one)	99.9%	99.3%	100.0%	99.3%	94.8%	98.7%
		False Detection - Kink (Detecting selection effect/kink when there is none)	0.5%	3.8%	1.1%	0.2%	0.2%	0.0%
	DGP1.4 Discontinuity/treatment effect = TRUE Kink/selection effect = TRUE	True Detection - Discontinuity (Detecting treatment effect when there is one)	99.2%	96.1%	99.1%	99.2%	95.8%	99.1%
		True Detection - Kink (Detecting selection effect/kink when there is one)	99.2%	95.5%	98.8%	99.1%	90.9%	97.3%
2 s.d	DGP1.1 Discontinuity/treatment effect = FALSE Kink/selection effect = FALSE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	0.1%	4.5%	4.1%	0.1%	2.9%	2.6%
		False Detection - Kink (Detecting selection effect/kink when there is none)	0.1%	2.9%	2.8%	0.0%	1.7%	1.6%
	DGP1.2 Discontinuity/treatment effect = FALSE Kink/selection effect = TRUE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	0.9%	13.4%	9.5%	0.2%	2.8%	2.4%
		True Detection - Kink (Detecting selection effect/kink when there is one)	0.2%	24.5%	36.7%	0.1%	23.7%	36.1%
	DGP1.3 Discontinuity/treatment effect = TRUE Kink/selection effect = FALSE	True Detection - Discontinuity (Detecting treatment effect when there is one)	0.5%	54.0%	60.7%	0.2%	0.7%	0.1%
		False Detection - Kink (Detecting selection effect/kink when there is none)	0.2%	0.4%	0.0%	0.2%	0.2%	0.0%
	DGP1.4 Discontinuity/treatment effect = TRUE Kink/selection effect = TRUE	True Detection - Discontinuity (Detecting treatment effect when there is one)	0.8%	63.1%	82.0%	0.5%	61.7%	81.4%
		True Detection - Kink (Detecting selection effect/kink when there is one)	0.5%	34.5%	46.5%	0.4%	0.7%	0.8%

Green: False-Detection < 1% / True-Detection > 99% | Yellow - False-Detection < 5% / True-Detection > 95% | Red: False-Detection > 5% / True-Detection < 95%

(Indicator saturation method applied at 1% nominal false-detection rate)

Extended Table 7 - Simulated False-Detection Rates & True-Detection Rates for Quadratic DGPs

(Results when signal detection Interval of 2.58 standard deviation of disturbance term used are included)

Magnitude of Discontinuity (if exists)	DGPs (Quadratic)	Parameter	signal detection interval: 1.96 s.d. of disturbance			signal detection interval 2.58 s.d. of disturbance		
			n = 200	n = 500	n = 1000	n = 200	n = 500	n = 1000
5 s.d	DGP2.1 Discontinuity/treatment effect = FALSE Kink/selection effect = FALSE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	0.9%	0.7%	0.3%	0.6%	0.6%	0.3%
		False Detection - Kink (Detecting selection effect/kink when there is none)	0.8%	0.7%	0.3%	0.5%	0.5%	0.3%
	DGP2.2 Discontinuity/treatment effect = FALSE Kink/selection effect = TRUE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	15.6%	34.5%	23.6%	13.9%	31.0%	22.6%
		True Detection - Kink (Detecting selection effect/kink when there is one)	92.5%	84.7%	89.9%	92.4%	84.3%	89.7%
	DGP2.3 Discontinuity/treatment effect = TRUE Kink/selection effect = FALSE	True Detection - Discontinuity (Detecting treatment effect when there is one)	100.0%	100.0%	100.0%	99.9%	100.0%	100.0%
		False Detection - Kink (Detecting selection effect/kink when there is none)	7.5%	0.1%	0.0%	7.0%	0.1%	0.0%
	DGP2.4 Discontinuity/treatment effect = TRUE Kink/selection effect = TRUE	True Detection - Discontinuity (Detecting treatment effect when there is one)	94.9%	98.9%	100.0%	94.8%	98.6%	100.0%
		True Detection - Kink (Detecting selection effect/kink when there is one)	97.2%	99.8%	100.0%	97.2%	99.8%	100.0%
4 s.d	DGP2.1 Discontinuity/treatment effect = FALSE Kink/selection effect = FALSE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	1.6%	0.5%	0.9%	1.1%	0.1%	0.5%
		False Detection - Kink (Detecting selection effect/kink when there is none)	1.4%	0.2%	0.7%	1.0%	0.1%	0.4%
	DGP2.2 Discontinuity/treatment effect = FALSE Kink/selection effect = TRUE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	19.1%	31.8%	28.2%	15.2%	27.7%	26.3%
		True Detection - Kink (Detecting selection effect/kink when there is one)	90.6%	86.3%	86.1%	89.7%	85.4%	86.0%
	DGP2.3 Discontinuity/treatment effect = TRUE Kink/selection effect = FALSE	True Detection - Discontinuity (Detecting treatment effect when there is one)	99.8%	100.0%	100.0%	98.0%	99.9%	100.0%
		False Detection - Kink (Detecting selection effect/kink when there is none)	12.8%	1.0%	0.2%	11.7%	0.8%	0.1%
	DGP2.4 Discontinuity/treatment effect = TRUE Kink/selection effect = TRUE	True Detection - Discontinuity (Detecting treatment effect when there is one)	94.3%	98.4%	99.9%	93.6%	98.2%	99.5%
		True Detection - Kink (Detecting selection effect/kink when there is one)	97.1%	99.7%	99.8%	96.5%	99.6%	99.8%
3 s.d	DGP2.1 Discontinuity/treatment effect = FALSE Kink/selection effect = FALSE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	2.0%	2.4%	2.3%	1.3%	1.1%	0.9%
		False Detection - Kink (Detecting selection effect/kink when there is none)	1.7%	1.6%	0.9%	1.1%	0.7%	0.7%
	DGP2.2 Discontinuity/treatment effect = FALSE Kink/selection effect = TRUE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	27.7%	35.0%	30.1%	22.3%	28.2%	26.7%
		True Detection - Kink (Detecting selection effect/kink when there is one)	82.4%	78.3%	80.8%	81.4%	78.0%	80.1%
	DGP2.3 Discontinuity/treatment effect = TRUE Kink/selection effect = FALSE	True Detection - Discontinuity (Detecting treatment effect when there is one)	94.2%	99.3%	99.5%	78.8%	94.0%	97.1%
		False Detection - Kink (Detecting selection effect/kink when there is none)	18.8%	4.6%	1.7%	8.6%	1.1%	0.4%
	DGP2.4 Discontinuity/treatment effect = TRUE Kink/selection effect = TRUE	True Detection - Discontinuity (Detecting treatment effect when there is one)	85.2%	96.9%	99.4%	77.2%	94.4%	97.7%
		True Detection - Kink (Detecting selection effect/kink when there is one)	94.4%	98.8%	99.1%	92.1%	98.8%	99.0%
2 s.d	DGP2.1 Discontinuity/treatment effect = FALSE Kink/selection effect = FALSE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	4.6%	3.9%	4.2%	2.1%	2.5%	2.8%
		False Detection - Kink (Detecting selection effect/kink when there is none)	2.8%	2.5%	2.3%	1.8%	1.8%	1.3%
	DGP2.2 Discontinuity/treatment effect = FALSE Kink/selection effect = TRUE	False Detection - Discontinuity (Detecting treatment effect/discontinuity when there is none)	56.3%	44.4%	43.8%	45.8%	34.2%	36.1%
		True Detection - Kink (Detecting selection effect/kink when there is one)	40.2%	41.9%	49.4%	34.3%	39.6%	48.5%
	DGP2.3 Discontinuity/treatment effect = TRUE Kink/selection effect = FALSE	True Detection - Discontinuity (Detecting treatment effect when there is one)	39.6%	47.6%	52.5%	8.8%	1.1%	0.6%
		False Detection - Kink (Detecting selection effect/kink when there is none)	11.2%	2.6%	1.0%	5.7%	0.7%	0.6%
	DGP2.4 Discontinuity/treatment effect = TRUE Kink/selection effect = TRUE	True Detection - Discontinuity (Detecting treatment effect when there is one)	39.6%	48.7%	55.9%	17.0%	12.2%	11.1%
		True Detection - Kink (Detecting selection effect/kink when there is one)	64.2%	82.0%	89.5%	59.4%	79.4%	89.2%

Green: False-Detection < 1% / True-Detection > 99% | **Yellow - False-Detection < 5% / True-Detection > 95%** | **Red: False-Detection > 5% / True-Detection < 95%**

(Indicator saturation method applied at 1% nominal false-detection rate)

Appendix B

R Codes – Monte Carlo Simulation

Following code generates the 1,000 replications of the simulation for the linear functional forms of DGPs:

```
#Select a Sample Size
# n <- 200
# n <- 500
n <- 1000

#Create empty table for storing data
disc.tbl <- array(0, dim = c(1000, 4, 4))
kink.tbl <- array(0, dim = c(1000, 4, 4))

#Define function
isatfun <- function(i){

  #generate the DGPs
  set.seed(k)
  e   <- rnorm(n,0,(1/i)^2)
  x   <- runif(n,-2,3)
  xsq <- x^2
  sl  <- seq(1,n,1)
  y1 <- x + e
  y2 <- ifelse((x<1 & x>=-2),(x + e),(-x + 2  + e))
  y3 <- ifelse((x<1 & x>=-2),(x + e),(x + 1 + e))
  y4 <- ifelse((x<1 & x>=-2),(x + e),(-x + 3 + e))
  dgp <- cbind(x, xsq, y1, y2, y3, y4)

  #create ordering in the data for application of time-series based methods
  dgp <- dgp[order(dgp[,1]),]
  dgp <- cbind(sl,dgp)
  dgp <- as.data.frame(dgp)

  #discontinuity detection using Andrews' test
  a1 <- Fstats(dgp$y1 ~ 1 + dgp$x, from = 0.15, to = 0.85)
  d1 <- a1$breakpoint
  c1 <- dgp$x[d1]

  a2 <- Fstats(dgp$y2 ~ 1 + dgp$x, from = 0.15, to = 0.85)
  d2 <- a2$breakpoint
  c2 <- dgp$x[d2]

  a3 <- Fstats(dgp$y3 ~ 1 + dgp$x, from = 0.15, to = 0.85)
  d3 <- a3$breakpoint
  c3 <- dgp$x[d3]

  a4 <- Fstats(dgp$y4 ~ 1 + dgp$x, from = 0.15, to = 0.85)
  d4 <- a4$breakpoint
  c4 <- dgp$x[d4]
```

```
#RD regressions based on estimated discontinuity point
r1e <- rdrobust(y1,x, c=c1, kernel = “uni”)
r2e <- rdrobust(y2,x, c=c2, kernel = “uni”)
r3e <- rdrobust(y3,x, c=c3, kernel = “uni”)
r4e <- rdrobust(y4,x, c=c4, kernel = “uni”)

#Bandwidth estimation
lb1 <- r1e$N_h[1]
lb2 <- r2e$N_h[1]
lb3 <- r3e$N_h[1]
lb4 <- r4e$N_h[1]

rb1 <- r1e$N_h[2]
rb2 <- r2e$N_h[2]
rb3 <- r3e$N_h[2]
rb4 <- r4e$N_h[2]

#Restricting the dataframe to bandwidth interval only and running indicator saturation
act.dgp <- dgp[(d1-lb1):(d1+rb1),]
T <- nrow(act.dgp)
x.sim <- sim(T)*act.dgp[,2]
colnames(x.sim) <- paste("x.sim", 2:nrow(x.sim), sep="")
reg <- as.matrix(act.dgp[,2])
m1 <- isat(act.dgp[,4], sis = 1, mxreg = reg, uis = x.sim, t.pval = 1/(2*T), mc = 0,
        parallel.options = 4, print.searchinfo=0)

act.dgp <- dgp[(d2-lb2):(d2+rb2),]
T <- nrow(act.dgp)
x.sim <- sim(T)*act.dgp[,2]
colnames(x.sim) <- paste("x.sim", 2:nrow(x.sim), sep="")
reg <- as.matrix(act.dgp[,2])
m2 <- isat(act.dgp[,5], sis = 1, mxreg = reg, uis = x.sim, t.pval = 1/(2*T), mc = 0,
        parallel.options = 4, print.searchinfo=0)

act.dgp <- dgp[(d3-lb3):(d3+rb3),]
T <- nrow(act.dgp)
x.sim <- sim(T)*act.dgp[,2]
colnames(x.sim) <- paste("x.sim", 2:nrow(x.sim), sep="")
reg <- as.matrix(act.dgp[,2])
m3 <- isat(act.dgp[,6], sis = 1, mxreg = reg, uis = x.sim, t.pval = 1/(2*T), mc = 0,
        parallel.options = 4, print.searchinfo=0)

act.dgp <- dgp[(d4-lb4):(d4+rb4),]
T <- nrow(act.dgp)
x.sim <- sim(T)*act.dgp[,2]
colnames(x.sim) <- paste("x.sim", 2:nrow(x.sim), sep="")
reg <- as.matrix(act.dgp[,2])
m4 <- isat(act.dgp[,7], sis = 1, mxreg = reg, uis = x.sim, t.pval = 1/(2*T), mc = 0,
        parallel.options = 4, print.searchinfo=0)
```

```
#store results
  return(list(m1$mean.results, m2$mean.results, m3$mean.results, m4$mean.results))
  }

for (k in 1:1000){
  simul <- lapply(5:2, isatfun)
  for (j in 1:4){
    for (i in 1:4){
      temp.sistbl <- simul[[i]][[j]] %>% tibble::rownames_to_column() %>%
                dplyr::filter(grepl("sis", rowname))
      temp.simtbl <- simul[[i]][[j]] %>% tibble::rownames_to_column() %>%
                dplyr::filter(grepl("sim", rowname))
      disc.tbl[k,i,j] <- sum(temp.simtbl$coef) + sum(temp.sistbl$coef)
      kink.tbl[k,i,j] <- sum(temp.simtbl$coef)
    }
  }
}

#Save simulation results in tables
dgp1.disc <- as.data.frame(disc.tbl[,,1]) %>% magrittr::set_names(c("5sd", "4sd", "3sd", "2sd"))
dgp2.disc <- as.data.frame(disc.tbl[,,2]) %>% magrittr::set_names(c("5sd", "4sd", "3sd", "2sd"))
dgp3.disc <- as.data.frame(disc.tbl[,,3]) %>% magrittr::set_names(c("5sd", "4sd", "3sd", "2sd"))
dgp4.disc <- as.data.frame(disc.tbl[,,4]) %>% magrittr::set_names(c("5sd", "4sd", "3sd", "2sd"))

dgp1.kink <- as.data.frame(kink.tbl[,,1]) %>% magrittr::set_names(c("5sd", "4sd", "3sd", "2sd"))
dgp2.kink <- as.data.frame(kink.tbl[,,2]) %>% magrittr::set_names(c("5sd", "4sd", "3sd", "2sd"))
dgp3.kink <- as.data.frame(kink.tbl[,,3]) %>% magrittr::set_names(c("5sd", "4sd", "3sd", "2sd"))
dgp4.kink <- as.data.frame(kink.tbl[,,4]) %>% magrittr::set_names(c("5sd", "4sd", "3sd", "2sd"))

dgp1.sis <- dgp1.disc - dgp1.kink
dgp2.sis <- dgp2.disc - dgp2.kink
dgp3.sis <- dgp3.disc - dgp3.kink
dgp4.sis <- dgp4.disc - dgp4.kink

#Save Data Environment
# save.image("D:\\Google Drive\\UVic MA Economics\\Thesis\\1. RDD Break
Detection\\Work\\Data\\linear.1-1000.200.rds")
# save.image("D:\\Google Drive\\UVic MA Economics\\Thesis\\1. RDD Break
Detection\\Work\\Data\\linear.1-1000.500.rds")
save.image("D:\\Google Drive\\UVic MA Economics\\Thesis\\1. RDD Break
Detection\\Work\\Data\\linear.1-1000.1000.rds")
```

R Codes – Application on Empirical Data

```
if(!require(pacman))install.packages("pacman")
pacman::p_load(tidyverse, strucchange, gets, rdrobust)

setwd("D:\\Google Drive\\UVic MA Economics\\Thesis\\Reference Papers\\Replication")
data <- read.csv("CIT_2019_Cambridge_polecon.csv")
# https://sites.google.com/site/rdpackages/replication/cit-2019-cup

#Plot
Y = data$Y
X = data$X
rdplot(Y, X, nbins = c(2500, 500), p = 0, col.lines = "red", col.dots = "black", title = "",
      x.label = "Islamic Party's Margin of Victory", y.label = "Female High School Completion Rate",
y.lim = c(0,70), cex.axis = 1.5,
      cex.lab = 1.5)
rdplot(Y[abs(X) <= 50], X[abs(X) <= 50], nbins = c(2500, 500), p = 4, col.lines = "red", col.dots =
"black", title = "",
      x.label = "Islamic Party's Margin of Victory", y.label = "Female High School Completion Rate",
y.lim = c(0,70), cex.axis = 1.5,
      cex.lab = 1.5)

#Detection when 50% Interval-Length is used
print("Using 50% cutoff score interval centered on true cutoff")
print("-----------------------------------------------------")
act.data <- data[order(data$X),]
cut.data3.2 <- act.data %>% filter(X >= -50) %>% filter(X <= 50)
paste0("No of observations = ", nrow(cut.data3.2))
a3.2 <- Fstats(Y ~ 1 + X + vshr_islam1994 + partycount + lpop1994 +
               merkezi + merkezp + subbuyuk + buyuk, data = cut.data3.2)
d3.2 <- a3.2$breakpoint
c3.2 <- cut.data3.2$X[d3.2]
paste0("estimated cutoff location = ",d3.2)
paste0("true cutoff location = ", cut.data3.2 %>% filter(X < 0) %>% nrow())
paste0("estimated cutoff value = ", c3.2)
paste0("true cutoff value = ", 0)
sctest(a3.2)

T <- nrow(cut.data3.2)
reg = as.matrix(cut.data3.2[,c(1,6,9,10,11,12,17,18)])
m3.2 <- isat(cut.data3.2[,2], iis = 0, sis = 1, mxreg = reg, t.pval = 10/(T), mc = 1,
        parallel.options = 4, print.searchinfo=0)
m3.2
cut.data3.2$X[1874] #estimated cutoff location using indicator saturation method.

#Bandwidth Restriction
#c = c3.2  #from Andrew's Test
c = 0.36  #from isat
out3.2 = rdrobust(Y, X, covs = Z, c = c, p = 1, kernel = "uniform", bwselect = "mserd", all = 1)
summary(out3.2)
```

```
lb <- out3.2$N_h[1]
rb <- out3.2$N_h[2]
cut.data <- cut.data3.2[(d3.2-lb):(d3.2+rb),] #cutoff at 220(Andrew's)/257(isat)

#Effect estimation
nrow(cut.data)
T <- nrow(cut.data)
reg = as.matrix(cut.data[,c(1,6,9,10,11,12,17,18)])
m3.2_1 <- isat(cut.data[,2], iis = 0, sis = 1, mxreg = reg, t.pval = 1/(T), mc = 1,
        parallel.options = 4, print.searchinfo=0)
m3.2_1

m3.2_5 <- isat(cut.data[,2], iis = 0, sis = 1, mxreg = reg, t.pval = 5/(T), mc = 1,
        parallel.options = 4, print.searchinfo=0)
m3.2_5

#using Andrew's cutoff
#sum(m3.2_1$coefficients[10:13])
#sum(m3.2_5$coefficients[13:14])

#using isat cutoff
sum(m3.2_1$coefficients[12:15])-sum(m3.2_1$coefficients[10:11])
sum(m3.2_5$coefficients[15:16])
```