Republic of Iraq
Ministry of Higher Education
and scientific Research
University of Al Qadisiyah
College of Administration and Economics
Department of Statistics

Variable Selection for Regression Discontinuity Designs with an Application

A Thesis submitted to
the Council of theCollege of Administration and Economics at University of Al - Qadisiyah in Partial Fulfilament of the Requirements for the Degree M.S.C in Statistics

By
Ashwaq AbdulSaada Kadhim

Supervised by
Dr. Bahr Kadhim Muhammad

2021 A.D.   **1443 A.H.**

بسم الله الرحمن الرحيم

﴿ وَمَا يَعْلَمُ تَأْوِيلَهُ إِلَّا اللَّهُ وَ الرَّاسِخُونَ فِي الْعِلْمِ يَقُولُونَ آمَنَّا بِهِ كُلٌّ مِنْ عِنْدِ رَبِّنَا ﴾

﴿ صدق الله العظيم ﴾

[آل عمران: 7]

Acknowledgments

DEDICATION

This work is dedicated to my beloved family, my dear husband, and my sons and daughters, for their constant support and always being with me.

Abstract

Regression discontinuity design (RDD) is a popular technique for causal inference with observational data. It has been applied to statistics, econometrics, political science, epidemiology, and related disciplines. The effects of treatment with (RDD) are often estimated with a small number of observations and when the correct functional form of the local regression is unknown. In this design, the study population is divided into two groups based on a specific cutoff point. In advance, according to the terms and requirements of the study. The estimation of the average effects of local processing in (RDD) is typically based on local linear regressions using the outcome variable, the processing task variable, and the continuous operation variable.

This study, combined some method variable selection, such as the Elastic net and MCP with RDD model. Three types of bandwidth, the IK (Iembens and kalyanman) bandwidth, cross-validation (CV) method, and The CCT (Calonico, Cattaneo & Titiunik) bandwidth were also used. In addition, covariates are often added to increase the efficiency of treatment effect estimates and to adjust for known imbalances. To evaluate the proposed methods, they were compared adaptive lasso method using simulation study employing the mean square errors to compare and select the best method .

These methods, were employed to study the issue of resorting to dialysis as a treatment, as a result of exposing to kidney failure to get rid of waste products and toxins in the blood such as urea, creatinine and others. Real data with a size of (79) people was used, represented by the percentage of urea in the blood as an

explanatory variable (Fi, the filtering variable) and the ability of the kidneys to perform their function as a response variable (Y_i).

The R program was used in the analysis process. It turns out that the best way to estimate the average effect of treatment and variable selection is the (MCP) method when using bandwidth (CCT) as it achieves the least (MSE) , and the results in the simulation study and real data indicate that the proposed methods have superior performance .

TABLE OF CONTENTS

<u>page</u>

Chapter Three (Application side)

Chapter Four (Conclusion and Recommendations)

References

List of Taple

Table   **page**

List of Figures

X

List of Abbreviations

Abbreviations	Meaning
RDD	Regression Discontinuity Design
LASSO	Least Absolute Shrinkage and Selection Operator
OLS	Ordinary Least Squares Method
MSE	Mean Squared Error
LATE	Local Average Treatment Effect
LLR	Local Linear Regression
SRD	Sharp Regression Discontinuity
V.S	Variable Selection
AMSE	Asymptotic mean squared error
CEF	Conditional Expectation Function
SCAD	Smoothly Clipped Absolute Deviation
MCP	Minimax Concave Penalty
IK	Imbens and Kalyanaraman
CV	Cross-validation
CCT	Calonico, cattaneo, and Titiunik
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
AL	Adaptive Lasso
BLUE	Best Linear Unbiased Estimate

CHAPTER ONE

INTRODUCTION, THE RESEARCH PROBLEM, THE AIM OF THE RESEARCH AND   REVIEW OF LITERATURE

1-1 Introduction

Regression Discontinuity Design (RDD) is one of the most valid non-experimental strategies for analyzing causal effects. as a way of estimating treatment effects in a non-experimental setting where treatment is determined by whether an observed "assignment" variable. RDD was first introduced by Donald Thistlethwaite and Donald Campbell (1960) RDD has many applications in statistics, econometrics, political science, epidemiology, and related disciplines.

RDD is a quasi-experimental pre test and post test design that extract the causal effects of interventions by assigning a cut off or threshold above or below which an intervention is assigned. By comparing observations lying closely on either side of the threshold, it is possible to estimate the average treatment effect in environments in which randomization is unfeasible. The RDD yields an unbiased estimate of the local treatment effect. It can be almost as good as a randomized experiment in measuring a treatment effect.

In the last two decades, causal inference in RDD has been a fertile area of research, and there has been a growing number of studies applying and extending RDD methods (Anastasopoulos, 2019) .
This chapter disply the problem of study, the goals of study and finally the literature review that related to this study.

1-2   The Problem of the Study

The problem of this study is centered on case that there are cutting point in the data. And in the same time there are many variables and therefore the difficulty of estimating and selecting the important variables using classical methods, and therefore there is a need to suggest a new methods that can deal with these problems.

1-3 The Aim of the Study

This study aims to employ some new variable selection methods (Adaptive Lasso, MCP, and Elastic Net) in estimating and selecting the variable for regression discontinuity designs for the purpose of using them in applied fields (economics, political sciences and related disciplines).

1-4 Literature Review

When highlighting the most important developments in RDD over the past years, there are several important studies in this area. Also,there are a good number of studies and researches that dealt with methods of estimating and selecting variables.

In 1960, a study was presented by Thistlewaite and Campbell, on discontinuity designs of regression as an important method of causal inference in the field of social sciences. They applied this design for the first time in award analysis on future academic results of students where awards are allocated based on an observed test score.

The main idea of the study is that the scores of the non-awarding participants were below the maximum, and the comparisons were good for those above the cutoff point (who took the awards), although this strategy was about 50 years ago, but it did not attract much of interest until recently (Thistlew aite and Campbell, 1960).

James Jackman et al. (1999), argued that RDD is a special case of selection over observations, and the discontinuity designs (RDD) estimator can be considered a specific form of conformance at one point (Lee, & Lemieux, 2010).

Hahn, et al. (2001) indicated that the main assumption of a correct RDD is that "all factors are continuous for X values, and for the purpose of estimating the local average treatment effect τ , a nonparametric procedure has been proposed and no basic linear procedure is assumed. The RDD is used more formally using the framework. The "potential outcomes" of the effects of the graph-assisted therapy

literature. It is common for there to be a pair of "potential" outcomes for each individual $Y_i(1)$ when the unit undergoes treatment and $Y_i(0)$ if the unit is not treated (Hahn, et al., 2001).

In 2002, Wilbert estimated the parameter of the effect of financial aid provided by colleges to students on student enrollment in those colleges using RDD , where the study was conducted on a group of East Coast College students, and the study reached a positive value for the effect parameter, indicating that the financial aid provided by East Coast College plays a fundamental role in the standards students compete against, especially students from low-income families, compared to other colleges (Van der Klaauw, 2002).

In 2003, Porter published a paper in which he addressed the derivation of the optimal convergence rate to estimate the effect of regression treatment in RDD models, where the optimal convergence rate was derived to estimate the effect of discontinuity regression treatment under the problems of bias (Porter, 2003) .

Zou (2006) developed adaptive lasso method for the purpose of maximizing the selection of the correct variable to solve problems of estimating low and high dimensions. This makes it ideal for selecting covariates in RDD and other causal inference procedures in which variable modification is appropriate. (Zou, 2006).

Guido Imbens and Thomas Lemieux (2008) published a paper titled "Regression Discontinuity Designs: A Guide to Practice" that stated that for the purpose of assessing causal effects in RDD. Treatment is at least partially assigned a value from the observed covariate that falls on either side of the fixed threshold. In the economic fields, interest in these methods was not greatly appreciated until the late 1990's, after which many studies were applied in economics as these methods were developed and expanded. (Imbens, & Lemieux, 2008) .

Dunning 2008, Robinson, et al., (2009) note that sociologists have come up with a basic knowledge of research design in causal inference. So, scientists have largely resorted to quasi-experimental designs that take advantage of random manipulation, and that the discontinuity design exploits situations in which experimental units are allocated to treatment based on the result of the cut. (Dunning 2008, Robinson, et al., 2009)

Cook, et al., 2008) stated that the RDD developed by Donald Campbell in the early 1960 had not received much attention from statisticia (Thistlewaite and Campbell 1960; Campbell and Stanley 1963). Then the RDD was developed by estimating the useful statistical properties of this design and applying it to various fields. This methodology was specifically developed in experimental practices (Cook, et al.,2008).

Meier, et al., (2008) created additional versions of lasso to solve high-dimensional regression problems, as they included adjustments to the penalty

included the Elastic Net , which combined Ridg's penalty and Lasso's penalty with a "group lasso " used for the purpose of selecting a large set of covariates (Meier, et al., 2008).

Imbens and Kalyanaraman, (2009) presented a study that included investigated the problem of optimizing the smoothing parameter (bandwidth) for estimating the discontinuity regression function. the researchers focused on the estimation method using the polynomial function, they also proposed a new mechanism for selecting the bandwidth using the data. (Imbens, & Kalyanaraman, 2009).

David and Thomas Lemieux, (2010) presented a concept and guide for users for discontinuity design to experimental researchers, where the basic theory of this design was presented to test its validity. In addition, they show the interpretation of the quasi-experimental design, which summarizes the different methods using in estimating discontinues effect coefficient of regression and the constraints imposed on the interpretation of these estimates.   The general concepts of the discontinuity

design of regression are discussed using a set of empirical examples (Lee, & Lemieux, 2010).

Cauchy and Sikhon (2011), show that the introduction to these designs was simple by customizing the treatment around a small window sill, and it appears

that the individuals or units on either side of this small window have assigned a random treatment to them. While RDD is simple in principle, in practice estimating the effect of a treatment can become complex, and the use of optimal bandwidth selection algorithms that use MSE miniaturization and validation as shown often requires a variety of parametric and non-parametric models, even before the introduction of covariates (Cauchy and Sekhon, 2011).

José, et al., (2012) used RDD for the purpose of measuring the impact of national research funds, in which they analyzed national research funds in promoting scientific production in economically developing countries. This study dealt with the Chilean National Fund for Science and Technology Research(FONDECYT), and the results had a significant and positive impact in terms of publications. (Benavente, et al., 2012).

Bloniarz, et al., (2016) proposed a method of shrinkage and variable selection to identify covariates, as this method was modified and expanded to estimate local average treatment effect (LATE) in RDD through the use of adaptive Lasso, a version of lasso with an advantage owning Oracle Properties (Bloniarz, et al., 2016) .

Cattaneo et al., (2017) discussed the effect of the front of the head on infant mortality using RDD model. Both parametric and non-parametric method was used, for the purpose of estimating the effect of treatment proposed by specialists on the number of deaths (Cattaneo, et al., 2017).

Jason Anastasopoulos, (2019) suggested the use of local average treatment effect (LATE) estimates in RDD on local linear regressions using the outcome variable, processing effect variable, and the continuous operating variable, and demonstrated that incorporation of the adaptive lasso into RDD treatment effect estimation can improve the efficiency of treatment effect estimates when covariates are included and can also provide a principled framework of treatment effect adjustment for RDDs. (Jason Anastasopoulos, 2019)

Matthias de Cattaneo et al, (2019) provided a practical guide to access the analysis and interpretation of RDD. They also discussed the principles of Sharp RDD which is characterized by the fact that the result is distributed continuously, has only one dimension, has one cut-off point and matches the treatment task. (Cattaneo et al.,2019).

Albino, (2020) presented a study for multiple RDD model, where he discussed how the number of untestable variables varies according to the degree of accuracy of data and the degree of concentration at the boundaries of discontinuity by using multiple thresholds to reduce these variables to a minimum and thus increase inference efficiency (Albino, 2020).

Nathan Kitelwell and Peter Siminsky, (2020) conducted a research for optimal model selection in RDD using placebo areas. They showed that the RDD is

attractive model because it serve to obtain an estimate of causal effects easily under relatively weak assumptions.   The possibility of RDD in simple portions of the operating and outcome variable (Calonico et al. 2015) provided an attractive atmosphere of transparency for this approach (Kettlewell, & Siminski, 2020).

Millie Blaise and Lalive Rafael, (2020) indicated that the RDD is receiving significant attention in many areas, among them labor economics, political economy, health economics, criminology, environmental economics, and development economics. They explained that there are two main requirements for discontinuity design. First, treatment must be a discontinuous function of acontinuous and observable variable, which is called the operating variable, assignment, or coercion. And second, there is a jump in norm that defines treatment tasks at some thresholds (Melly, & Lalive, 2020) .

In this study, we suggested the use of some variable selection methods (MCP and Elastic Net) with RDD model, which will be mentioned later in the second chapter and their application in simulation and real data in the third chapter using the (R) program because of the efficiency and superiority of the two methods compared to the adaptive Lasso method with (RDD).

CHAPTER TWO

THEORETICAL SIDE

2-1 Introduction

Variable selection methods are one of the well-sophisticated field in the modern statistics.

The most widely used approach in variable selection methods models is to construct the functional relationship between two or more variables, i.e. to predict the response variable from one or more explanatory variables. In this chapter litirature reviews of important methods that deal with variable estimation and selection methods are adopted, with one of the most commonly used models, which is the regression discontinuity design model.

2-2 Regression-Discontinuity Design (RDD)

RDD model is divided into two groups on the basis of a specific threshold limit or the so-called cutoff point (Thistlethwaite and Campbell, 1960). This point is determined in advance according to the study conditions and requirements.

The importance of calling it a discontinuity design comes from the fact that the treatment effect will lead to a 'jump' or discontinuity 'in the regression function point of the relationship between $X_i = F_i$ (an explanatory variable) and the dependent variable Y_i,(Lee, & Lemieux, 2010) .

Simply it can be said that the RDD divides the study population into two groups, the bottom and the highest cutoff point, as the case of the randomized trial. The RDD assumes the equivalence of the two groups in the absence of the treatment effect.   In experimental designs, the random allocation ensures that the

treatment and comparison groups are equal in the beginning, and the difference can be attributed to the treatments i.e. the effect of the treatments . The units defined in this design may be specific individuals or units such as hospital wards, counties ...so on . The basic RDD model can be depicted as shown in Figure (2-1) .
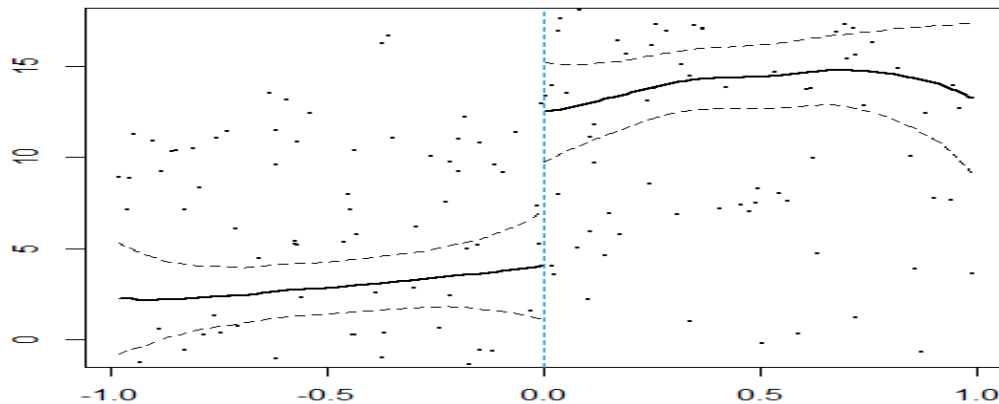
Figure (2-1): The cut point for RDD

Several researchers have provided practical evidence that facilitates the understanding of (RDD), especially that which contains certain features such as the fact that the distribution of the result is continuous, and has only one dimension, and one cutoff point, and has the characteristic to comply with the treatment in an integrated manner, that is, The actual treatment is received by all units that have points equal to or greater than cut points, and the treatment is not received by units that have points less than cut points. These characteristics are called Sharp Segression Discontinuity (SRD), but recent studies came out with a new design called regression Fuzzy Regression Discontinuity and its abbreviation (FRD) as well as multi-segment designs (RD) and geographic designs (RD) with effect variables discontinuous , but in this study we'll deal with Sharp Regression Discontinuity (SRD). (Cattaneo, et al., 2019).

2-3 Assumptions of the Discontinuity Design Analysis.

The Cutoff Criterion: A pre-determined and precisely defined cutting standard that shall followed on the classification variable. On the basis all pilot units are classified without exception, unless some necessary adjustments are required. This leads to the emergence of a problem that threatens the selection process, as estimating of the treatment effect of the experimental units are likely to be biased. (Cappelleri, & Trochim, 2015).

The Form of the Functional Relationship before-after : The form of the functional relationship known and correctly defined. Is It a linear model, a polynomial model, a logarithmic model, or something similar. There are difficulties for determining the appropriate model, in the case that some nonlinear curves represent the shape of the functional relationship between before - and after cutpoint.

The Pre-test Variance of the Comparison Group: There shall be a sufficient range of values for the pre-test in the comparison group to be able to appropriately estimate the regression model before - after for that group.

Continuity for the Pretest: both groups shall have the same continuous distribution for the pretest.

Implementation of the Treatment (the program): It is assumed that all experimental units undergo treatment equally or uniformly - because if this does not happen. It will complicate the analysis for the user. (Treasury, 2007).

2-4 Advantages of RDD

- RDD results is an unbiased estimate of the effect of a local treatment, when properly implemented. (Cappelleri, & Trochim, 2015)
- It is used to measure the effect of treatments as in randomized trials, and a good estimate, will be got .
- RDD, as a quasi-experimental method, does not require prior randomization.
- When properly executed, RDD studies generate treatment effect estimates similar to randomized studies.

2-5 Disadvantages of RDD

- The statistical analysis of this design is complex. (Treasury, 2007)
- It requires a large sample size to detect the effect of treatment, and data collection is difficult.
- It is more difficult to interpret the results of the RDD than the results of other designs.
- The effect of discontinuity is estimated, so the number of observations is less than the random trial.
- Classifications are sensitive to the shape of the non-linear functional relationship with the presence of interactions, and their behavior in a specific way, so that part of it is linear and part nonlinear.

2-6 Local Linear Regression (LLR)

Local linear regression is a non-parametric method that is used to continuously estimate the average treatment effect in RDD model (Imbens, & Lemieux, 2008; Cattaneo, et al. 2019) .

This method is considered as one of the important nonparametric methods in estimating because of the need to estimate the regression function in a consistent manner at a certain point. The estimator (LLR) is preferred over the rest of the nonparametric methods in the case of discontinuity regression because it has distinctive properties at the limits and has the best estimation rate very clearly (Porter, 2003). Neighborhood idea h is the basis of the LLR. Where h bandwidth is chosen in this method, on either side of the cut-off point, the observations closest to the cut-off point (c) are given more weight than those farther away Observations $(c - h$ and $c + h)$.

The average weight to weigh the adjacent observed data is determined by the kernel which is a statistical technique for estimating the reality of the function. The kernel function K (u): R → R. It has the following properties (Mutair, Hafez Muhammad, 2011) .

1. $0 \leq k(u) \leq \infty$, $K(u)$ is a continuous function with non-negative real values
2. $\int_{-\infty}^{\infty} K(u)du = 1$

3. K(u) is a symmetric function around zero, $\int u\, K(u)\, du = 0$ and $\sigma^2 = \int u^2\, K(u)\, du > 0$.

2-7 Parametric and Non-Parametric Regression Model.

The main problem that faces any non-experimental method is the careful selection of experimental units or observations in the same conditions of the experiment to be studied due to the lack of sufficient information about the experiment or the inability to implement it. (Lee, & Lemieux, 2010).

Determining the type of the model and the form of the relationship between the variables in the model will give an accurate and realistic estimation process, as this is applied to the RD design, as there may not be enough information about the form of the relationship between the classification variable and the response variable. (Lee, & Lemieux, 2014). The relationship between the response variable and the classification variable when using quasi-experimental designs is a counterfactual relationship because there is a regular difference between the classification variable and the response variable. As a result, a variety of tests are conducted that include different models and functions and the estimated values of the studied phenomenon are calculated by one of the parametric estimation methods (the least squares method) and nonparametric estimation methods (the kernel function), as explained below :

2-8 Parametric Linear Regression Model:

2-8-1 Simple Parametric Linear Regression Model in RDD.

The simplest parametric method for implementing the RDD is by computing the estimated the parameters of the two simple linear regression models of observations within the best chosen bandwidth for the design to the right and left of the cutoff point, of the response variable (Y_i) on the rating variable (F_i).

The value of the cutoff point is subtracted from each observation of the classification variable on both sides of the cutoff point, where the explanatory variable (rating variable, F_i) is transformed into the variable $(F_i - c)$ to obtain the value of the fixed boundary at the cut point . F_i (forcing variable). (Imbens, & Lemieux, 2008).

Regression models can be written on both sides of the cutoff point as follows:

$$Y_i = \alpha_l + \beta_l (F_i - c) + \epsilon_i \qquad \ldots\ldots (2\text{-}1)$$

The model (2-1) represents the simple linear regression from the left side of the cutoff point when $(F_i < c)$.

$$Y_i = \alpha_r + \beta_r (F_i - c) + \epsilon_i \qquad \ldots\ldots (2\text{-}2)$$

The model (2-2) represents the simple linear regression from the right side of the cutoff point when $(F_i \geq c)$.

Where : α_l, α_r represent the intersections of the two regressions at the cut-off point. β_l , β_r represent the slope of the regression line from the left and right of the cut-off point.

F_i: (forcing variable) .

2-8-2 Polynomial Parameter Linear Regression Model in RDD.

Polynomial linear regression models are one the popular methods for obtaining a direct estimate of the effect of a treatment or intervention. This is done by building a regression model for all observations that fall on both sides of the cutoff point by adding the treatment variable or the dummy variable (T_i) ,

(intervention), and in this method the polynomial regression models are estimated for various shapes with interaction and without interaction ,between the rating variable and the outcome. A variety of functional forms can be tested to determine which fits the data best, so that bias will be minimized. For example, the following models are often tested in the parametric analysis of the RD design, as shown below (Jacob, et al., 2012):

$$Y_i = \alpha + \tau\, T_i + \beta_1(F_i - c) + \epsilon_i \quad \dots \quad (2\text{-}3)$$

Where: Y_i : Response variable. τ: Local average treatment effect.

T_i : Treatment variable (is a binary treatment indicator function which equals 1 when $F_i \geq$ c and equals 0 when $F_i \geq c$). F_i : Forcing variable.   c: Cutoff point .
ϵ_i : The limit of the random error.

   The equation $(2-3)$ is the polynomial model of the first order without interaction, whereas:

$$Y_i = \alpha + \tau\, T_i + \beta_1(F_i - c) + \beta_2(F_i - c)\, T_i + \epsilon_i \dots (2\text{-}4)$$

  The equation (2-4) is the polynomial model of the first order with interaction.

$$Y_i = \alpha + \tau\, T_i + \beta_1(F_i - c) + \beta_2(F_i - c)^2 + \epsilon_i \dots (2\text{-}5)$$

   The equation (2-5) is the polynomial model of the second order without interaction.

$$Y_i = \alpha + \tau\, T_i + \beta_1(F_i - c) + \beta_2(F_i - c)^2 + \beta_3\,(F_i - c)\, T_i + \beta_4\,(F_i - c)^{2\,T_i} +$$
$$\epsilon_i \dots (2\text{-}6)$$

The equation (2-6) is the polynomial model of the second order with interaction.

$$Y_i = \alpha + \tau\, T_i + \beta_1(F_i - c) + \beta_2(F_i - c)^2 + \beta_3\,(F_i - c)^3 + \epsilon_i \ldots (2-7)$$

The equation (2-7) is the polynomial model of the third order without interaction.

$$Y_i = \alpha + \tau\, T_i + \beta_1(F_i - c) + \beta_2(F_i - c)^2 + \beta_3\,(F_i - c)^3 + \beta_4\,(F_i - c)\,T_i +$$
$$\beta_5\,(F_i - c)^2\,T_i + \beta_6\,(F_i - c)^3\,T_i + \epsilon_i \ldots (2-8)$$

The equation (2-8) is the polynomial model of the third order with interaction. (Linden, et al 2006).

The cutoff point value was subtracted from the classification variable observations to get rid of the linear multiplicity problem, and the rating variable (intervention) value represents the effect parameter (gap) value.

2-9   Non-Parametric Regression Model:

2-9-1 The Simple Linear Non-Parametric Regression Model in RDD.

The simplest nonparametric method for implementing the RDD is by computing the estimated values of the response variable of the two simple linear regression models of observations within the best chosen bandwidth for the design to the right and left of the cutoff point (Lee, & Lemieux, 20 10). The difference between estimated values to the response variable to observations on cutoff point and equal to it, is the amount of the impact factor value that represents the value of the treatment effect on the observations thereof (Calonico, et al., 2015).

2-9-2 Nonparametric Polynomial Linear Regression Model in RDD.

Polynomial linear regression is a nonparametric method for estimating the treatment effect on discontinuity design. Hahn (2010), Porter (2003) Lee and Lemieux (2010) focused on kernel functions to estimate nonparametric regression. One of the most important challenges in the nonparametric regression method is the selection of the appropriate kernel function and that the simplest method non-parametric it is the estimation of the regression equation for each side of the cutoff point and obtaining the estimated value of the observation that falls on the cut point (Lee, & Lemieux, 2010). Another important challenge in the nonparametric estimation method is choosing the appropriate bandwidth which requires a balance between accuracy and bias. The simple nonparametric regression model is one of the simplest and most used nonparametric regression models. Then, it is represented by the following formula:

$$Y_i = m(x_i) + \varepsilon i \qquad \dots (\,2\text{-}9\,)$$

$$E(y_i) = \widehat{m}(x_i) \qquad \dots (2\text{-}10)$$

where: Y_i: response variable.

$m\ (x_i)$: represents the unknown regression function to be estimated by nonparametric method . ε_i : the random error that is assumed to be normally distributed with a mean of zero and a constant variance (σ^2) i.e. $\varepsilon_i \sim N(\,0,\,\sigma^2)$. Estimation methods are based on non-parametric smoothing methods, that are the main input to non-parametric regression. It aims to find the best curve that matches or approaches the curve of the response variable. One of the nonparametric smoothing methods used in the smoothing of the response variable

(Y_i) is the kernel function, as mentioned in paragraph (2-6) (Imbens, & Lemieux, 2008) . The variable (u) when applying the kernel function for the purpose of smoothing out the discontinuity design taking the following form:

$$u_i = \frac{Fi - c}{h} \qquad , \quad i = 1,2, \ldots, n$$

where; F_i : forcing variable, c: cutoff point and h: bandwidth parameter.

2-10   Local Average Treatment Effect (LATE)

Discontinuity designs are a framework for causal estimation of the local average treatment effect with the observed data using a forcing variable F_i, i = 1, ...n based on a known threshold value. If $F_i > f$, the experimental units that receive treatment takes the value (1) i.e., $T_i = 1$, while the untreated experimental unit takes the value (0) i.e., $T_i = 0$ (Anastasopoulos, 2019). The effect variable is assumed to be in the case of continuity, and therefore the sharp discontinuity design will enhance this condition by reducing the causal effect of the LATE estimates around a narrow window of threshold $f - \varepsilon < f < f + \varepsilon$.

Where it is assumed, within this window, that the experimental units have been allocated to treatment randomly. This design used frequently in political science Public Policy Literature .

$$LATE = \tau = \lim_{F_i \downarrow 0} E[Y(1)_i | F_i = f + \varepsilon] - \lim_{F_i \uparrow 0} E[Y(0)_i | F_i = f - \varepsilon] \ldots (2 - 11)$$

Where:

LATE = τ: Local average treatment effects.

$Y_i(1)$: The outcome variable for the experimental unit receiving treatment.

$Y_i(0)$: The outcome variable for the experimental unit that do not receive treatment

F_i : *The* forcing variable or rating variable.

f : Threshold value   .

ε : narrow window of threshold (a small number of observations)
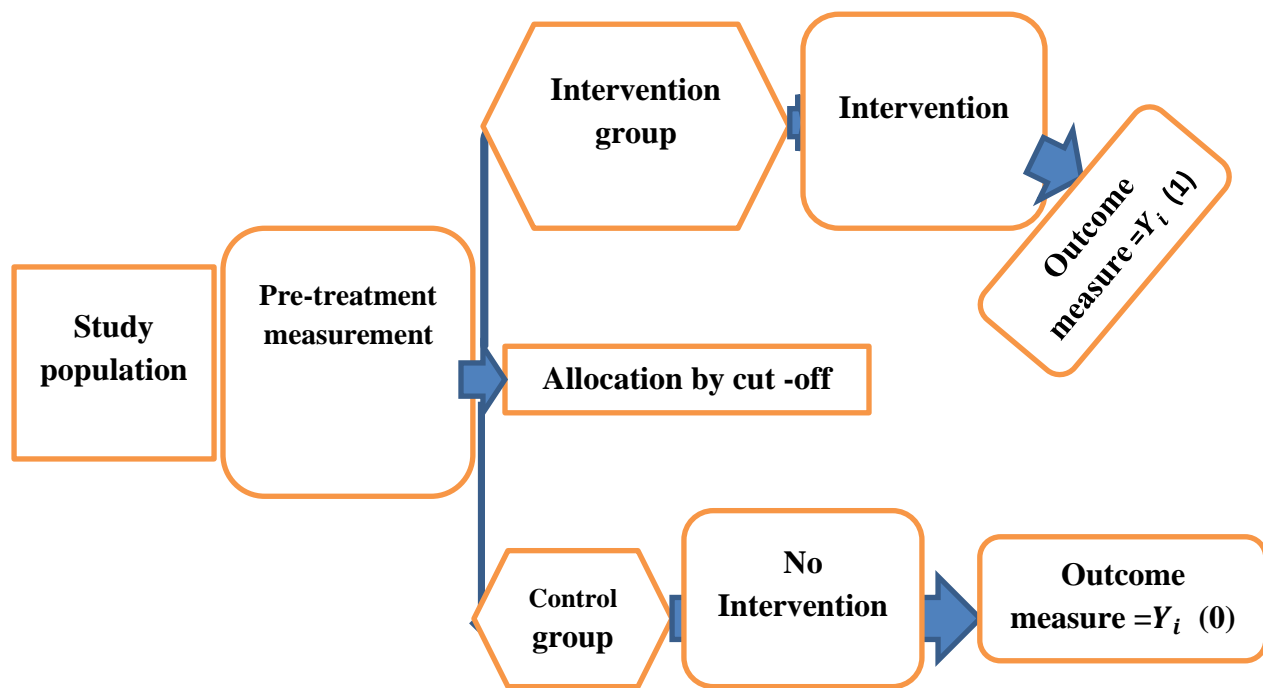
Figure (2-2): A schematic diagram for regression discontinuity design (RD),
(From the researcher's work).

The simplest method to estimate the treatment effect is by using local linear regression (LLR) in a neighborhood of the cutpoint $F_i \in [c - h, c + h]$,which is determined through optimal bandwidth selection procedures designed to minimize

cross-validated MSE (Imbens, & Kalyanaraman, 2012). Where (c) denotes the cutoff point, (h) denotes the bandwidth.

$$\hat{Y}_i = \alpha + \hat{\tau}T_i + \delta_1 F_i + f(T_i, F_i) \quad \dots (2\text{-}12)$$

where:

$\hat{Y}_i$: Observation outcome variable estimator i.

$\hat{\tau}$:   Estimated local average treatment effect.

T_i: An indication whether or not to receive treatment:

$$T_i = \begin{cases} 1 & \text{receiving treatment} \\ 0 & \text{Not receiving treatment} \end{cases} \quad \dots (2\text{-}13)$$

δ_1: Coefficient of the forcing variable.

$f(T_i, F_i)$: Is a function of the force variable which is in the form of nonparametric kernel or a polynomial .

One of the most common models for local linear regression (LLR) is the following model: (Anastasopoulos, 2019) .

$$\hat{Y}_i = \alpha + \hat{\tau} T_i + \delta_1 F_i + \delta_2 (F_i . T_i) + X\beta \quad \dots (2\text{-}14)$$

where:

δ_2: Coefficient $\left(F_i . T_i\right)$. X : Matrix of covariates $(n \times P)$ $.\beta$: Coefficients vector $\beta_{p \times 1}$

A set of X covariates were added in equation (2-14), to increase the accuracy of the local average treatment effect.   Calonico, et al., (2019) derived the covariate adjusted estimator of $\hat{\tau}$ and demonstrate that working on modifying this covariate before treatment results in more efficient estimates $\hat{\tau}$.

(Bloniarz , et al., 2016) proposed the selection of covariates using the Lasso method. This method was developed and expanded to include estimating the local average treatment effect in RDD through the use of adaptive Lasso which is a modified version of lasso that possesses Oracle properties (Zou, 2006) . It ensures that natural selection is consistent for estimating (τ) and the selection of the variable. It indicates the asymptotic unbiasedness of $\hat{\tau}$.

2-11 The Sharp RDD

An RDD model feature is that all experimental units under study receive a score known as (running variable, forcing variable or index). Treatment is assigned to experimental units whose scores are higher than the known cutoff point.   Also, treatment is not assign to units whose scores are lower than the cutoff point (Cattaneo, et al. 2019). An important feature of RDD is that it is defined by three components (score, cutoff and treatment). Suppose we have a random sample (n) of observations, taken from a large sample of size (N), $Y_i(1)$ denotes the potential outcome of unit (i) that received treatment for each $i = 1, 2, \ldots , $ n and $Y_i(0)$ indicates the likely outcome of unit (i) that doesn´t receive treatment.   This can be summarized as follows :

$$Y_i = \begin{cases} Y_i(0) & \text{if} \quad T_i = 0 \\ Yi(1) & \text{if} \quad T_i = 1 \end{cases} \ldots (2\text{-}15)$$

The basic idea for designing the sharp discontinuity (SRD) is that the treatment value T_i can only be determined using the F_i value on either side of the known cutoff point (c) .(Imbens, & Kalyanaraman, 2012).

$$T_i \;=\; 1 \quad if \quad F_i \;\geq c$$

Emphasis in SRD design is on average treatment effect for units with variable values equal to cut.

$$\tau_{SRD} = E[Y_i(1) - Y_i(0)|\, F_i = c] \quad \dots\dots\dots\dots\dots \text{ (2-16)}$$

where:

τ_{SRD}: Average treatment effect of SRD design.

2-11-1 Sharp Discontinuity Design Estimation Method

The focus is on practical appreciation of the treatment effect in designing a sharp RD of regression function at a single boundary point. The nonparametric regression of the kernel function does not work properly at the boundary points, and as a result the rate of convergence is slower at those   points. The regression is non-parametric at the boundary (Imbens, & Lemieux, 2008).

$$\hat{\tau}_c^{SRD} = \widehat{m}_l(\text{c}) - \widehat{m}_r(c) = \frac{\sum_{i:F_i < c} Y_i^{obs} . k(\frac{F_i - c}{h})}{\sum_{i:F_i < c} \quad . k\left(\frac{F_i - c}{h}\right)} - \frac{\sum_{i:F_i \geq c} Y_i^{obs} \; . \; k\left(\frac{F_i - c}{h}\right)}{\sum_{i:F_i \geq c} \quad . \; k\left(\frac{Fi - c}{h}\right)} \quad \dots \text{ (2-17)}$$

where:

$\hat{\tau}_c^{SRD}$: Estimation average treatment effect at the cut-off point in SRD design.

$\widehat{m}_l(c)$: The nonparametric regression function estimated to the left of the cutoff point. $\widehat{m}_r(c)$: The nonparametric regression function estimated to the right of the cutoff point. $Y_i{}^{obs}$:   Observed data.

$$Y_i^{obs} = Y_i(0)(1 - T_i) + Y_i(1)\, T_i... \textbf{(2-18)}$$

SRD design is called any RD design in which the treatment assignment is identical to the treatment state already received for all units while the RDD in which compliance with the treatment assignment is incomplete is called Fuzzy regression discontinuity (FRD). Here, the focus on the SRD design need one degree and one cut point (Cattaneo et al., 2019).   In Figure (2-3), the drawing shows that the conditional probability of receiving treatment in the SRD design is by looking at the degree $P\ (T_i\ =\ 1|F_i\ =\ c\ )$, for different values of the running variable F_i as shown below, and this probability changes from zero to one at the cut point. Since the specific treatment in the SRD model and the treatment received are identical, reflecting both the treatment task and the receive treatment .
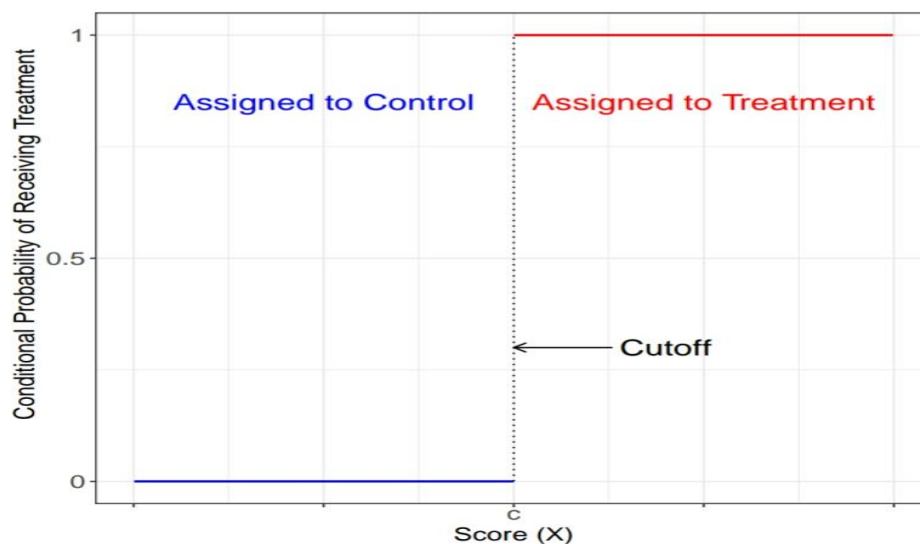
Figure (2-3): The conditional probability of receiving treatment in the Sharp RD design (Cattaneo et al., 2019)

2-12 Bandwidth Selection

Bandwidth is an unrestricted parameter (Free parameter) that has a clear role in the estimation process as it greatly affects bias and variance, as the more bandwidth increasing, the bias increaseing and the variance decreaseing and vice versa . So, it will have a clear effect on smoothing the curve and the rate of its approach to the original curve. (Imbens, & Lemieux, 2008). The basic idea of choosing a bandwidth for the SRD is a trade-off between bias and variance for $(\hat{\tau}\ _c^{SRD})$. There are several methods for choosing the optimal value of the bandwidth that have been used by many researchers such as Cross Validation(CV) and plug-in method, ..., so on. In this study we will deal the methods using by the researcher.

2-12-1 Iembens and Kalyanaraman (IK) Method.

The IK method was suggested by (Imbens and Kalyanaraman, 2009). The bandwidth estimator has been developed and its approximate characteristics are discussed. This estimator is fully supported by data. The different components of optimum bandwidth depend on the substitution of consistent estimators as well as plug-in estimators. The researchers explained that the optimal choice of the Bandwidth Optimal is by substituting the six unknown quantities shown in the equation (2-19), which will ultimately lead to the consistent estimators.

$$h_{opt}^{IK} = \arg\min_h AMSE\ (h) = CK \ . \left(\frac{\sigma_+^2\ (c)\ +\ \sigma_-^2\ (c)}{f(c).\ (m_+^{(2)}(c)\ \_ \ m_-^{(2)}(c))} \right)^{1/5} . \ N^{-1/5} \ ...(2\text{-}19)$$

h_{opt}^{IK}: The Bandwidth Optimal using (IK) method.

Whereas: $AMSE\ (h)$: the asymptotic mean squared error, its a function to bandwidth (h), represented by following formula :

$$\text{AMSE}(h) = C_1\ h^4(\ m_+^{(2)}(c)\ -\ m_-^{(2)}(c))^2 + \frac{C_2}{N.h}\ \left(\frac{\sigma_+^2\ (c)}{f(c)} + \frac{\sigma_-^2\ (c)}{f(c)}\right) \dots . (2-20)$$

In equation (2-20) the first term corresponds to square bias while the second term corresponds to variance. In this approximation, C1 and C2 are constants and represent the kernel functions .(Imbens, & Kalyanaraman, 2009).

whereas:

$$C_1 = \frac{1}{4}\ \left(\frac{v_2^2\ -\ v_1\ v_2}{v_2\ v_0\ -\ v_1^2}\right)^2 , \quad C_2 = \frac{1}{4}\ \frac{v_2^2 \pi_0\ -\ 2\ v_1\ v_2 \pi_1 + v_1^2\ \pi_2}{(v_2\ v_0\ -\ v_1^2\ )^2}$$

$$v_j = \int_0^\infty u^j\ K(u)du\ , \qquad \pi_j = \int_0^\infty\ u^j K^2(u)du,$$

$$CK = \left(\frac{C_2}{4.C_1}\right)^{1/5} \qquad \dots . (2\text{-}21)$$

In the event that the right-hand term is not equal to the left-hand term for the second derivative $[(m_+^{(2)}(c) \neq m_-^{(2)}(c)]$, the main term of expansion the bias is of the rank (h^4). When right term is equal to the left term$[(m_+^{(2)}(c) = m_-^{(2)}(c)]$, the bias converts to zero faster, allowing the estimate of $\hat\tau^{SRD}$ to be obtained at a faster convergence rate. Derivative $m\ (x)$, where $m(x)$ represents the conditional mean.

$$m(x) = E[Y_i|F_i\ =\ f]$$

$(m_+^{(2)}(c)\ ,\ m_-^{(2)}(c)$: represent right term and left term to second derivative $m\ (x)$ to threshold (c) for conditional mean .

$\hat{\sigma}^2 (x)$: represents the conditional variance function, $\hat{\sigma}^2 (x){=}Var\ (Y_i\,|\,F_i{=}\ f\ )$ is bounded in an open neighbourhood of $F = c$, right bound is continuous and so is the left at the cutoff point c.

$(\sigma_+^2\ (c)\ ,\ \sigma_-^2\ (c))$ are the left and right limits at the threshold c, $\sigma_+^2\ (c){>}0$, $\sigma_-^2\ (c) > 0$.

In this approximation, C1 , C2 and CK are constants and represent the kernel functions. $f(\cdot)$ denoted to marginal destribution to forcing variable, which is a continuous and limited away from zero to threshold (c). When obtaining the six unknown estimators in equation (2-19), the optimum bandwidth estimate is according to the following formula:

$$\hat{h}_{opt}^{IK}= \arg \min_h AMSE(h) = CK \left(\frac{\hat{\sigma}_+^2(c)+ \hat{\sigma}_-^2(c)}{\hat{f}(c).(\hat{m}_+^2\ (c)_-\hat{m}_-^2\ (c))}\right)^{1/5} N^{-1/5}\ ....\ (2\text{-}22)$$

$\hat{h}_{opt}^{IK} =$ The Bandwidth Optimal estimator using (IK) method.

2-12-2 The Cross-Validation (CV) Method.

The Cross-validation (CV) approach proposed by miller and Ludwig, (2007). This method is considered one of the best and most used methods of selecting the bandwidth, and it is called the method (leave – one - out) in which one observation is excluded from the values of the observations, As it is the main part of the process of balance between both the variance and the bias, as the more the variance value decreaseing, the value of the bandwidth increases and the bias value begins to increase. The bandwidth affects the shape of the curve if it is small and close to it (Ludwig, & Miller, 2007) .

Criterion Cross-validation (CV) is calculated according to the following formula :

$$CV_Y(h) = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y})^2$$

Where: $CV_Y(h)$ = criterion value cross -validation of the packet (h) .

Y_i : The real value of observation i . $\hat{Y}$: the estimated value of the response variable.

The package width that has the lowest v

.3alue for the Cross-validation criterion (CV) is chosen according to the following formula:

$$h_{CV} = \min_{h>0} CV_Y(h) \dots (2\text{-}23)$$

2-12-3   Calonico, Cattaneo & Titiunik (CCT) Method.

The CCT method was proposed by Calonico et al.(2014). Thus, the steps for this method are summarized as follows :

Step1: Estimating the (asymptotic variance)limits by finding the initial bandwidth
(Vn, Cn) denoted by (V) where (Calonico et al.,2014):

$$\hat{V} = 2.58 \, \omega \, n^{-\frac{1}{5}} \dots (2\text{-}24)$$

where:

$$\omega = min \left\{ S_x , \frac{IQR_x}{1.349} \right\} \dots (2\text{-}25)$$

(S_x) Denotes the sample variance, (IQR_x) indicates the interquartile range, and

the bandwidth ($\hat{C}_n$) where :

$$\hat{C}_n = \begin{pmatrix} 1/(2q+5) \\ v \quad , \quad p \ , \ q \end{pmatrix} . \ n^{-1/(2q+5)} \dots \text{(2-26)}$$

v : denotes the derivative of interest (τ).

p: denotes the order of the local polynomial point estimator.

q : denotes the order of the local polynomial bias estimator.

Step 2: finding the bandwidth ($\hat{b}_{CCT}$) and it is calculated according to the following formula (Ali, et al., 2020) :

$$\hat{b}_{CCT} = \begin{pmatrix} 1/(2q+5) \\ 0 \ , \ 1 \ , \ 2 \end{pmatrix} \ \dots (2-27)$$

$\hat{b}_{CCT}$: Estimated experimental bandwidth by (CCT) method.

$$\hat{C}_{0,1,2} = \frac{5n \ \hat{v}_n^5 \ , \hat{V}_{2,2} \ (\hat{v}_n)}{2 \ \beta_{2,2}^2 \{ \left(\acute{e}_3 \ \hat{\beta}_{+,3} \ (Cn) + \acute{e}_3 \ \hat{\beta}_{-,3} \ (\hat{C}_n) \right)^2 + 3\hat{V}_{3,3} \ (\hat{C}_n) \}} \ \dots \text{(2-28)}$$

Step 3: finding the basic bandwidth ($\hat{h}_{CCT}$) (Calonico et al. 2014) according to the following formula:

$$\hat{h}_{CCT} = \begin{pmatrix} 1/5 \\ 0 \ , 1 \ , 0 \end{pmatrix} . \ n^{-1/5} \ \dots \textbf{(2-29)}$$

$$\hat{C}_{0,1,0} = \frac{n \ \hat{v}_n^5 \ , \hat{V}_{0,1} \ (\hat{v}_n)}{4 \ \beta_{0,1}^2 \{ \left(\acute{e}_2 \ \hat{\beta}_{+,2} \ (\hat{b}_{CCT}) + \acute{e}_2 \ \hat{\beta}_{-,2} \ (\hat{b}_{CCT}) \right)^2 + 3\hat{V}_{3,3} \ (\hat{C}_n) \}} \dots \textbf{(2-30)}$$

2-13 Variable Selection

It means choosing the variables to be included in the model to be estimated. It is a method widely using for multiple regression and classification for selecting subsets of variables, in which subsets of the original set of variables are selected to obtain the smallest subset . It can be used to model the problem.

The purpose of the variable selection process is to obtain a clear concept of the data and the variables by providing information about the important variables as well as their relationship with each other.

There are two types of variable selection methods: the traditional methods and the regularization methods. Examples of traditional methods include incremental selection (Efroymson, 1960), Akaike Information Criteria (AIC) (Akaike, 1973) and Bayesian Information Criteria (BIC) (Schwartz, 1978). We note the instability and great variance in the traditional methods if compared to the methods of regularization. Examples of regularization methods are lasso (Tibshirani, 1996), Elastic net (Zou and Hastie, 2005) and Adaptive Lasso (AL), (Zou, 2006), MCB (Zhang, 2010), and others.

These methods have high stability compared to the traditional methods because parameter estimation and selection of variables are performed simultaneously. These methods depend on its application on the data and the purpose of the

study.

The regularization methods used in this study are briefly presented as follows:-

2-13-1 Adaptive Lasso Regression Model.

Zou (2006) showed that the lasso estimator is inconsistent in the selection of variables and to overcome this limitation he proposed the Adaptive Lasso (AL) method (Zou, 2006).. This method puts elastic weights in a penalizes function (L_1-norm) so that the amount of shrinkage applied to the estimators of the regression coefficient different, unlike the lasso method, it imposes the same penalizes on the estimators of all the regression parameters. The estimator of AL is obtained as follows:

$$\hat{\beta}_{AL} = arg \min_{\beta} \sum_{i=1}^{N}(Y_i - X\beta)^2 - \lambda \sum_{j=1}^{p} W_j|\beta_j|, \quad \lambda \geq 0 \, ... \quad (2-31)$$

Where : $\lambda\sum_{j=1}^{p} W_j|\beta_j|$, is called adaptive lasso penalty. That bias increases and variance decreases when λ increases. The parameter (λ) is adjusted through k-fold validation.

Adaptive lasso is developed for the purpose of maximizing the selection of the correct variable to solve problems of estimating low and high dimensions. This makes it ideal for selecting covariates in RDD and other causal inference procedures in which variable modification is appropriate. AL differs from other Lasso varieties, by including a set of weights as shown in the equation (2-32) so that the weights are obtained from estimates (OLS) of the coefficients (Anastasopoulos, 2019) :

$$w_j = \frac{1}{|\beta_j|^\gamma} \quad ... \quad (2-32)$$

where; β_j: Represents the estimated coefficients in OLS method.

γ: The tuning parameter represents, where $\gamma > 0$.

γ: tuned parameter by (Zou, 2006) using the cross - verification method, and it was tuned to (0.5, 1, 2) when performing the simulations.

2-13-2 Adaptive Lasso with RDD Model

Adaptive lasso with the estimated value of (λ) becomes attractive for causal inference, because it possesses the characteristics of an Oracle, as it works continuously to select a correct subset of variables from the sum of the large, so that this group is characterized by its approximate unbiasedness and as a natural selection, this is good if the Lasso model is used as a variable selection and not a shrinkage (Zou, 2006). Most of the time this will be true in covariate adjustment of local average treatment effect (LATE) in the RDD model and in most causal inference procedures. Bloniarz et al. (2016) and Wager et al. (2016) indicated that the estimation process takes place in two steps, where the lasso is used as a variable selection and final parameter values are estimated using OLS allows us to obtain (the best linear unbiased estimate, (BLUE)) in addition to the standard errors that fit with the model for easy interpretation.

(Bloniarz et al., 2016), have argued that adaptive lasso when adjusting the covariates local average treatment effect (LATE) improves the accuracy of the estimates in addition to using it as a way of selecting the "principled" model to avoid some pitfalls in order to identify treatment effects for RDD. They recommended a process that includes four steps for estimating the effect of treatment in RDD when inserting covariates based on a series of simulations that they performed. Where these steps will be mentioned in paragraph (2-14) .

2-14 Estimation (Local Average Treatment Effect(LATE) According to the Adaptive Lasso Model in RDD

Step 1: Selection of the Covariate Variable Researcher before Performing Treatment: In estimating RDD, the purpose of including pre-treatment covariates, is to obtain more accurate treatment effect estimates (Bloniarz , et al., 2016); Calonico et al., 2018). The most important part of the accuracy depends mainly on the bandwidth, or on the low variance in the model, and it may be due to both. Researchers may face some questions about the covariates variables in how to include them in the model before performing the treatment, or when choosing the optimal bandwidth, that making these decisions may lead to important consequences when estimating the efficacy of the local average treatment effect (LATE) in the final stages, especially when the covariates that included are closely related to the effect variable and in the case of local linear regression, in the case of the sample size small which is common in estimating RDD. (Bloniarz , et al., 2016).   Making preliminary decisions regarding the covariates variables that shall be included before performing a treatment, it shall always be based on expert judgment and the researcher's expectations they are closely related to the problem at hand.

$$Y_i = \alpha + \tau T_i + \delta_1 F_i + \delta_2(F_i.T_i) + XB + \epsilon_i \ldots (2-33)$$

Where the model (2-33) represents the estimated local linear regression, through which the treatment effect is estimated $\hat{\tau}$ where the observations are

in some neighborhood. The α value of the forcing variable lies around the cutoff point, such that i $\in F_\alpha \pm$ c.

Step 2: An Adaptive Lasso Regulation Application: After choosing the model (2-33), the appropriate model in which a set of the covariates we denote by the symbol $X^*$is chosen from the original set of covariates X which leads to the reduction of the variance from Local average treatment effect [var $(\hat{\tau})$]. In the case of obtaining the lowest (MSE) when using local linear regression model we can achieve the lowest var $(\hat{\tau})$. In the event that X^s is a subset of the covariates from the original set X, we aspire to choose $X^s \subseteq X$ such that :

$$\text{Var} (\hat{\tau}| X^s) \leq \text{Var} (\hat{\tau}| X) \dots \quad (2\text{-}34)$$

We aspire to obtain the minimum for the (SRD) when estimating the parameters of the model as shown in the equation (2-35).

$$\arg\min_{\Theta} \sum_{i=1}^{N} (Y_i - [\alpha + \tau T_i + \delta_1 F_i + \delta_2(F_i.T_i) + X^s\beta])^2 \quad \dots \quad (2-35)$$

Where: $\Theta = (\tau, \delta_1, \delta_2, \beta)$ represents the vector of the estimated coefficients in SRD. The selection of subgroups is an unstable method because any slight change in the data can lead to large changes in the chosen model. Also, it is not computationally possible in light of a large number of explanatory variables. The researchers proposed many punitive methods that shrink the estimators of some regression coefficients and make others equal to zero. Then, these methods can estimate and choose the regression model variables and treat the problem of linear multiplicity. Researchers are primarily interested in choosing a model that the sample size is small. Adaptive lasso model is a natural choice and it is a modified version of lasso. It has the advantage of having Oracle properties

(Anastasopoulos,2019). Oracle's characteristics ensure that the adaptive lasso model estimates are approximated in terms of bias and in terms of efficiency when we are working on defining combinations of non-zero variables simultaneously, i.e. choosing the right combination. When excluding covariates in RDD adaptive lasso modification is required to perform RDD, the treatment effect, forcing variable, and variables included in the kernel selected are not penalized. But a penalty is required for any additional covariates. It is easy to estimate a modified version of the adaptive lasso to exclude covariates in the RDD design where the weights of the parameters are set to zero. The weights of the added covariates repesent the weights for the adaptive lasso model. Model (2-36) represents the full initial model to be estimated: (Anastasopoulos, 2019)

$$arg \min_{\theta} \sum_{i=1}^{N} [Y_i - (\alpha + \tau T_i + \delta_1 F_i + \delta_2(F_i.T_i) + X\beta)]^2 + \lambda \left[\sum_{j=3}^{p} W_j|\beta_j| \right] \dots$$

..(2-36)

Step 3: Automatic Model Selection

After estimating the parameters from the adaptive lasso model in Equation (2-37), the covariates variables are excluded before calculating the optimum bandwidth . They are shrunk to zero from the model using the best value for λ^{RDD} and the optimum weights, the resulting model is used to estimate the optimum bandwidth (Anastasopoulos, 2019). The choice of optimal MSE bandwidth and point estimators varies when covariate adjustment is used. A bias correction techniques and standard error constructions also differ.

Thus the robust treatment effects will be as follows:

$$E(Y_i|T_i, F_i, X^s) = \alpha + \tau\, T_i + \delta_1 F_i + \delta_2(F_i . T_i) + X^s\, \beta \quad \dots \text{(2-37)}$$

Where $E(Y_i|T_i, F_i, X^s)$ is Conditional expectation function.

$X^S \subseteq X$ is the set of covariates chosen by the adaptive lasso. Since optimal bandwidth selection algorithms such as Imbens Kalyanaraman use cross-validated MSE as criteria for selecting the "best" bandwidth. The estimated bandwidth values chosen after manipulating the covariates with adaptive lasso shall be less than or equal to the estimated bandwidth values of the MSE model when using the full model in step 1.

In the estimation of (RDD) we can combine this method with the covariates before making the choice of bandwidth, which will alter the optimal bandwidth chosen, or after bandwidth selection if the bandwidth is set to a predetermined value (e.g. 1%, 5% etc) .

Step 4: CCT Robust's Regular Estimation of Model selection

When robust CCT estimation is used, definitive estimates of treatment effect can be obtained (Calonico et al. 2018). After performing steps 1 to 3 involving selecting the optimal conditional expectation function (CEF), $E\,(Y_i|T_i, F_i, X^s)$, using local linear regression LLR , as well as estimating the optimal bandwidth h_0^* based on the conditional expectation function CEF .(Anastasopoulos, 2019).

2-15 Minimax Concave Penalty (MCP) Variable Selection in the RDD Model.

The Minimax Concave Penalty (MCP) is another alternative to get less biased regression coefficients in sparse models. Zhang (2010) proposed the MCP method, that estimates and selects linear regression variables simultaneously using the MCP penalty function, overcomes the Lasso method in terms of its inconsistency in the selection of variables. The MCP estimator is obtained by the following formula: (Choon, 2012).

$$\hat{\beta}_j^{MCP} = arg \min_p \|Y - X\beta\|^2 + \sum_{j=1}^{p} P_{\lambda,\gamma}^{MCP} \ldots (2-38)$$

Where: $\sum_{j=1}^{p} P_{\lambda,\gamma}^{MCP}$ the MCP penalty function.

The MCP function takes the following form:

$$p_\lambda(|\beta|) = \begin{cases} \lambda\left(|\beta| - \frac{|\beta|^2}{2\lambda\gamma}\right), & |\beta| < \lambda\gamma \\ \frac{\lambda^2\gamma}{2}, & |\beta| \geq \lambda\gamma \end{cases} \ldots (2-39)$$

Where : $\gamma > 1$ (Breheny, 2016).

Many concave penalties are based on λ , as well as the inclusion of a fine-tuning coefficient (γ) that controls concavity to penalty (that is, how quickly the penalty is reduced). MCP possess an interval of λ values over which all the estimates are flat – over this region, the estimates are the same as those of ordinary least squares regression.

The Adaptive Lasso and MCP methods diffir from Lasso in that they allow estimated parameters to reach larg values more quickly than Lasso, since all of these methods shrink most of the coefficients towards zero, but these two methods (Adaptive Lasso and MCP) operate on apply less shrinkage to non-zero coefficients; this indicates less bias.

By doing the same previous steps for the adaptive lasso model with RDD, but by combining the MCP penalty with the RDD model, we get the following model :

$$arg\ min_{\theta}\ \sum_{i=1}^{N}[Y_i - (\alpha + \tau T_i + \delta_1 F_i + \delta_2\ (F_i, T_i) + X\beta)]^2 + \sum_{j=3}^{p} P_{\lambda,\gamma}^{MCP} \ ... \ (2\text{-}40)$$

2-16   Elasticnet Variable Selection in the RDD Model.

Zou and Hastie (2005) suggested the Elastic Net. It is a regulated regression method as it is known in linear regression models because it is a combination of the lasso model penalty L_1 and the Ridge model penalty L_2. This method overcomes the limitations of the lasso method which uses a penal function based on the following equation:

$$\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$$

When (p > n) in the case of a set of highly correlated variables, lasso selects one of these variables and discards the other.

To get around these limitations, this method adds a square fraction to the penalty ($\|\beta\|^2$), its represents ridge regression when used alone which reduces the sum of the squares remaining for the criterion (L_2).

Ridge regression represents a type of common of regular linear regression introduced by (Hoerl and Kennard, 1970b), it´s one of the first contraction methodsfor linear regression models that added the L_2 penalty. To the problem of OLS miniaturization .

The ridge estimation can be achieved through the following equation:

$$\hat{\beta}_{Ridge} = arg \min_{\beta} \sum_{i=1}^{N} (Y_i - X\beta)^2 - \lambda \sum_{j=1}^{P} \beta_j^2 \ \dots (2-41)$$

Where, $\lambda \sum_{j=1}^{P} \beta_j^2$ is called the ridge regression penalty.

We cannot use Ridge regression for selecting the variable because it does not shrink the coefficients towards zero (Tibchirani et al., 2015).

The Elastic Net is an alternative approach for variable selection process. The estimates are determined by the Elastic Net method through the following formula: (Zou, 2005).

$$\hat{\beta} = arg \min_{\beta}(\|Y - X\beta\|^2 + \lambda_2\|\beta\|^2 + \lambda_1\|\beta\|_1) \ \dots (\mathbf{2-42})$$

Where:

λ_1 : The tuning parameter represents the Lasso penalty (L_1).

λ_2 : The tuning parameter represents the Ridge regression penalty (L_2).

When we use Elastic Net regression, "α" represents a ratio $\lambda_1 : \lambda_2$.

λ_1 represents the value of "α" for the portion of the Lasso regression , λ_2 represents the value of (α) for the portion of the regression of Ridge. When the ratio is equal to zero, it acts as Ridge regression, and when the ratio is one, it acts as a Lasso regression. Where any value between 0 and 1 is a combination of Ridge regression and Lasso regression.

Whereas :

$$P_\alpha\ (\beta) = \frac{(1-\alpha)}{2} |\beta|_2^2 + \alpha|\beta_j|_1 = \sum_{i=1}^{P} \left(\frac{(1-\alpha)}{2} \beta_j^2 + \alpha|\beta_j| \right) \dots (2-43)$$

The penalty term $P_\alpha(\beta)$ is combination between the L_1 norm of β and the squared L_2 norm of β. Then, it is called an Elastic Net penalty. (Friedman,et al., 2010).

By combining Elastic Net penalty of variable selection with RDD model after applying the previously mentioned steps we obtain the following formula : -

$$arg\ \min_\theta \sum_{i=1}^{N}[Y_i - (\alpha + \tau T_i + \delta_1 F_i + \delta_2(F_i.T_i) + X\beta]^2 + \sum_{i=3}^{P} \left(\frac{(1-\alpha)}{2} \beta_j^2 + \alpha|\beta_j| \right) \dots (2\text{-}44).$$

Where $0 \le \alpha \le 1$ is a penalty weight .

CHAPTER THREE

SIMULATION STUDY

AND

APPLICATION

3. Application side

3.1 Simulation Study:

In this section, a Monte Carlo simulation study is designed to evaluate the performance of our new proposed methods, RDD combined with both MCP and Elastic Net with different bandwidths (IK, CV and CCT) and compared with existing method, namely the adaptive lasso - RDD. The performances of these methods are evaluated based on (MSE) criterion. A good method is one that has a smallest value of MSE. Programming (R) are used for the data analysis.

The main model for simulation study is given as follows (Anastasopoulos, 2019):

$$Y_i = \alpha + \tau T_i + \delta_1 F_i + \delta_2 (F_i . T_i) + XB + \epsilon_i \quad ....(3\text{-}1)$$

The steps of simulation study is presented as following:

Step1: The sample of study is generated in different size (50, 100, 250, 500) with p = 50 covariates include (s) non-zero variables and (p - s) zero variables.

Step2: the correlation between variables is taken as (r = 0.75).

Step3: Two types of variables were generated, where the first type is a treatment variable that generated according to the uniform distribution where a = -1 is the lower bound and b = 1 is the upper bound when the cutoff point is (0) with parameter value (10, 2) (processing variables and processing parameters). That is, the uniform distribution depends on the cut-off point, so the lower and upper bounds of the distribution depend on one dimension on both sides of the cut-off point . The second type data $\acute{x}\beta$ were generated according to a multivariate normal distribution defined by some mean parameters ($\mu = (\mu_1, \mu_2, ..., \mu_p)$) and the covariance matrix Σ. (Szakonyi, David. 2018).

Step4: The random error term (ε) was generated according to the standard normal distribution N (0,1), and the data were generated based on the equation (3-1) and repeat each experiment (IT=1000) for all of the simulation experiments for the purpose of obtaining stable results .

Step 5: Reduce the number of variable by using variable selection method such that Elastic net and MCP method.

Step 6: Apply the bandwidth by using (CCT, CV and IK).

Step 7: Apply the RDD model depend on steps 1 to 6 .

Step 8: Finally, compute the MSE for all the methods for comparison.

Example 1: Samples size (n=50,100,250,500) , number of variables (p =15) ,(s=5),(p-s=10) and $\rho = 0.75$.

In this simulation, it has been suggested that the real vector of the coefficient $\beta = (\underbrace{\beta_1, \beta_2, ..., \beta_s}_{S}, \underbrace{0,0, ...,0}_{P-S})$, The β is selected as $\beta = (\underbrace{0.5, 1, 1.5, 2, 3}_{S}, \underbrace{0,0, ...,0}_{P-S})$ and cut off point (0 , 0.5 , 2) .

Table (3-1): MSE values for methods of study for n=50, p= 15, s=5 and ρ=0.75.

methods	Cut off point =0.0			methods	Cut off point =0.5			methods	Cut off point =2		
	IK	CCT	CV		IK	CCT	CV		IK	CCT	CV
Ad lasso	0.0672	0.0530	0.0597	Ad lasso	0.0689	0.0563	0.0583	Ad lasso	0.0869	0.0625	0.0720
Elistic.Net	0.0587	0.0484	0.0591	Elistic.Net	0.0643	0.0549	0.0581	Elistic.Net	0.0799	0.0561	0.0657
MCP	0.0576	0.0464	0.0576	MCP	0.0627	0.0463	0.0538	MCP	0.0789	0.0540	0.0606

From Table (3-1) with n = 50 , p= 15, s=5 and ρ =0.75 , we notice the superiority of our suggested method (MCP) over the Elastic Net and adaptive lasso through MSE values. In addition, it's clear to see that the best method of the bandwidth is CCT at all of cut- off point. Figure (3-1) illustrates this through MSE.

Figure (3-1): Shows three cases of cut-off point (0.0, 0.5 and 2) for the preference of our suggested methods over the adaptive lasso method in the RDD model. according to the (MSE) criterion. Also, the figure at the right, shows the effect of the treatment used to change the value of the dependent variable be fore and after the cut-off, it is noted that the number of important variables appeared far from zero, as is true coefficients that assumed by the simulation and the figure that shows the features in the methods Elastic Net and MCP.

Figure (3-1): MSE Values, Cut off Point and Coefficients for Methods of study   for n=50, p= 15, s=5 and ρ =0.75

Table(3- 2): MSE values for methods of study for n=100, p= 15, s=5 and ρ =0.75

Methods	Cut off point =0.0			Methods	Cut off point =0.5			Methods	Cut off point =2		
	IK	CCT	CV		IK	CCT	CV		IK	CCT	CV
Ad lasso	0.0661	0.0520	0.0586	Ad lasso	0.0687	0.0555	0.0583	Ad lasso	0.0863	0.0624	0.0710
Elistic.Net	0.0575	0.0474	0.0564	Elistic.Net	0.0642	0.0548	0.0581	Elistic.Net	0.0795	0.0567	0.0647
MCP	0.0566	0.0454	0.0576	MCP	0.0624	0.0444	0.0538	MCP	0.0777	0.0520	0.0546

From Table (3- 2) with n = 100, p= 15, s =5 and ρ =0.75 , it is noticed that superiority of our suggested method (MCP) over the Elastic Net and adaptive lasso through MSE values. In addition, it's clear to see that the best method of the bandwidth is CCT at all of cut- off point.   Figure (3-2) illustrates this through MSE.

Figure (3-2): Shows three cases of cut-off point (0.0, 0.5 and 2) for the preference of our suggested methods over the adaptive lasso method in the RDD model. according to the (MSE) criterion. Also, the figure at the right, shows the effect of the treatment used to change the value of the dependent variable before and after the cut-off, it is noted that the number of important variables appeared far from zero, as is true coefficients that assumed by the simulation and the figure that shows the features in the methods Elastic Net and MCP.
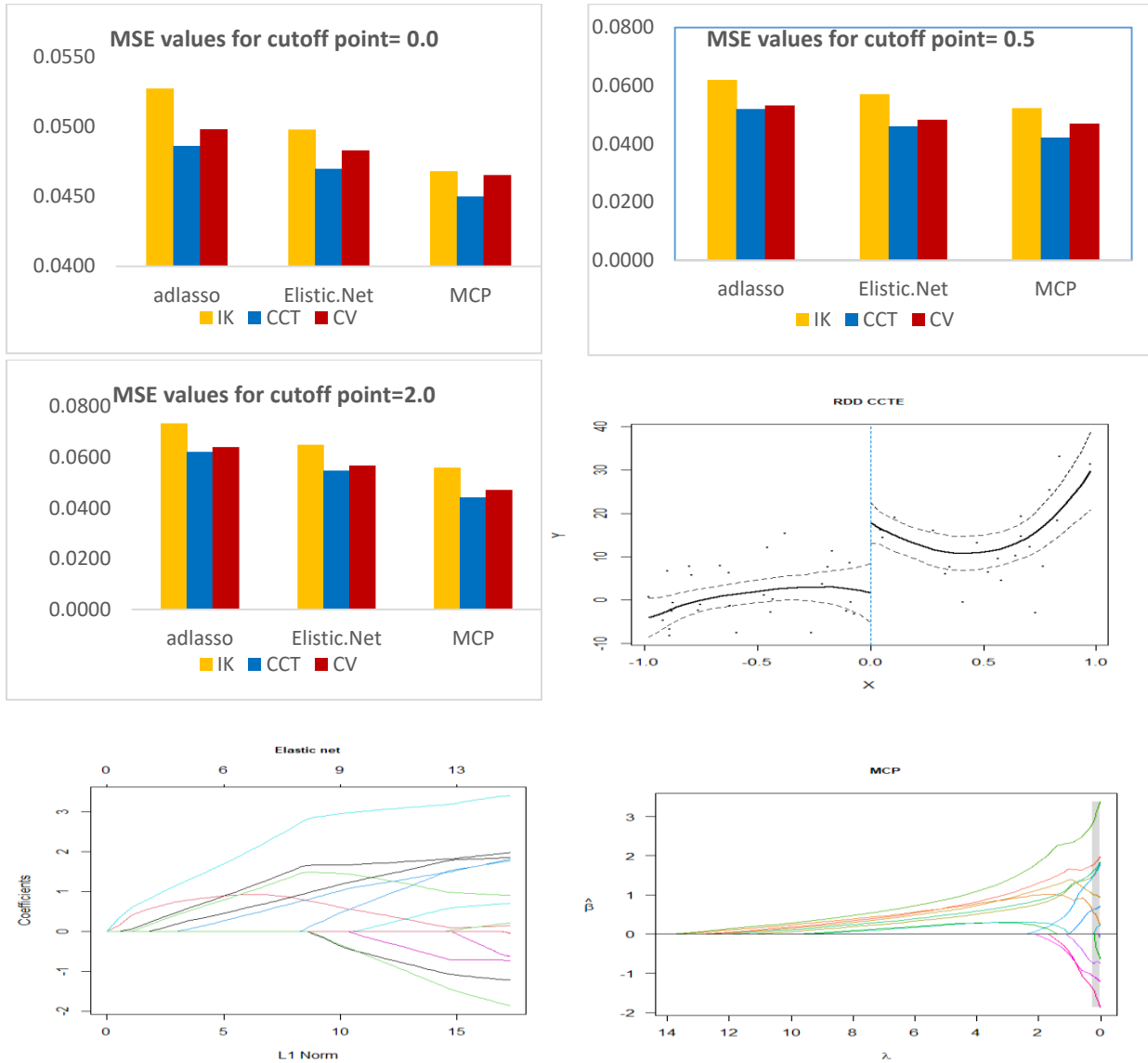

Figure (3-2): MSE Values, Cut off Point and Coefficients for Methods of Study for n=100 , p= 15, s = 5 and ρ =0.75 .

Table (3-3): MSE values for methods of study for n=250, p=15, s=5 and ρ =0.75

Methods	Cut off point =0.0			Methods	Cut off point =0.5			Methods	Cut off point =2		
	IK	CCT	CV		IK	CCT	CV		IK	CCT	CV
Ad lasso	0.0642	0.0510	0.0640	Ad lasso	0.0686	0.0544	0.0581	Ad lasso	0.0860	0.0613	0.0712
Elistic.Net	0.0544	0.0464	0.0539	Elistic.Net	0.0641	0.0542	0.0580	Elistic.Net	0.0781	0.0534	0.0643
MCP	0.0538	0.0444	0.0532	MCP	0.0622	0.0442	0.0527	MCP	0.0762	0.0510	0.0544

From Table (3-3) with n = 250 , p= 15, s=5 and ρ =0.75 , we notice the superiority of our suggested method (MCP) over the Elastic Net and adaptive lasso through MSE values. In addition, it's clear to see that the best method of the bandwidth is CCT at all of cut- off point.   Figure (3-3) illustrates this through MSE.

Figure (3-3): Shows three cases of cut-off point (0.0, 0.5 and 2) for the preference of our suggested methods over the adaptive lasso method in the RDD model. according to the (MSE) criterion. Also, the figure at the right, shows the effect of the treatment used to change the value of the dependent variable before and after the cut-off, it is noted that the number of important variables appeared far from zero, as is true coefficients that assumed by the simulation and the figure that shows the features in the methods Elastic Net and MCP.
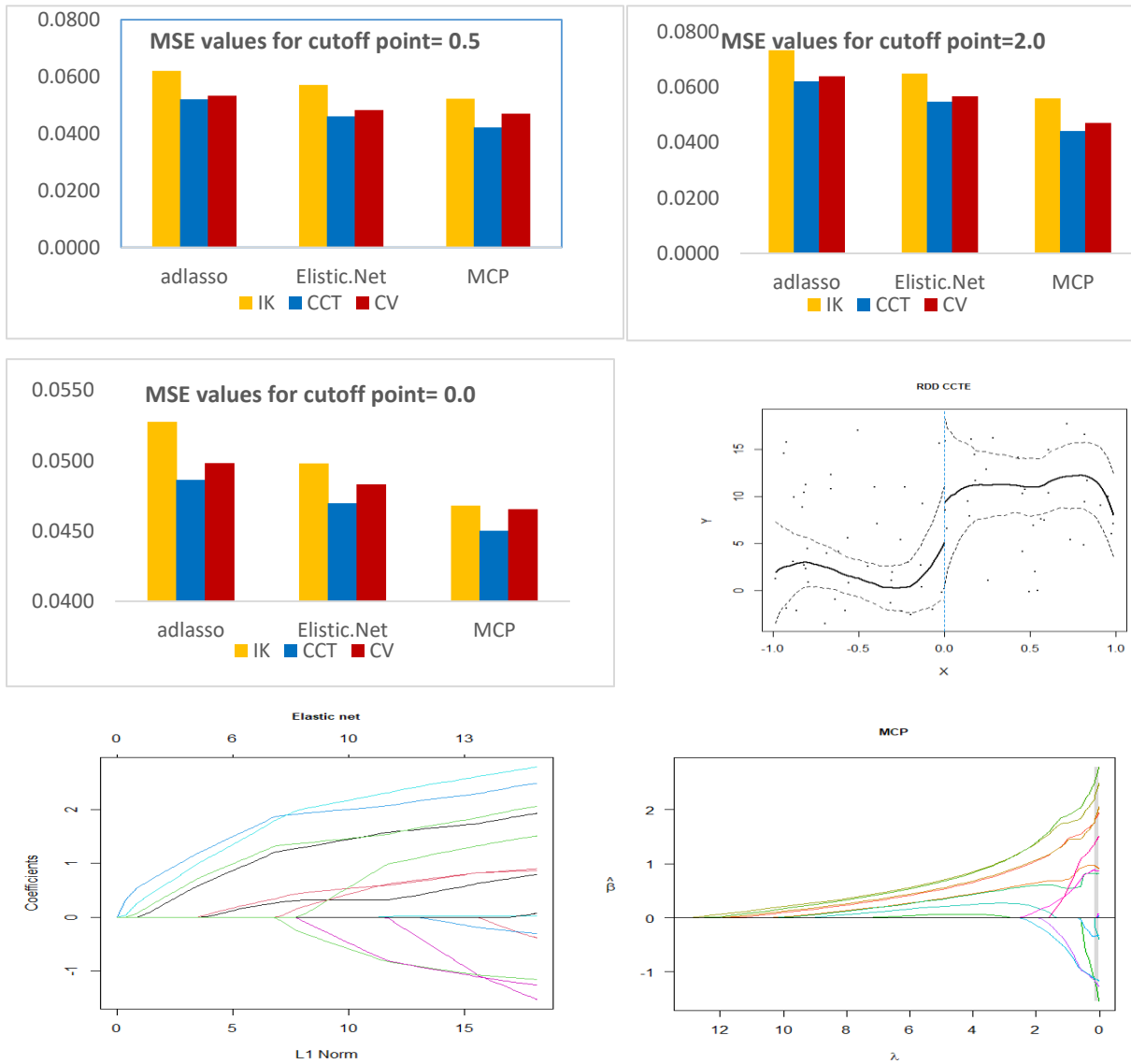

Figure (3-3): MSE Values, Cut off Point and Coefficients for Methods of Study for n=250 , p= 15, s=5 and ρ =0.75

Table (3-4): MSE values for methods of study for n=500, p =15, s =5 and ρ=0.75

Methods	Cut off point =0.0			Methods	Cut off point =0.5			Methods	Cut off point =2		
	IK	CCT	CV		IK	CCT	CV		IK	CCT	CV
Ad lasso	0.0561	0.0541	0.0556	Ad lasso	0.0502	0.0455	0.0470	Ad lasso	0.0639	0.0516	0.0528
Elistic.Net	0.0461	0.0447	0.0457	Elistic.Net	0.0491	0.0449	0.0469	Elistic.Net	0.0566	0.0447	0.0466
MCP	0.0436	0.0422	0.0431	MCP	0.0461	0.0440	0.0459	MCP	0.0546	0.0407	0.0446

From Table (3- 4) with n = 500 , p = 15, s = 5 and ρ =0.75 , we notice the superiority of our suggested method (MCP) over the Elastic Net and adaptive lasso through MSE values. In addition, it's clear to see that the best method of the bandwidth is CCT at all of cut- off point.   Figure (3-4) illustrates this through MSE.

Figure (3-4): Shows three cases of cut-off point (0.0, 0.5 and 2) for the preference of our suggested methods over the adaptive lasso method in the RDD model. according to the (MSE) criterion. Also, the figure at the right, shows the effect of the treatment used to change the value of the dependent variable before and after the cut-off, it is noted that the number of important variables appeared far from zero, as is true coefficients that assumed by the simulation and the figure that shows the features in the methods Elastic Net and MCP.
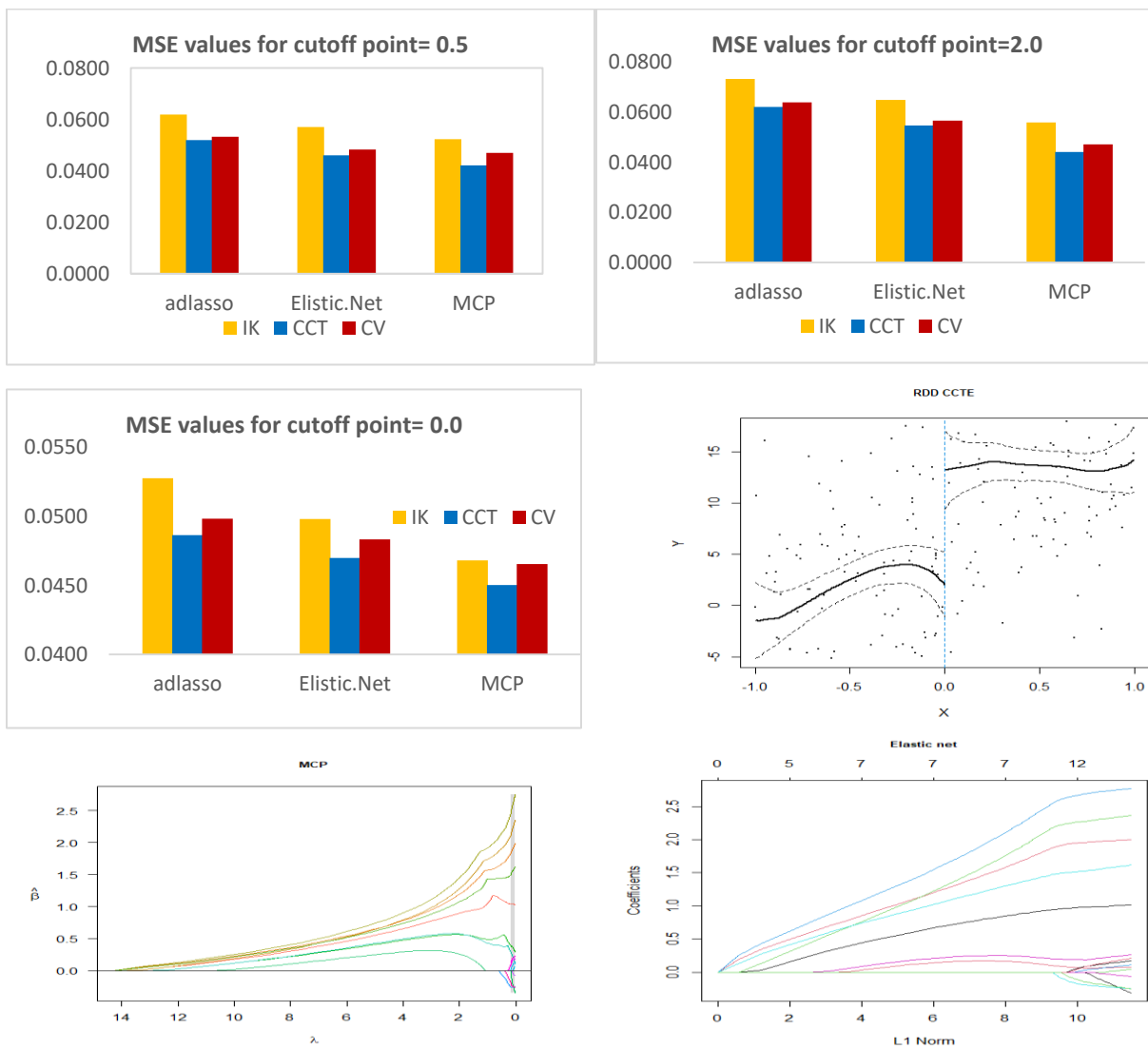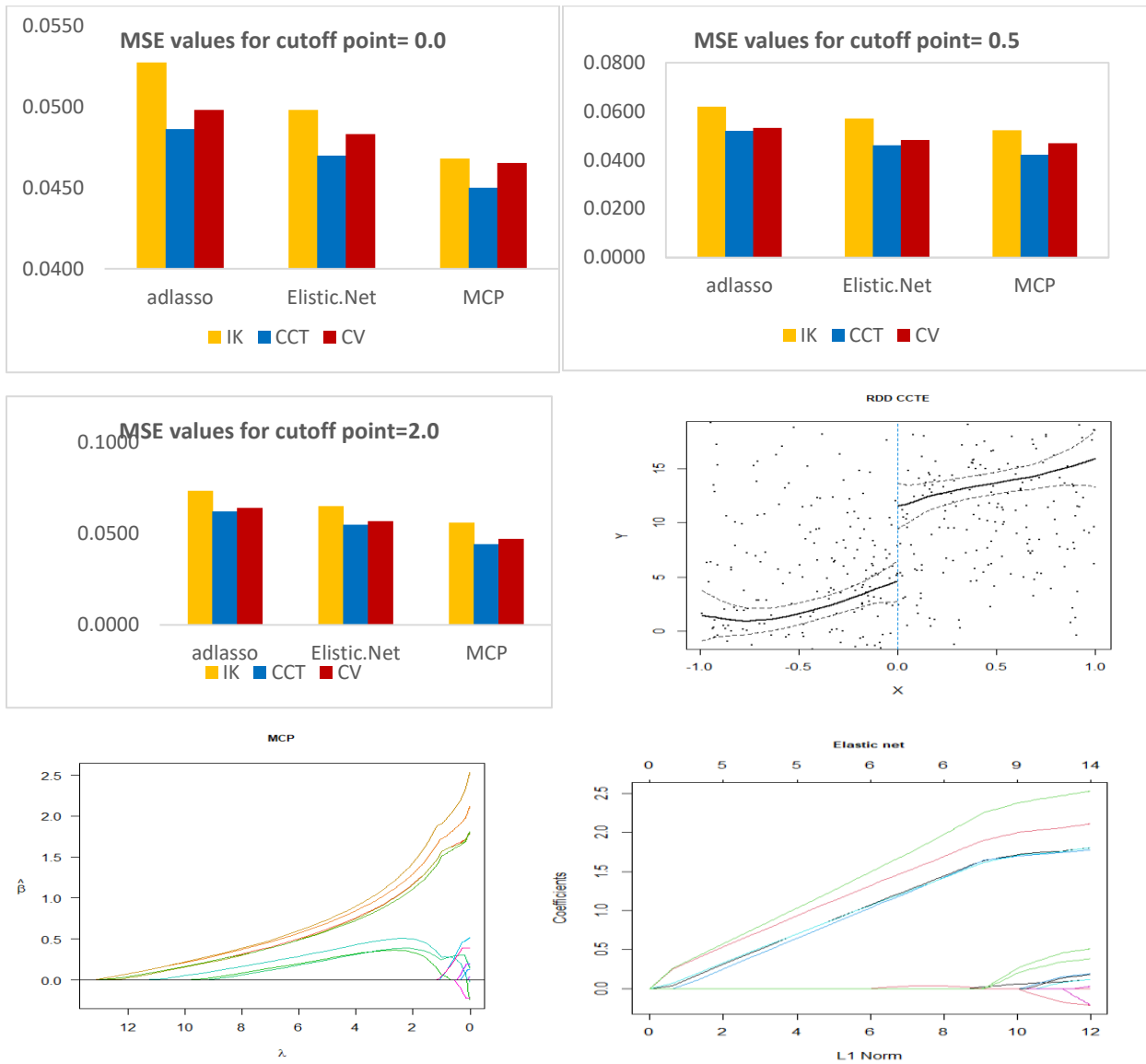

Figure (3-4): MSE Values, Cut off Point and Coefficients for Methods of Study for n=500 , p= 15, s =5 and ρ =0.75

Example 2:   Samples size (n=50,100,250,500) , number of variables (p =25)

,(s=10),(p-s=15) and $\rho = 0.75$, and β is selected as $\beta = (\underbrace{1.5,1.5, \dots , 1.5}_{S}, \underbrace{0,0, \dots ,0}_{P-S})$

and cut off point (0 , 0.5 , 2).

Table (3-5): MSE values for methods of study for n=50, p = 25, s =10 and ρ =0.75

Methods	Cut off point =0.0			Methods	Cut off point =0.5			Methods	Cut off point =2		
	IK	CCT	CV		IK	CCT	CV		IK	CCT	CV
Ad lasso	0.2733	0.2468	0.2589	Ad lasso	0.2063	0.1525	0.1673	Ad lasso	0.2910	0.2525	0.2525
Elistic.Net	0.2183	0.2033	0.2075	Elistic.Net	0.1611	0.1174	0.1235	Elistic.Net	0.2202	0.1758	0.1857
MCP	0.1740	0.1532	0.1662	MCP	0.1413	0.1009	0.1127	MCP	0.1462	0.1178	0.1273

From Table (3- 5) with n = 50 , p= 25, s=10 and ρ =0.75 , we notice the superiority of our suggested method (MCP) over the Elastic Net and adaptive lasso through MSE values. In addition, it's clear to see that the best method of the bandwidth is CCT at all of cut- off point.   Figure (3- 5) illustrates this through MSE.

Figure (3-5): Shows three cases of cut-off point (0.0, 0.5 and 2) for the preference of our suggested methods over the adaptive lasso method in the RDD model. According to the (MSE) criterion. Also, the figure at the right, shows the effect of the treatment used to change the value of the dependent variable before and after the cut-off it is noted that the number of important variables appeared far from zero, as is true coefficients that assumed by the simulation and the figure that shows the features in the methods Elastic Net and MCP.

Figure (3- 5): MSE Values, Cut off Point and Coefficients for Methods of study for n=50 , p= 25, s=10 and ρ =0.75

Table (3-6): MSE values for methods of study for n=100, p= 25, s=10 and ρ 0.75

Methods	Cut off point =0.0			Methods	Cut off point =0.5			Methods	Cut off point =2		
	IK	CCT	CV		IK	CCT	CV		IK	CCT	CV
Ad lasso	0.1269	0.1150	0.1216	Ad lasso	0.1410	0.1062	0.1244	Ad lasso	0.2118	0.1549	0.1745
Elistic.Net	0.1057	0.0928	0.0947	Elistic.Net	0.1182	0.0966	0.1053	Elistic.Net	0.1510	0.1119	0.1313
MCP	0.0970	0.0848	0.0917	MCP	0.1082	0.0858	0.1007	MCP	0.1142	0.0818	0.1017

From Table (3- 6) with n = 100 , p= 25, s=10 and ρ =0.75 , we notice the superiority of our suggested method (MCP) over the Elastic Net and adaptive lasso through MSE values. In addition, it's clear to see that the best method of the bandwidth is CCT at all of cut- off point.   Figure (3-6) illustrates this through MSE.

Figure (3-6) Shows three cases of cut-off point (0.0, 0.5 and 2) for the preference of our suggested methods over the adaptive lasso method in the RDD model. according to the (MSE) criterion. Also, the figure at the right, shows the effect of the treatment used to change the value of the dependent variable before and after the cut-off, it is noted that the number of important variables appeared far from zero, as is true coefficients that assumed by the simulation and the figure that shows the features in the methods Elastic Net and MCP.
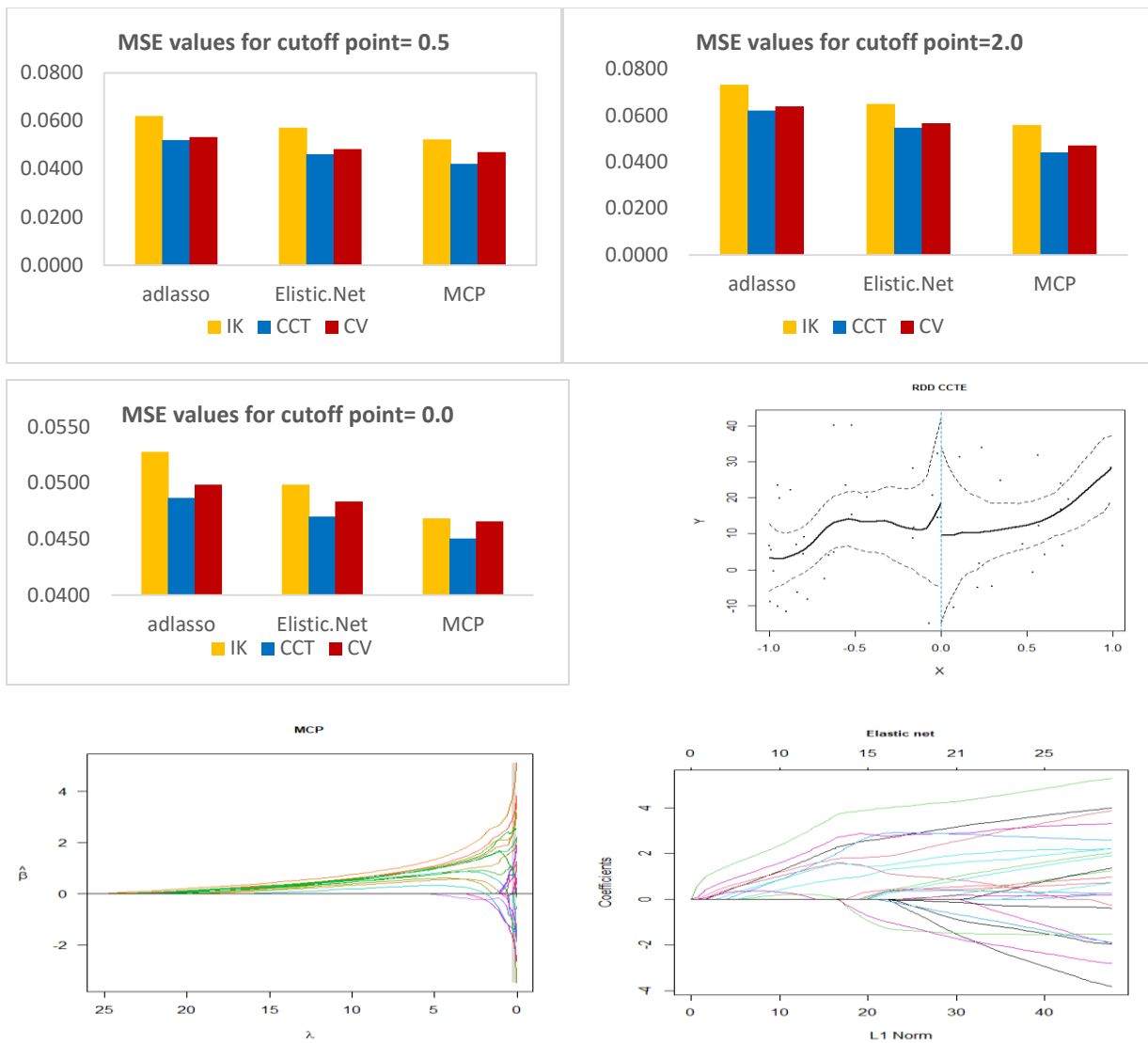

Figure (3-6): MSE Values, Cut off Point and Coefficients for Methods of Study for n=100 , p= 25, s=10 and ρ =0.75

Table (3-7): MSE values for methods of study for n=250, p = 25, s=10 and ρ =0.75

Methods	Cut off point =0.0			Methods	Cut off point =0.5			Methods	Cut off point =2		
	IK	CCT	CV		IK	CCT	CV		IK	CCT	CV
Ad lasso	0.0689	0.0674	0.0684	Ad lasso	0.1007	0.0751	0.0814	Ad lasso	0.1599	0.1336	0.1388
Elistic.Net	0.0649	0.0639	0.0648	Elistic.Net	0.0932	0.0695	0.0734	Elistic.Net	0.0974	0.0815	0.0846
MCP	0.0642	0.0634	0.0641	MCP	0.0886	0.0639	0.0699	MCP	0.0760	0.0635	0.0667

From Table (3-7) with n = 250 , p= 25, s=10 and ρ =0.75 , we notice the superiority of our suggested method (MCP) over the Elastic Net and adaptive lasso through MSE values. In addition, it's clear to see that the best method of the bandwidth is CCT at all of cut- off point.   Figure (3-7) illustrates this through MSE.

Figure (3-7) shows three cases of cut-off point (0.0, 0.5 and 2) for the preference of our suggested methods over the adaptive lasso method in the RDD model. according to the (MSE) criterion. Also, the figure at the right, shows the effect of the treatment used to change the value of the dependent variable before and after the cut-off, it is noted that the number of important variables appeared far from zero, as is true coefficients that assumed by the simulation and the figure that shows the features in the methods Elastic Net and MCP.
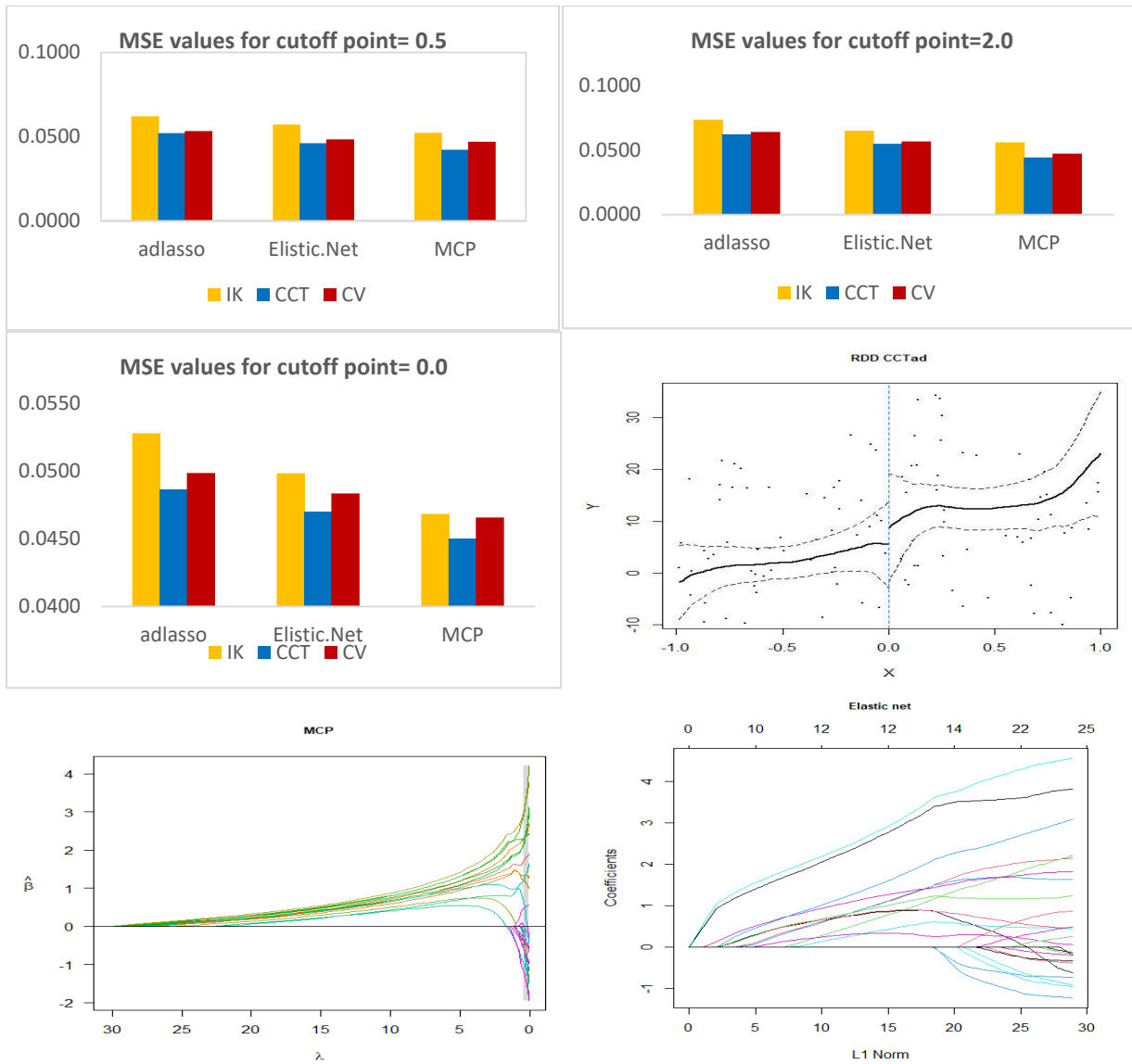

Figure (3-7): MSE Values, cut off Point and Coefficients for Methods of Study for n=250 , p= 25, s=10 and ρ =0.75

Table(3-8): MSE values for methods of study for n=500,p= 25, s=10 and ρ=0.75

Methods	Cut off point =0.0			Methods	Cut off point =0.5			Methods	Cut off point =2		
	IK	CCT	CV		IK	CCT	CV		IK	CCT	CV
Ad lasso	0.0471	0.0461	0.0466	Ad lasso	0.0594	0.0458	0.0485	Ad lasso	0.0626	0.0513	0.0596
Elistic.Net	0.0460	0.0453	0.0457	Elistic.Net	0.0524	0.0378	0.0435	Elistic.Net	0.0608	0.0467	0.0545
MCP	0.0453	0.0449	0.0451	MCP	0.0474	0.0327	0.0384	MCP	0.0475	0.0427	0.0444

From Table (3- 8) with n = 500, p= 25, s =10 and ρ =0.75 , we notice the superiority of our suggested method (MCP) over the Elastic Net and adaptive lasso through MSE values. In addition, it's clear to see that the best method of the bandwidth is CCT at all of cut- off point.   Figure (3-8) illustrates this through MSE.

Figure (3-8): Shows three cases of cut-off point (0.0, 0.5 and 2) for the preference of our suggested methods over the adaptive lasso method in the RDD model. according to the (MSE) criterion. Also, the figure at the right, shows the effect of the treatment used to change the value of the dependent variable before and after the cut-off, it is noted that the number of important variables appeared far from zero, as is true coefficients that assumed by the simulation and the figure that shows the features in the methods Elastic Net and MCP.
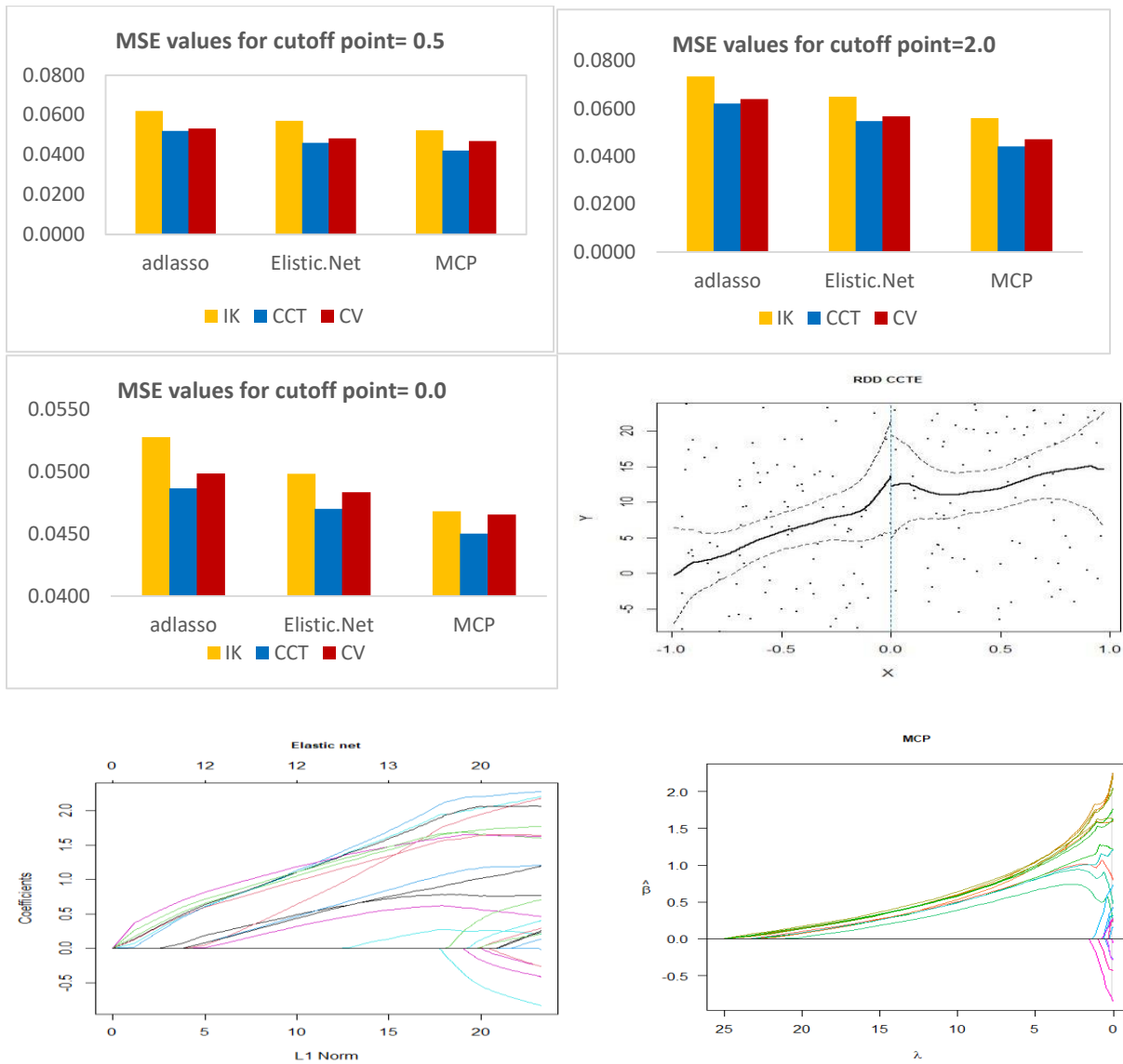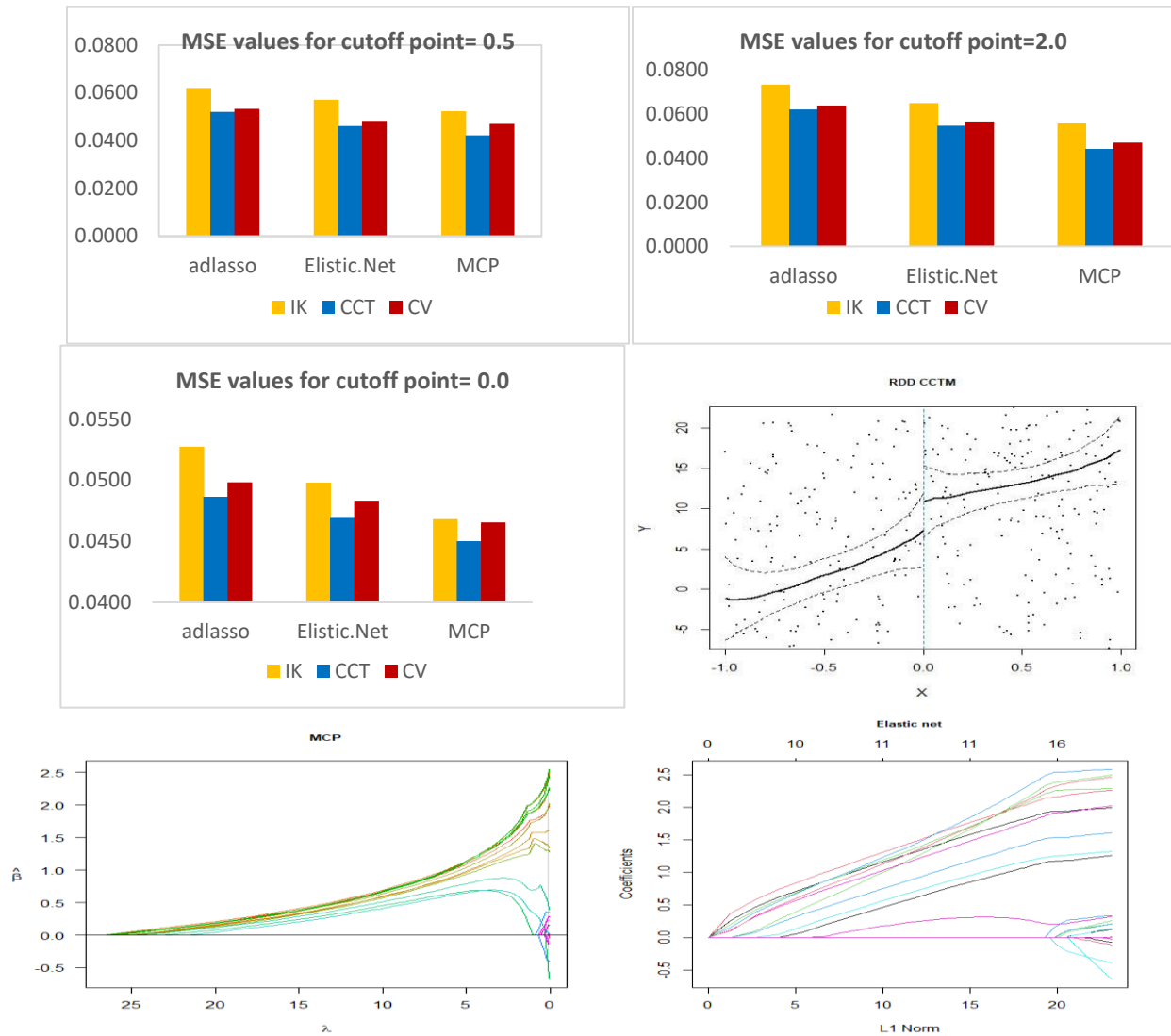

Figure (3-8): MSE Values, Cut off Point and Coefficients for Methods of Study for n=500 , p= 25, s=10 and ρ =0.75.

Example 3: Samples size (n=50,100,250,500) , number of variables (p =50) ,(s=20),(p-s=30) and $\rho = 0.75$, and β is selected as $\beta = (\underbrace{2,2,2\ldots,2}_{S}, \underbrace{0,0,\ldots,0}_{P-S})$ and cut off point (0 , 0.5 , 2).

Table(3-9): MSE values for methods of study   for n=50, p = 50, s=20 and ρ=0.75

Methods	Cut off point =0.0			Methods	Cut off point =0.5			Methods	Cut off point =2		
	IK	CCT	CV		IK	CCT	CV		IK	CCT	CV
Ad lasso	0.3492	0.2235	0.2766	Ad lasso	0.1674	0.1313	0.1436	Ad lasso	0.5102	0.4470	0.4731
Elistic.Net	0.2834	0.1991	0.2398	Elistic.Net	0.1488	0.1223	0.1364	Elistic.Net	0.3554	0.2902	0.3112
MCP	0.2307	0.1443	0.1974	MCP	0.1173	0.0858	0.0999	MCP	0.2421	0.1620	0.1837

From Table (3-9) with n = 50, p= 50, s=20 and ρ =0.75 , we notice the superiority of our suggested method (MCP) over the Elastic Net and adaptive lasso through MSE values. In addition, it's clear to see that the best method of the bandwidth is CCT at all of cut- off point.   Figure (3-9) illustrates this through MSE.

Figure (3-9): Shows three cases of cut-off point (0.0, 0.5 and 2) for the preference of our suggested methods over the adaptive lasso method in the RDD model. according to the (MSE) criterion. Also, the figure at the right, shows the effect of the treatment used to change the value of the dependent variable before and after the cut-off, it is noted that the number of important variables appeared far from zero, as is true coefficients that assumed by the simulation and the figure that shows the features in the methods Elastic Net and MCP.
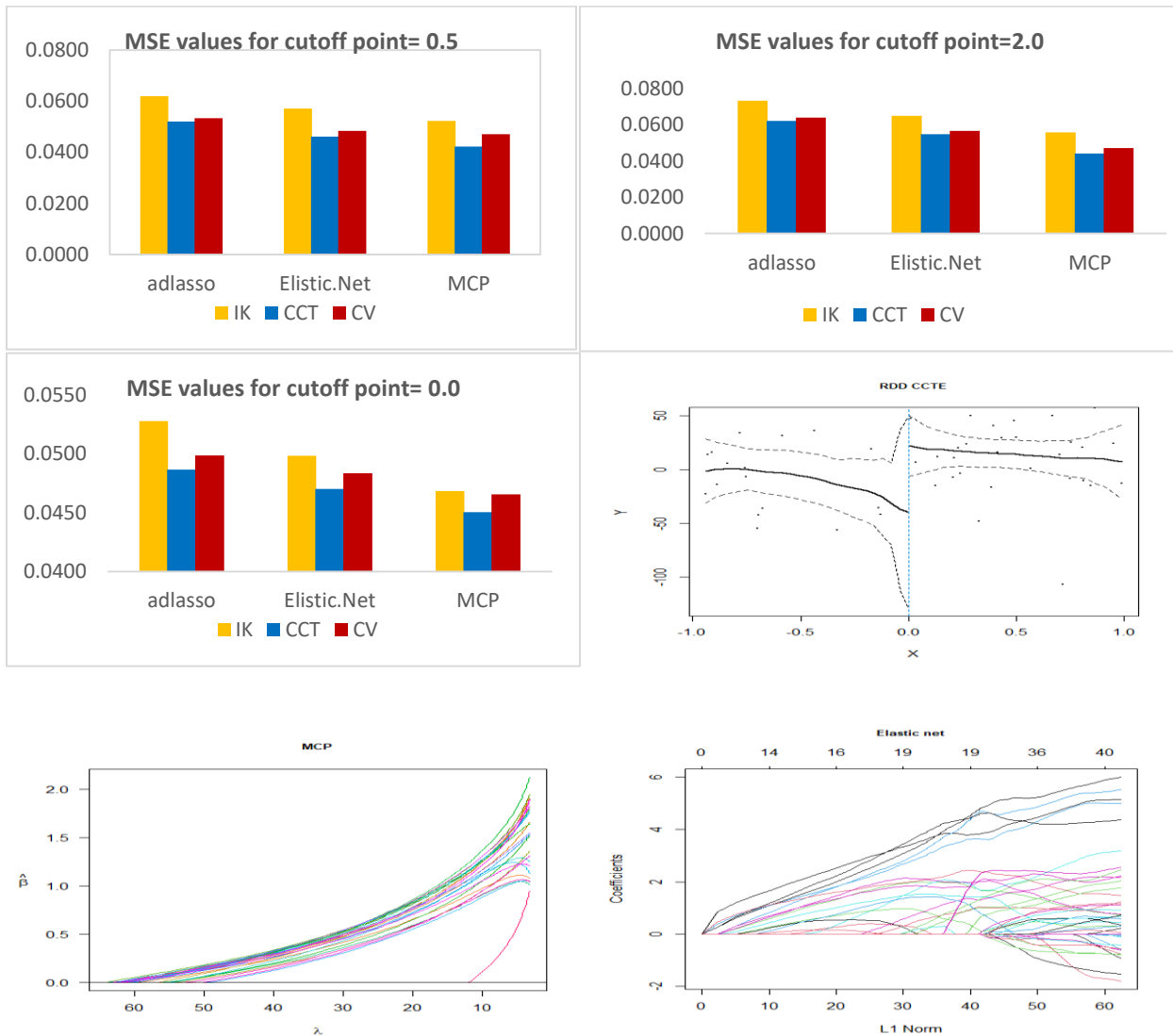

Figure (3-9): MSE Values, Cut off Point and Coefficients for Methods of Study for n=50 , p= 50, s =20 and ρ =0.75

Table(3-10): MSE values for methods of study for n=100, p= 50, s=20 and ρ=0.75

Methods	Cut off point =0.0			Methods	Cut off point =0.5			Methods	Cut off point =2		
	IK	CCT	CV		IK	CCT	CV		IK	CCT	CV
Ad lasso	0.3483	0.2142	0.2318	Ad lasso	0.1243	0.1125	0.1192	Ad lasso	0.3781	0.3108	0.3414
Elistic.Net	0.2703	0.1428	0.2634	Elistic.Net	0.1464	0.1202	0.1380	Elistic.Net	0.2548	0.2118	0.2313
MCP	0.2140	0.1367	0.1951	MCP	0.1147	0.0849	0.0903	MCP	0.1748	0.1193	0.1401

From Table (3-10) with n = 100, p = 50, s=20 and ρ =0.75 , we notice the superiority of our suggested method (MCP) over the Elastic Net and adaptive lasso through MSE values. In addition, it's clear to see that the best method of the bandwidth is CCT at all of cut- off point.   Figure (3-10) illustrates this through MSE.

Figure (3-10): Shows three cases of cut-off point (0.0, 0.5 and 2) for the preference of our suggested methods over the adaptive lasso method in the RDD model. according to the (MSE) criterion. Also, the figure at the right, shows the effect of the treatment used to change the value of the dependent variable before and after the cut-off, it is note that the number of important variables appeared far from zero, as is true coefficients that assumed by the simulation and the figure that shows the features in the methods Elastic Net and MCP.
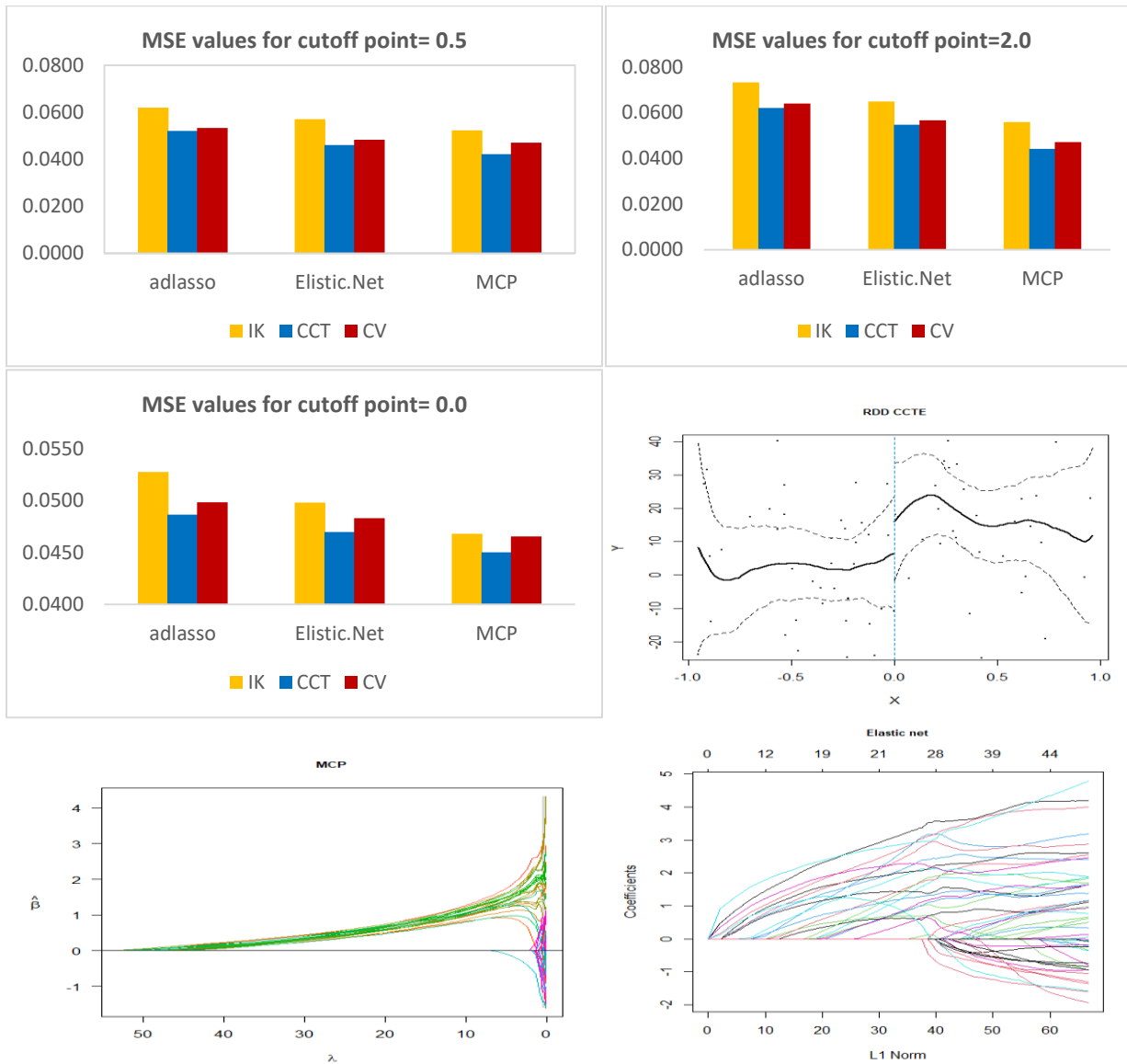

Figure (3-10): MSE Values, Cut off Point and Coefficients for Methods of Study for n=100 , p= 50, s =20 and ρ =0.75

Table(3-11): MSE values for methods of study for n=250,p $=$ 50,s $=$20 and ρ=0.75

Methods	Cut off point =0.0			Methods	Cut off point =0.5			Methods	Cut off point =2		
	IK	CCT	CV		IK	CCT	CV		IK	CCT	CV
Ad lasso	0.0864	0.0719	0.0758	Ad lasso	0.1062	0.0827	0.0863	Ad lasso	0.1557	0.1261	0.1292
Elistic.Net	0.0752	0.0634	0.0659	Elistic.Net	0.0939	0.0685	0.0762	Elistic.Net	0.1212	0.0958	0.1012
MCP	0.0641	0.0559	0.0609	MCP	0.0854	0.0630	0.0718	MCP	0.0834	0.0617	0.0700

From Table (3-11) with n = 250, p= 50, s=20 and ρ =0.75 , we notice the superiority of our suggested method (MCP) over the Elastic Net and adaptive lasso through MSE values. In addition, it's clear to see that the best method of the bandwidth is CCT at all of cut- off point.   Figure (3-11) illustrates this through MSE.

Figure (3-11): Shows three cases of cut-off point (0.0, 0.5 and 2) for the preference of our suggested methods over the adaptive lasso method in the RDD model. according to the (MSE) criterion. Also, the figure at the right, shows the effect of the treatment used to change the value of the dependent variable before and after the cut-off, it is noted that the number of important variables appeared far from zero, as is true coefficients that assumed by the simulation and the figure that shows the features in the methods Elastic Net and MCP.
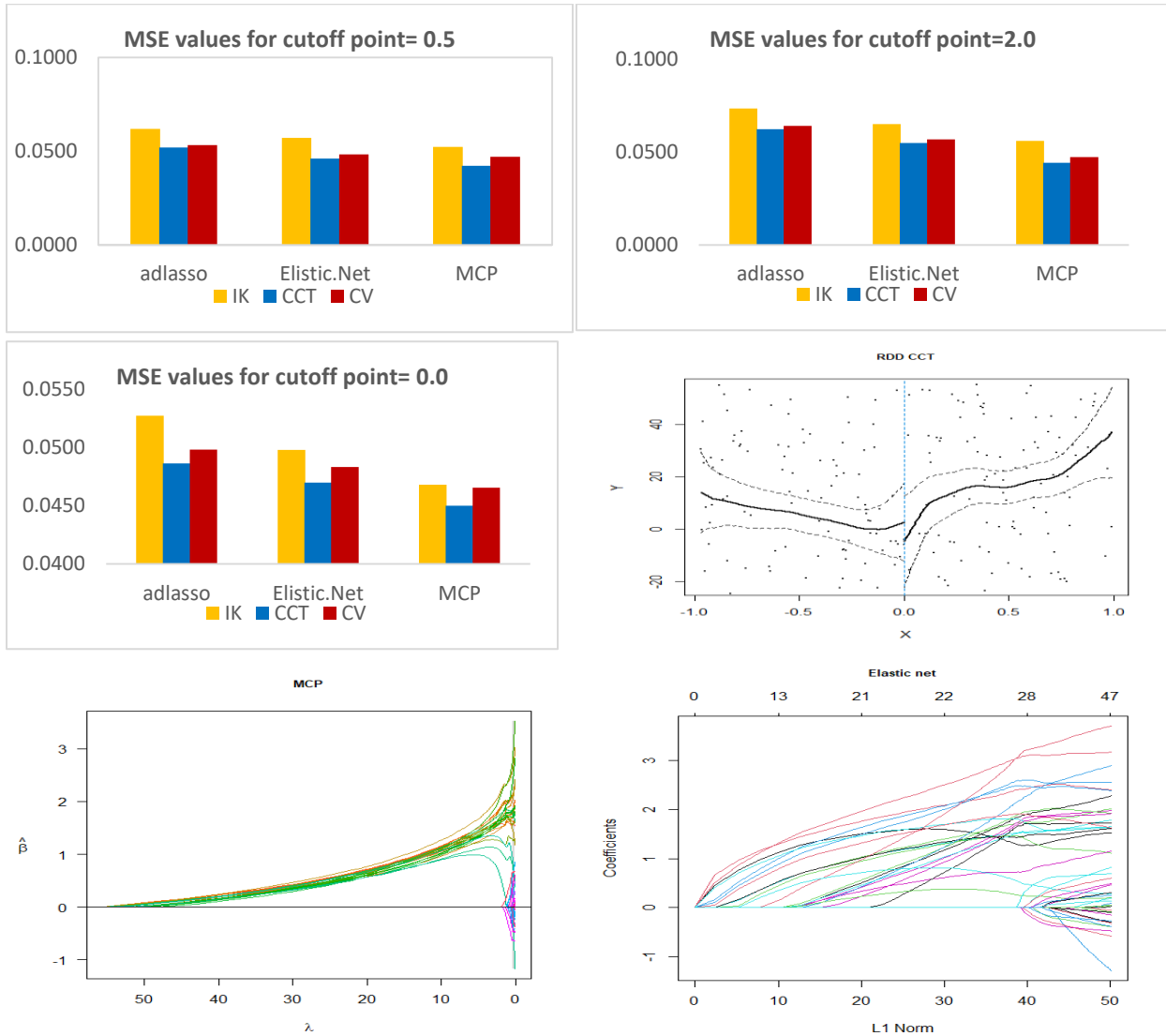

Figure (3-11): MSE Values, Cut off Point and Coefficients for Methods of Study for n=250 , p= 50, s=20 and ρ =0.75

Table(3-12): MSE values for methods of study for n=500, p= 50, s=20 and ρ=0.75

Methods	Cut off point =0.0			Methods	Cut off point =0.5			Methods	Cut off point =2		
	IK	CCT	CV		IK	CCT	CV		IK	CCT	CV
Ad lasso	0.0528	0.0486	0.0498	Ad lasso	0.0620	0.0520	0.0533	Ad lasso	0.0734	0.0621	0.0640
Elistic.Net	0.0498	0.0470	0.0483	Elistic.Net	0.0571	0.0461	0.0483	Elistic.Net	0.0650	0.0548	0.0567
MCP	0.0468	0.0450	0.0466	MCP	0.0523	0.0422	0.0470	MCP	0.0559	0.0442	0.0472

From Table (3-12) with n = 500 , p= 50, s =20 and ρ =0.75 , we notice the superiority of our suggested method (MCP) over the Elastic Net and adaptive lasso through MSE values. In addition, it's clear to see that the best method of the bandwidth is CCT at all of cut- off point.   Figure (3-12) illustrates this through MSE.

Figure (3-12): Shows three cases of cut-off point (0.0, 0.5 and 2) for the preference of our suggested methods over the adaptive lasso method in the RDD model. according to the (MSE) criterion. Also, the figure at the right, shows the effect of the treatment used to change the value of the dependent variable before and after the cut-off, it is noted that the number of important variables appeared far from zero, as is true coefficients that assumed by the simulation and the figure that shows the features in the methods Elastic Net and MCP.
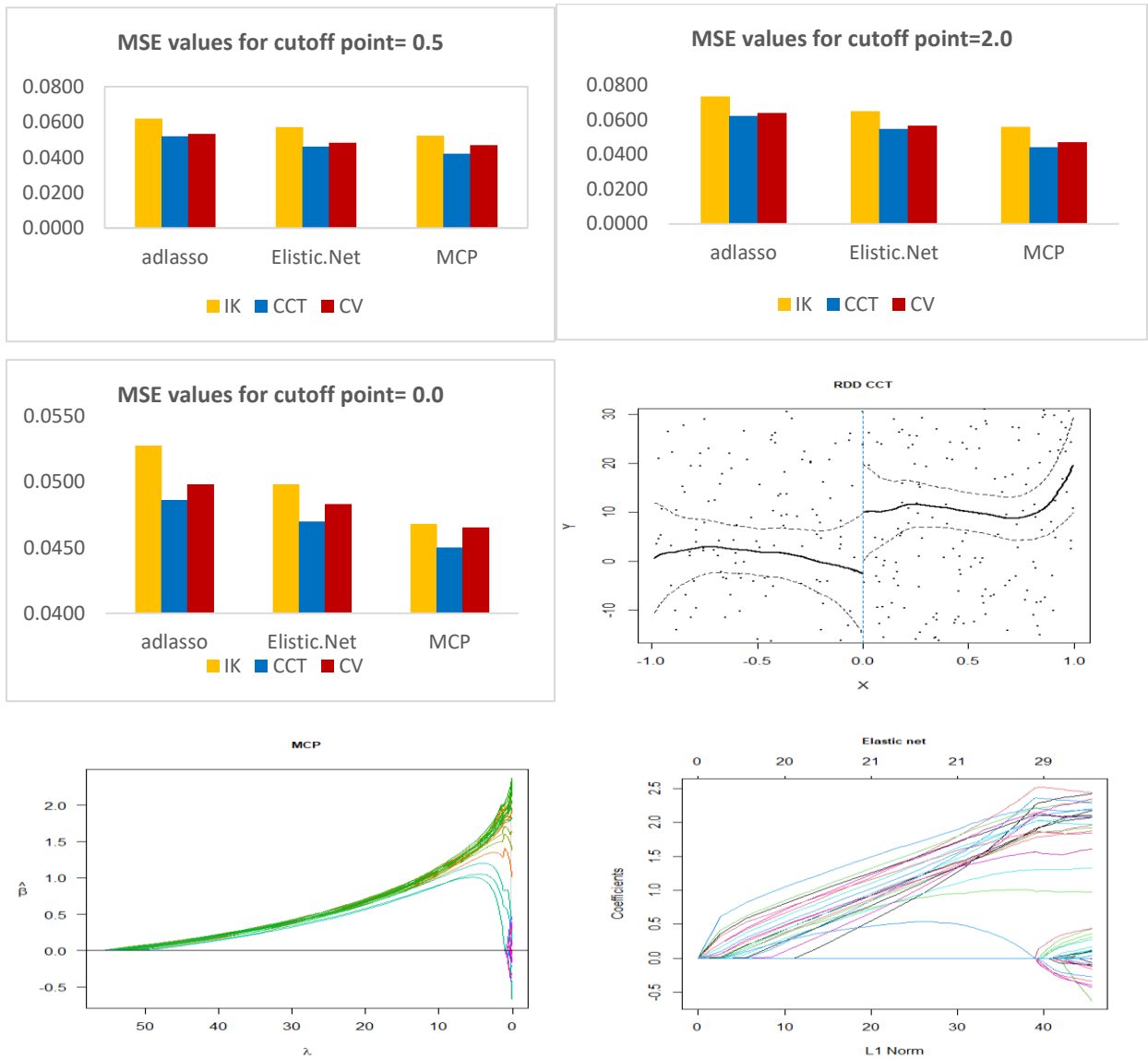

Figure (3-12): MSE Values, Cut off Point and Coefficients for Methods of Study for n=500 , p = 50, s=20 and ρ =0.75

3.2 Real Data:

In this section, real data on real Kidney failure from the Department of Nephrology (Hemodialysis) at Al-Diwaniyah Teaching Hospital, in cooperation with specialized doctors in the hospital, were used, Factors believed to cause renal failure in people have been identified, including (kidney stone size, An increase in blood pressure, Diabetes., patient's age, Use of analgesic drugs, Smoking, genetics, Drought, heart disease and Body fat percentage (BFP).

Kidney failure "is a term in medicine for cases of kidney failure to adequately filter metabolic wastes from the blood, resulting in the kidneys not being able to do their job".   The process of dialysis is performed by medical devices that filter the blood, where the blood is entered through a tube to the device, where it is filtered and then returned to the body again, and this process takes hours and must be performed several times a week.

Data on cases of kidney failure were recorded before and after patients underwent dialysis. As for patients who did not undergo the process because of their dependence on medical drugs, their case kidneys were recorded once before the dialysis process.

A sample of (79) people who visited the nephrology consultant in the aforementioned hospital was selected, and it was represented by the percentage of blood urea as an explanatory variable (the classification variable F_i) and the ability of the kidneys to perform their functions as a response variable (Y_i). The specialized doctors set the cut-off point on the values of the classification variable (65) or more, and on the basis of which the patients with renal failure were divided into two groups, one of them is the group that undergoes dialysis known as

the treatment group (intervention group) and the other is the group that is not subject to dialysis (control group)

The table below illustrates the most important factors believed to be the cause of kidney failure:

Table (3-13) Description of variables

y_i	**Kidney performance reading**
x_1	**Kidney stone size**
x_2	**An increase in blood pressure .**
x_3	**Diabetes.**
x_4	**Patient's age**
x_5	**Use of analgesic drugs**
x_6	**Smoking**
x_7	**Genetics**
x_8	**Drought**
x_9	**Heart disease**
x_{10}	**Body fat percentage(BFP)**

Kidney Performance Reading (yi):

Identifying the ability of the kidneys to meet the needs of the body and perform its functions fully by reading the degree of kidney disease before and after the dialysis process, by conducting laboratory tests for urine or blood, or by imaging tests, or by taking a sample of kidney tissue for examination.

Kidney Stone Size (x_1) :

Where the presence of stones in the kidneys leads to blockage of the urinary tract and failure of urine passage outside the body and can lead to acute kidney injury.

High Blood Pressure (x_2):

When damage occurs in one of the kidneys, the level of Blood pressure is affected and rises, because if the kidneys are unable to perform their function, the volume of blood increases, which increases pressure on the walls of blood vessels, and thus raises the level of blood pressure.

Diabetic Kidney Disease (x_3) :

represents a serious complication of the kidneys and is also known as kidney disease caused by diabetes. About 25% of people who have diabetes eventually develop kidney disease.

Patient's Age (x_4) :

The life expectancy of a dialysis patient varies from one patient to another depending on the health condition, and the extent to which the patient follows the instructions issued by the doctor or the health care to which he is subject.

But if a person suffers from a case of kidney failure like any other disease, it may be affected by some variables surrounding the patient. There are some variables that can be controlled such as lifestyle, diet, type of treatment that is used and others that cannot be controlled such as age, gender, race and genes. Adialysis patient can live for decades, a dialysis patient can live 20 years on.

Using Analgesic Drugs (x_5):

It is necessary to know the patient's condition on the basis of which the patient is given the appropriate sedative for him and to a certain extent, after examinations of the kidneys first, as well as the liver to determine the quantity,

and give him what suits his health condition. It is necessary to carry out all the tests, know the medications the patient is taking, and know the kidney tests.

Smoking (x_6) :

Cigarette smoking is a major cause of deteriorating kidney function. And it can increase the incidence of complications because the components of cigarettes can cause atherosclerosis, which increases the risk of heart attacks, and can also cause stroke. At the same time, less blood and oxygen reach the kidneys, which accelerates the deterioration of kidney function.

Genetics (x_7) :

hereditary nephritis is an inherited disorder that leads to inflammation of the glomeruli when kidney function is impaired, blood leaking into the urine, deafness, and eye abnormalities.

Drought (x_8):

Dehydration occurs when you use or lose more fluid than you take in and there is not enough water or other fluids in your body to perform its normal functions. If you don't replace lost fluids, you will become dehydrated. Anyone is susceptible to dehydration, but the condition is only serious in young children and the elderly.

Heart Diseases (x_9):

Heart diseases are the most common cause of death for patients with kidney failure especially patients undergoing dialysis for several reasons. The most important of which is that most patients undergoing dialysis may suffer from some other diseases that cause increased effort from the heart and reduce the amount of blood and oxygen going to the heart.

.

Body Fat Percentage (BFP) (x_{10}):

Excess body fat directly contributes to the risk of diabetes and high blood pressure, and is one of the main causes of chronic kidney disease.

To analyze the real data and through the simulation results mentioned above which are related to comparing the proposed methods in the RDD model to choose the important variables and select the appropriate model which showed that the MCP method was the best method by comparing with the MSE standard, so this method was applied to analyze the data and the results were shown in the table below:

Table (3-14) shows a summary of the estimated model statistics

Coefficients:	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	57.984653	12.187017	4.758	3.33e-05 ***
Tr	14.615125	10.088966	1.449	0.15634
Xl	1.606405	0.166808	9.630	2.25e-11 ***
X_r	-0.123572	0.284363	-0.435	0.66655
SX_1	1.226357	0.552742	2.219	0.03309 *
SX_2	0.049961	0.104511	0.478	0.63559
SX_3	-0.128582	0.043339	2.967	0.00539 **
SX_4	0.295110	0.119868	2.462	0.01889 *
SX_5	0.006762	0.010852	0.623	0.53727
SX_6	1.017917	0.348897	2.918	0.00613 **
SX_7	-0.545192	0.216901	2.514	0.01671 *
SX_8	2.584653	0.955602	2.705	0.01049 *
SX_9	0.016762	0.041000	0.409	0.68516
SX_{10}	0.731522	0.289897	2.523	0.01632 *

From the table (3-14) it is clear that the variables (X_1, X_3,X_4,X_6,X_7,X_8,X_{10})

represented by (kidney stone size , diabetes., patient's age, Smoking (no smoking),genetics, drought and Body fat percentage (BFP) had a significant effect through t-values.   (Tr, XI, X_r) represent the model variables (RDD) respectively.

Table(3-15): Represents a value (MSE, R², Adjusted R², F-statistic) in the model

MSE	R²	Adjusted R²	F-statistic	P-value
0.25	0.8663	0.8166	17.44	0.000

Through the table (3-15) that includes the value of (MSE=0.25), it is clear that the method (MCP).

With a coefficient of determination R-square (0.8663) this indicates that the explanatory variables (the classification variable Fi) explained 86% of the response variable and can be considered as an indicator of the fit of the estimated model to the data. adjusted R-squared was (0.8166), which is another measure to judge the extent to which the estimated model matches the data, since the value of (P-value =0.000) and based on the value of (F-statistic =17.44) the model was significant, with a level of significance of 5%.

Figure (3-13): Shows the data spread pattern before and after the cut-off point. As for the figure on the right, it shows the cross-validation process in choosing the shrinkage coefficient values. Also, the figure at the bottom left, shows the effect of the treatment used to change the value of the dependent variable before and after the cut-off using Bandwidth (CCT) in addition to a diagram of the model using the (MCP) method.
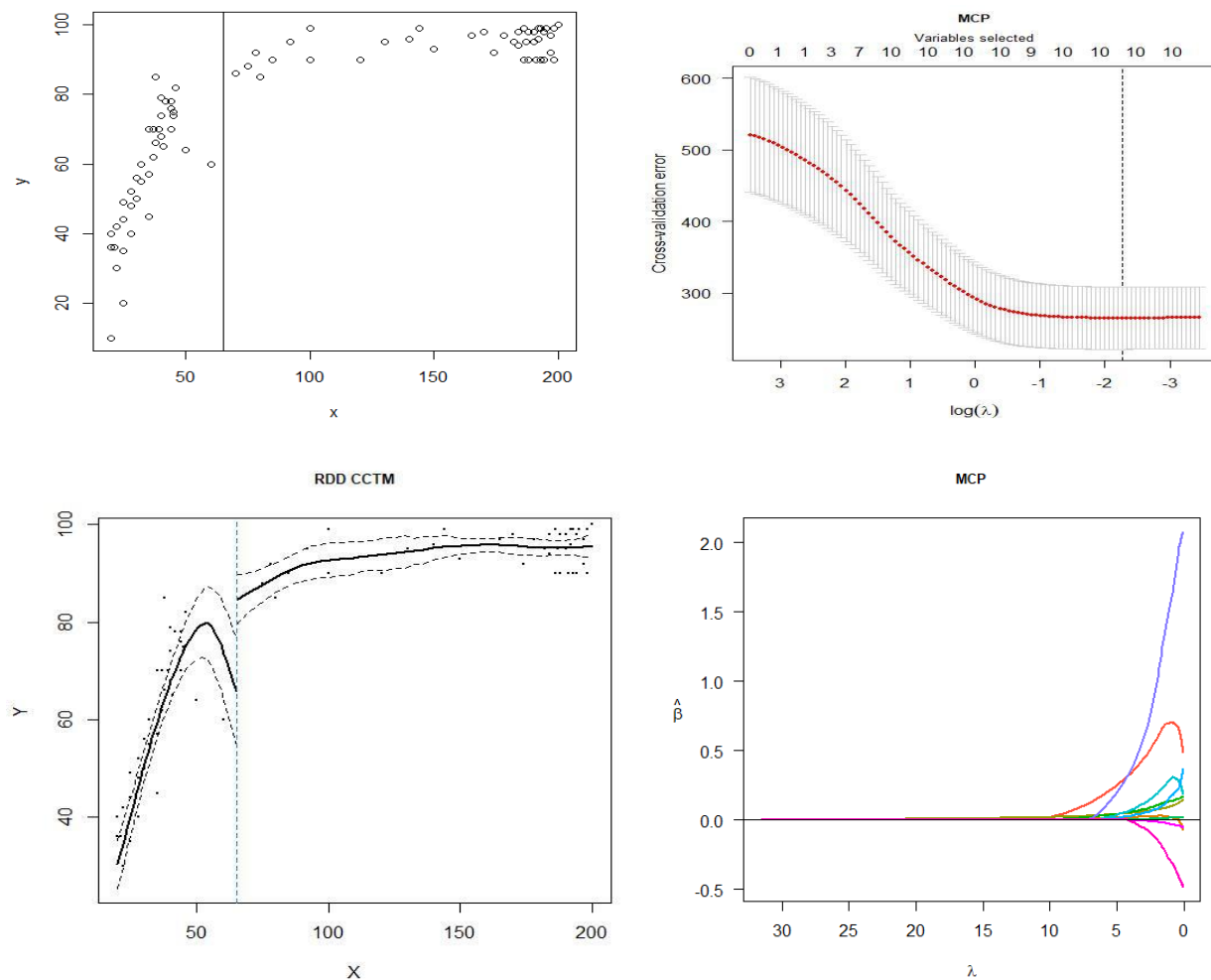

Figure (3-13) :A schematic diagram of real data using the (MCP) method.

CHAPTER FOUR

CONCLUSION AND RECOMMENDATIONS

4.1 Conclusion:

1 − This study introduced new techniques for choosing a discontinuity regression design (RDD), in which the procedures of the MCP method and the Elastic network method were proposed to estimate the coefficients with VS.

2 − The proposed procedures are based on the standard (MSE) for the purpose of distinguishing the proposed methods. Moreover, the advantages of the new procedures in both simulation and analysis of real data are also explained in Chapter 3.

3 − The results showed that these procedures performed well in terms of VS and estimation parameters. In particular, the MCP method is by far the best among the above procedures as it achieves the least (MSE) .

4 − The conclusions of this study state that the statisticians are assisted by the technique of organization methods in statistics, using this new technique to ensure accurate and useful results for correct prediction.

5 − The best bandwidth method used is the CCT method, followed by the CV bandwidth method Bandwidth and then IK by comparing with the value of the mean square error .

6 − Through the simulation study and analysis of real data, the important factors affecting kidney patients were known, including (kidney stone size , diabetes, patient's age, smoking (no smoking) , genetics, drought and body fat percentage (BFP) and when they undergo dialysis.

7 −   The larger the sample size, the less is obtained (MSE) .

4.2 Recommendations and Future Work

The regularization procedures in this study will provide researchers with promising hope, to introduce and extend new procedures for estimating coefficients and V.S in the of regression discontinuity design (RDD).

There are many other possible extensions, such as using Group Lasso, SCAD and other methods of organizing in design RDD.

Experimenting more than one discontinuity point, that is, more than one treatment, which is called multi- discontinuity regression.

Using multiple discontinuity regression, that is, for more than one explanatory variable, as this leads to the use of more than one discontinuity point.

Using fuzzy discontinuity regression (FDR) to estimate the effect of treatments when there is not enough information about the choice of cut- off point with variable selection method.

REFERENCES

References:

[1] Akaike, (1974). A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), 716-723.

[2] Albino, Dominic. (2020). "Prctical Validity of the Continuity Assumption in Regression Discontinuity Designs : Granularity, Deghree of Focus and Multiple ThreSholds ". University of Cannecticut.

[3] Ali, et al., (2020). Kernel estimation of returns of retirement funds of employers based on monetary earnings (subscriptions and compensation) via regression discontinuity in Iraq. Periodicals of Engineering and Natural Sciences *(PEN)*, *8*(3), 1752-1766.

[4] Anastasopoulos, (2019). Principled estimation of regression discontinuity designs with covariates: a machine learning approach. arXiv preprint arXiv:1910.06381.

[5] Benavente, et al., (2012). The impact of national research funds: A regression discontinuity approach to the Chilean FONDECYT. Research Policy, 41(8), 1461-1475.

[6] Bloniarz, et al., (2016). Lasso adjustments of treatment effect estimates in randomized experiments. Proceedings of the National Academy of Sciences, 113(27), 7383-7390.

[7] Breheny, (2016). Adaptive lasso, MCP, and SCAD.

[8] Calonico, et al., (2018). On the effect of bias estimation on coverage accuracy in nonparametric inference. Journal of the American Statistical Association, 113(522), 767-779.

[9] Calonico, et al., (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. Econometrica, 82(6), 2295-2326.

[10]   Calonico, et al., (2015). rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs. R J., 7(1), 38

[11]   Calonico, et al., (2019). Regression discontinuity designs using covariates. Review of Economics and Statistics, 101(3), 442-451.

[12]   Cappelleri, et al., (2015). Regression Discontinuity Design. International Encyclopedia of the Social & Behavioral Sciences, 2nd edition, Volume 20, 152-159

[13]   Cattaneo, et al. (2019). A practical introduction to regression discontinuity designs: Foundations. Cambridge University Press.

[14]   Cattaneo, et al., (2017). Comparing inference approaches for RD designs: A reexamination of the effect of head start on child mortality. Journal of Policy Analysis and Management, 36(3), 643-681.

[15]   Cattaneo, et al., (2019) "The Regression Discontinuity Designs" Arxiv: 1906. 04242v1 [econ. EM] 10-Jun (2019).

[16]   Caughey, & Sekhon, (2011). Elections and the regression discontinuity design: Lessons from close US house races, 1942–2008. Political Analysis, 19(4), 385-408.

[17]   cook, et al., (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management, 27(4), 724-750.

[18]   Choon, (2012). Minimax concave bridge penalty function for variable selection.

[19]   Dunning, (2008). Institutional reform, FDI and the location competitiveness of European transition economies. Foreign direct investment, location and competitiveness, Oxford: Elsevier, 175-201.

[20]   Friedman, (2010). Regularization paths for generalized linear models via coordinate descent. Journal of statistical software, 33(1), 1.

[21]   Hahn, (2001). Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica, 69(1), 201-209 .

[22]   Hoerl, & Kennard, (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.

[23]   Imbens, & Lemieux, (2008). Regression discontinuity designs: Aguide to practice. Journal of econometrics, 142(2), 615-635.

[24]   Imbens, & Kalyanaraman, (2009). Optimal bandwidth choice for the regression discontinuity estimator (Working Paper No. 14726). National Bureau of Economic Research. doi, 10, w14726.

[23]   Imbens, & Kalyanaraman, (2012). Optimal bandwidth choice for the regression discontinuity estimator. The Review of economic studies, 79(3), 933-959.

[25]   Jacob, et al., (2012). A practical guide to regression discontinuity. *MDRC*.

[26]   Kettlewell, & Siminski, (2020). Optimal Model Selection in RDD and Related Settings Using Placebo Zones. Life Course Centre Working Paper, (21).

[27]   Lee, & Lemieux, (2010). Regression discontinuity designs in economics. Journal of economic literature, 48(2), 281-355.

[28]   Lee, & Lemieux, (2014). Regression discontinuity designs in social sciences. Regression Analysis and Causal Inference, H. Best and C. Wolf (eds.), Sage.

[29]   Linden, et al., (2006). Evaluating disease management programed effectiveness: an introduction to the regression discontinuity design. Journal of evaluation in clinical practice, 12(2), 124-131.

[30]   Ludwig, et al., (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. The Quarterly journal of economics, 122(1), 159-208.

[31]   Meier, et al. (2008). The group lasso for logistic regression. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70(1), 53-71.

[32]   Melly, & Lalive, (2020). Estimation, Inference, and Interpretation in the Regression Discontinuity Design. Universität Bern, Faculty of Business, Economics and Social Sciences, Department of Economics.

[33]   Mutair, Hafez Muhammad, 2011 "Comparison of Some Non-Parametric Regression Methods Using Simulations", Master's Message, College of Computer Science and Mathematics, University of Al-Qadisiyah.

[34]   Porter, (2003). Estimation in the regression discontinuity model. Unpublished Manuscript, Department of Economics, University of Wisconsin at Madison, 2003, 5-19.

[35]   Robinson, et al., (2009). Instructional climates in preschool children who are at-risk. Part II: Perceived physical competence. Research quarterly for exercise and sport, 80(3), 543-551.

[36] Schwartz, & Wilde, (1978). Intervening in markets on the basis of imperfect information: A legal and economic analysis. U. Pa. L. Rev., 127, 630.

[37] Szakonyi, David. 2018. "Businesspeople in Elected Office: Identifying Private Benefits from Firm-Level Returns." American Political Science Review 112 (2): 322–338.

[38] Thistlethwaite, & Campbell, (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. Journal of Educational psychology, 51(6), 309.

[39] Tibshirani, (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society : Series B (Methodological), 58(1), 267-288.

[40] Treasury, (2007). The Magenta Book: guidance notes for policy evaluation and analysis. London: HM Treasury (Magenta Book Background Papers.

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

[42] Zhang, (2010). Nearly unbiased variable selection under minimax concave penalty. Annals of Statistics 38, 894–942

[43] Zou, (2006). The adaptive lasso and its oracle properties. Journal of the American statistical association, 101(476), 1418-1429.

[44] Zou, & Hastie, (2005). Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology), 67(2), 301-320 .

المستخلص

يعتبر تصميم عدم استمرارية الانحدار (RDD) أحد الأساليب الشائعة للاستدلال السببي للبيانات المشاهدة . تم تطبيقه في الإحصاء والاقتصاد القياسي والعلوم السياسية وعلم الأوبئة والتخصصات ذات الصلة. غالبًا ما يتم تقدير تأثيرات المعالجة باستخدام RDD بعدد قليل من المشاهدات وعندما يكون الشكل الدالي الصحيح للانحدار المحلي غير معروف. في هذا التصميم يتم تقسيم مجتمع الدراسة الى مجموعتين على اساس نقطة قطع محددة . مقدمة حسب شروط ومتطلبات الدراسة. عادة ما يعتمد تقدير متوسط تأثيرات المعالجة الموضعية في RDD على الانحدارات الخطية الموضعية باستخدام متغير النتيجة ، ومتغير مهمة المعالجة ، ومتغير التشغيل المستمر.

في هذه الدراسة , تم دمج بعض طرق اختيار المتغيرات مثل (Elastic net) ونموذج (MCP) مع نموذج RDD .كما تم استخدام ثلاث انواع من عرض النطاق الترددي (Iembens and kalyanman) IK bandwidth, cross-validation (CV) method, and The CCT (Calonico, Cattaneo & Titiunik) bandwidth بالإضافة إلى ذلك ، غالبًا ما يتم إضافة المتغيرات المشتركة لزيادة كفاءة تقديرات تأثير المعالجة والتكيف مع الاختلالات المعروفة. لتقييم الطرق المقترحة تمت مقارنتها بطريقة Adaptive Lasso باستخدام دراسة المحاكاة. استخدم متوسط أخطاء التربيع للمقارنة واختيار أفضل طريقة.

وقد تم توظيف هذه الاساليب لدراسة موضوع اللجوء الى غسيل الكلى كعلاج ، نتيجة للتعرض للفشل الكلوي للتخلص من الفضلات والسموم الموجودة في الدم مثل اليوريا والكرياتنين وغيرها . حيث تم استخدام بيانات حقيقية بحج م (79) شخصا ممثلة بنسبة اليوريا في الدم كمتغير توضيحي (Fi متغير التصفية) وقدرة الكلى على اداء وظيفتها كمتغير استجابة (Yi) .

تم استخدام برنامج R في عملية التحليل, تبين ان افضل طريقة لتقدير متوسط تأثير المعالجة واختيار المتغير هي طريقة (MCP) عند استخدام النطاق الترددي (CCT) لأنها تحقق أقل (MSE), والنتائج في دراسة المحاكاة والبيانات الحقيقية تشير الى ان الطرق المقترحة لها اداء متفوق .

اختيار متغير لتصاميم عدم استمرارية الانحدار مع التطبيق

رسالة مقدمة إلى
مجلس كلية الإدارة والاقتصاد - جامعة القادسية
وهي جزء من متطلبات نيل درجة الماجستير
في الإحصاء

من قبل الطالبة

أشواق عبد السادة كاظم

بإشراف

م. د. بحر كاظم محمد

١٤٤٣ﻫ ٢٠٢١م