Elastic net Variable selection regularization method for Regression Discontinuity Designs with Application

Bahr kadhim Mohammed ,   Ashwaq Abdul Sada Kadhim

Department of Statistics, University of Al-Qadisiyah, Iraq.

E-mails: bahr.mahemmed@qu.edu.iq
   ashwaq.abdul.sadah.kazem@gmail.com

ABSTRACT

The classical method faced a big problem with estimating and selecting important variables when the dataset has a cut-off point. Therefore, there is a need to propose a new method to solve these problems. In this paper we suggested a new approach by combining the Regression Discontinuity Designs (RDD) with the elastic net. Local linear regression (LLR) method was used to estimate the effect of processing on the cut-off region of the observations within the optimum bandwidth selection for the RDD design to obtain the best model. Three models were used to determine the bandwidth, IK method, Cross-Validation (CV) method and CCT method. A simulation study and real data are conducted to investigate the performance of the proposed method. The mean squared errors (MSE) is used to choose the best model.

Keywords: Regression Discontinuity Designs (RDD), Elastic net**,** variable selection, Local linear regression, bandwidth selection, IK, CV, CCT.

1. Introduction

RDD is a quasi-experimental pre test and post test, design that extract the causal effects of interventions by assigning a cutoff 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. Variable selection methods are one of the well-sophisticated field in the modern statistics. In this study, we deal with one of the most commonly used models in this area, which is the Regression-Discontinuity Design model. First applied by (Donald Thistlethwaite and Donald Campbell (1960) to the evaluation of scholarship programs, the RDD has become increasingly popular in recent years.

In regularization methods the Variable Selection (V.S) is implemented with the process of the parameter estimation.   Examples of regularization approaches are the Lasso (Tibshirani, 1996) explained that ridge regression and Lsso regression, each method in which the penalty is applied to each additional variable added to the OLS, , Elastic Net (Zou and Hastie, 2005) which combined Ridg's penalty and Lasso's penalty with a "group lasso " used for the purpose of selecting a large set of covariates. , adaptive Lasso (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 di   mensions , MCP (Zhang, 2010) 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 and others . Anastasopoulos, L. J. (2019) employ method adaptive lasso with rdd model .

In this paper, we will employ one of the variable selection methods, which is elastic net method with RDD, where use local linear regression (LLR) in the cut-off region of the observations within the optimal bandwidth range chosen for the RDD design on either side of the cut-off point $F_i \in$ (c - h, c + h) to obtain the lowest MSE. Three models were used to select the bandwidth, The IK method proposed by (Imbens and Kalyanaraman, (2009), The Cross-validation (CV) approach proposed by (miller and Ludwig, (2007) the CCT method was proposed by (Calonico et al. (2014) where the MSE criterion was adopted to compare the proposed method and some previous methods, where we used this criterion to determine the performance of those methods.

This paper is organized as follows: We present in Section 2 basics about discontinuity regression designs and Local Linear Regression (LLR) ;   In section 3 basics about Bandwidth Selection and some methods that were used by the researcher   ; In section

4 we present a method for selecting a variable using elastic net method ; In section 5 we explain the selection of the variable by combining each of the elastic net penalty function and model (RDD). In Section 6 we summarize the results of the simulation study and present the data for the sample analysis. A brief conclusion is included in Section 7.

2 . Regression-Discontinuity Design (RDD)

RDD model is divided into two groups on the basis of a specific threshold limit or the so-called breakpoint (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 (RD) stems from the fact that the treatment effect will lead to a 'jump' or discontinuity 'in the regression function point of the relationship between Classification variable (F_i)(anexplanatory viable) in a discontinuity design with response variable Y_i . (Lee, D. S., & Lemieux, T., 2010)

When estimating RDD, covariates should be included before treatment, for the purpose of obtaining the most accurate treatment effect estimates (Cattaneo 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. Making preliminary decisions regarding the covariates variables that must be included before performing a treatment should always be based on expert judgment and the researcher's expectations that are closely related to the problem at hand.

The simplest method to estimate the treatment effect is by using local linear regression (LLR) in the cut-off region of the observations within the optimal bandwidth range chosen for the RDD design on either side of the cut-off point $F_i \in$ (c - h, c + h) to obtain the lowest MSE. Three models were used to select the bandwidth, The IK method proposed by (Imbens and Kalyanaraman, (2009)), The Cross-validation (CV) approach proposed by (miller and Ludwig, (2007)), the CCT method was proposed by (Calonico et al. (2014)) .

$$Y_i = \alpha + \tau T_i + \gamma F_i + \delta(F_i.T_i) + XB + \epsilon_i \quad .... (1)$$

where:

α = the average value of the outcome for those in the treatment group after controlling for the rating variable. $\hat{\tau}$: Estimated local average treatment effect. F_i: The forcing variable or rating variable for observation i, centered at the cut-point. T_i: An indication whether or not to receive treatment:

$$T_i = \begin{cases} 1 & \text{receiving treatment} \\ 0 & \text{Not receiving treatment} \end{cases}$$

γ: Coefficient of the forcing variable. $f(T_i, F_i)$: It is a function of the force variable which is in the form of a nonparametric kernel or a polynomial of order p^{th}. δ: Coefficient $(F_i.T_i)$, β: Coefficient vector $\beta_{p\times1}$, X: Matrix of covariates $(n \times P)$ (Anastasopoulos, J. (2019)).
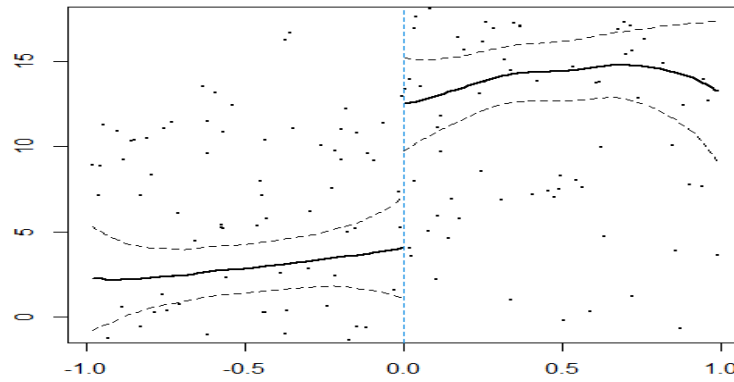

Figure 1: illustrating how the RD design got its name (Regression Discontinuity Design), in which a treatment effect is suggested when a "jump" or discontinuity in the regression lines occurs at the cutoff point.

2-1 Local Linear Regression (LLR)

Local linear regression is a non-parametric method that is used to continuously estimate the treatment effect in RDD model (Porter, 2003). Neighborhood idea h is the basis of the LLR. Where h bandwidth is chosen. In this method, points within the radius h of x_0 are determined. Points near x_0 are given greater weights than those further away from x_0. The average weight to weigh the adjacent observed data by the

kernel which is a statistical technique for estimating the reality of the function.   The kernel function K (u): R → R, , has the following properties.

1. $0 \le k(u) \le \infty$, $K(u)$ 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, $\int u\,K(u)\,du = 0$ $and\ \sigma^2 = \int u^2\,K(u)\,du > 0$

3. 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 increases, the bias increases and the variance decreases and vice versa, and as a result it will have a clear effect on smoothing the curve and the rate of its approach to the original curve.   The basic idea of choosing a bandwidth in the SRD is a trade-off between bias and variance for $(\hat{\tau}_c^{\,SRD})$ where the greater the bandwidth, the greater the bias and the less variance. There are several methods for choosing the optimal value of the bandwidth that have been used by many researchers such as cross validation and plug-in method, and many other methods (Hill, R. Carter and Kang-sun Lee., 2001). The methods that were used by the researcher will be covered in this research **:**

3.1 IK method

The IK method was suggested by (Imbens and Kalyanaraman, 2009). The researchers explained that the optimal choice of the Bandwidth Optimal is by substituting the six unknown quantities shown in the equation below, which will ultimately lead to the consistent estimators.

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

When obtaining the six unknown estimators in equation (2), the optimum bandwidth estimate is according to the following formula:

$$\hat{h}_{opt}^{IK} = \arg\min_h AMSE(h) = CK \left(\frac{\hat{\sigma}_+^2(c) + \hat{\sigma}_-^2(c)}{f(c).(\hat{m}_+^2\ (c) - \hat{m}_-^2\ (c))} \right)^{1/5} N^{-1/5} \ \text{.... (3)}$$

3.2 The Cross-validation 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 decreases, the value of the bandwidth increases and the bias value begins to increase. The package width that has the lowest value for the Cross-validation criterion (CV) is chosen according to the following formula:

$$h_{CV} = \ \min_{h>0} CV_Y\ (h) \qquad\qquad (4)$$

3.3 The CCT method.

The CCT method was proposed by (Calonico et al.(2014)).   We estimate the bounds of (asymptotic variance) by finding the initial bandwidth (Vn, Cn) denoted by (V) where:

$$\hat{V} = 2.58\ \omega\ n^{-\frac{1}{5}} \qquad \text{…..…… (5)}$$

where: $\qquad\qquad \omega\ = min\ \{\ S_x\ ,\frac{IQR_x}{1.349}\}\ \ \text{………. (6)}$

(S_x) Denotes the sample variance, $(\ IQR_x\ )$ indicates the interquartile range, and the bandwidth $(\hat{C}_n)$ where :

$$\hat{C}_{n=} \begin{pmatrix} 1/(2q+5) \\ v\quad,\quad p\ ,\ q \end{pmatrix} .\ n^{-1/(2q+5)} \ \ \text{.... (7)}$$

We find the bandwidth $(\hat{b}_{CCT})$ and it is calculated according to the following formula:

$$\hat{b}_{CCT} = \begin{pmatrix} 1/(2q+5) \\ 0\ ,\ 1\ ,\ 2 \end{pmatrix} \qquad \text{…. … … ….. (8)}$$

$$\hat{C}_{0,1,2} = \frac{5n\ \hat{v}_n^5\ ,\hat{V}_{2,2}\ (\ \hat{v}_n\ )}{2\ \beta_{2,2}^2\{\ \left(\ \acute{e}_3\ \hat{B}_{+,3}\ (Cn) + \acute{e}_3\ \hat{B}_{-,3}\ (\hat{C}_n)\right)^2 + 3\hat{V}_{3,3}\ (\hat{C}_n)\}} \text{……..(9)}$$

We find the basic bandwidth $(\hat{h}_{CCT})$ according to the following formula:

$$\hat{h}_{CCT}=\begin{pmatrix} 1/5 \\ 0 ,1,0 \end{pmatrix}.\ n^{-1/5} \quad \ldots\ldots\ldots (10)$$

$$C_{0,1,0}=\frac{n\ \hat{v}_n^5\ ,\mathcal{V}_{0,1}\ (\hat{v}_n)}{4\ \beta_{0,1}^2\{\ \left(\acute{e}_2\ \hat{\beta}_{+,2}\ (\mathfrak{b}_{CCT})+\acute{e}_2\ \hat{\beta}_{-,2}\ (\mathfrak{b}_{CCT})\right)^2+3\mathcal{V}_{3,3}\ (C_n\ )\}}\ \ldots.(11)$$

4.Variable selection in the elastic net regularization

Elastic net 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:

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

In the case of a group of highly correlated variables, the lasso selects one variable from the set of variables and ignores the other. To overcome these limitations, this method adds a square portion to the penalty $((\ \|\beta\|^2))$. , which 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), it's one of the first contraction methods for 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:

$$\hat{\beta}_{Ridge}\ =arg\min_{\beta}\sum_{i=1}^{N}(Y_i-X\beta)-\lambda\sum_{j=1}^{P}\beta_j^2\quad\ldots.\ (12)$$

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. 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, H., & Hastie, T. (2005)).

$$\hat{\beta} = arg \min_{\beta} (\|Y - XB\|^2 + \lambda_2 \|\beta\|^2 + \lambda_1 \|\beta\|_1) \quad \ldots \text{ (13)}$$

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. The Elastic net method solves the problem:

$$\min_{\beta_0 \beta} \left(\frac{1}{2n} \sum_{i=1}^{N} (Y_i - \beta_0 - X_i^T \beta)^2 + \lambda \, P_\alpha(\beta) \right) \quad \ldots \text{ (14)}$$

n : is the number of observations. Y_i : is the response at observation i.

X_i : is data, a vector of p values at observation i. λ : is a positive regularization parameter corresponding to one value of Lambda.

The parameters β_0 and β are scalar and p- vector respectively.

Whereas:

$$P_\alpha(\beta) = \frac{(1-\alpha)}{2} |\beta|_2^2 + \alpha |\beta_j|_1 = \sum_{i=1}^{P} \left(\frac{(1-\alpha)}{2} \right) \beta_j^2 + \alpha |\beta_j| \quad \ldots \text{ (15)}$$

The penalty term $P_\alpha(\beta)$ combination between the L_1 norm of β and the squared L_2 norm of β, and it is called an elastic net penalty. In this paper, we will employ Elastic net method with model (RDD) for the purpose of estimating and selecting the variable by integrating the Elastic net penalty function with the model (RDD) according to the following formula:

$$arg\ \min_\theta \sum_{i=1}^N [Y_i - (\alpha + \tau T_i + \gamma F_i + \delta(F_i.T_i) + X\beta]^2 + \sum_{i=3}^N \left(\frac{(1-\alpha)}{2}\right)\beta_j^2 + \alpha|\beta_j| \dots \quad (16)$$

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

6. Application side

6.1 SIMULATION STUDY:

In order to illustrate employ of the elastic net with RDD the simulation study is used by using the program R, that is explain as follows:

Step1: A sample was generated in the following sizes (50, 100, 150, 250) and p = 50 variables include (s) nonzero variables. That is mean, null variables are p-s.

Step2: Correlations have been formed between the variables from 1 to k; (1 ,…,k). Where k represents the number of variables related to the amount of correlation (r = 0.75).

Step3: Two types of variables were created, where the first type is a treatment variable. It has been generated according to a uniform distribution. Sample size n with terms a = - 1 represents the minimum and b = 1 represents the upper bound. With a parameter value of (10 , 2) (treatment variables and the two treatment parameters).The variables of the second type $\acute{x}\beta$ (the rest of the variables) were generated according to the normal distribution with a parameter vector μ of (zero) and with degree $1 \times p$ and a common variance matrix of sigma (σ_{ij}) of degree $(p \times p)$ where the main diameter

elements of this matrix are (1) As for the rest of the elements, it is equal to (Rou) when $i \neq j$ and that i, j is less than k where $i, j < k$)) and zero when $(i, j > k)$.

Step4: The random error term (e) was generated according to the standard normal distribution $N(0,1)$, and the data were generated based on the following model (TrBeta). and repeat each experiment (IT=1000) for all of the simulation experiments.

Step 5: Calculate the MSE.

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

$$\beta = (\underbrace{0.5 , 1, 1.5. \ 2,, 3,}_{S} \underbrace{0,0, \dots ,0}_{P-S})$$

Table 1: MSE values for methods of study for n=100, p= 15, s=5 and ρ =0.75 .

Cut of point	Methods	bandwidth		
		IK	CCT	CV
0.0	Ad Lasso	0.14515	0.12002	0.14023
	Elistic.Net	0.10527	0.08955	0.10153
0.5	Ad Lasso	0.14786	0.12047	0.13066
	Elistic.Net	0.13003	0.10421	0.11485
2.0	Ad Lasso	0.23035	0.18039	0.20433
	Elistic.Net	0.16986	0.14191	0.15056

In this example 1, From Table 1 with n = 100, p= 15, s=5 and ρ =0.75 , we notice the superiority of our suggested method (Elastic Net) over the 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.
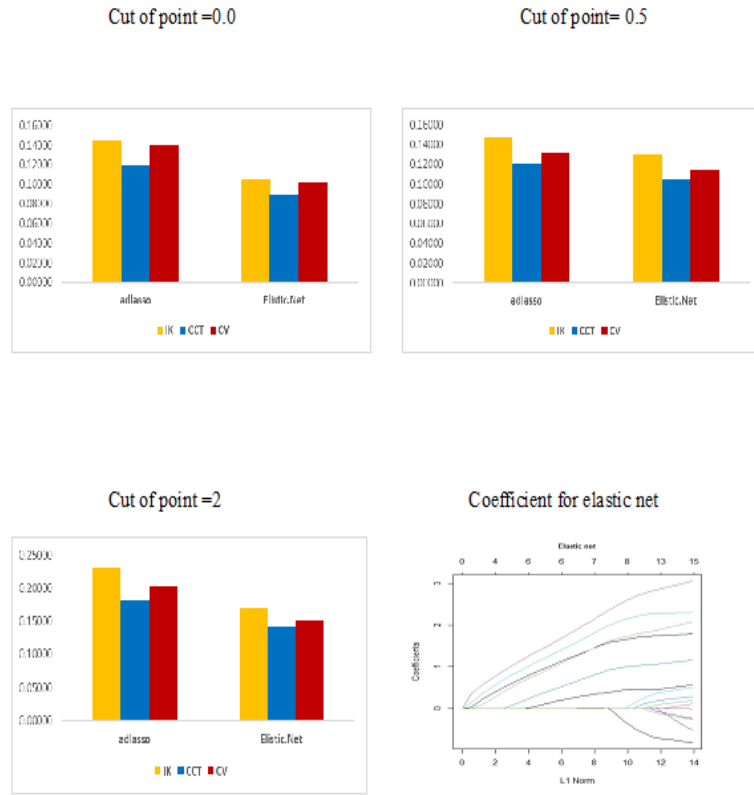
Cut of point =0.0

Cut of point= 0.5

Cut of point =2

Coefficient for elastic net

Figure 2: MSE values for methods of study for n=100, p= 15, s=5 and ρ =0.75

Figure (2) shows three cases of cut-off point (0.0, 0.5 and 2) for the preference of our suggested method over the adaptive lasso method in the RDD model. according to the (MSE) criterion, we 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 method Elastic net.

Example 2: Samples size (n=100) , number of variables (p=25) ,(s=10),(p-s=15) and $\rho = 0.75$. where

$$\beta = (\underbrace{0.5 , 1, 1.5. \ 2,, 3,}_{S} \underbrace{0,0, ... ,0}_{P-S})$$

Table 2: MSE values for methods of study for n=100, p= 25, s=10 and ρ=0.75

Cut of point	Methods	bandwidth		
		IK	CCT	CV
0.0	Ad Lasso	0.12690	0.11497	0.12158
	Elistic.Net	0.09777	0.09202	0.09472
0.5	Ad Lasso	0.14102	0.10623	0.12439
	Elistic.Net	0.10661	0.08821	0.10530
2.0	Ad Lasso	0.21182	0.14489	0.17451
	Elistic.Net	0.13187	0.10103	0.13132

In this example 2, From Table 2 with n= 100, p= 25, s=10 and ρ =0.75, we notice the superiority of our suggested method (Elastic Net) over the adaptive lasso through MSE values. We also note that the best method of the bandwidth is CCT.
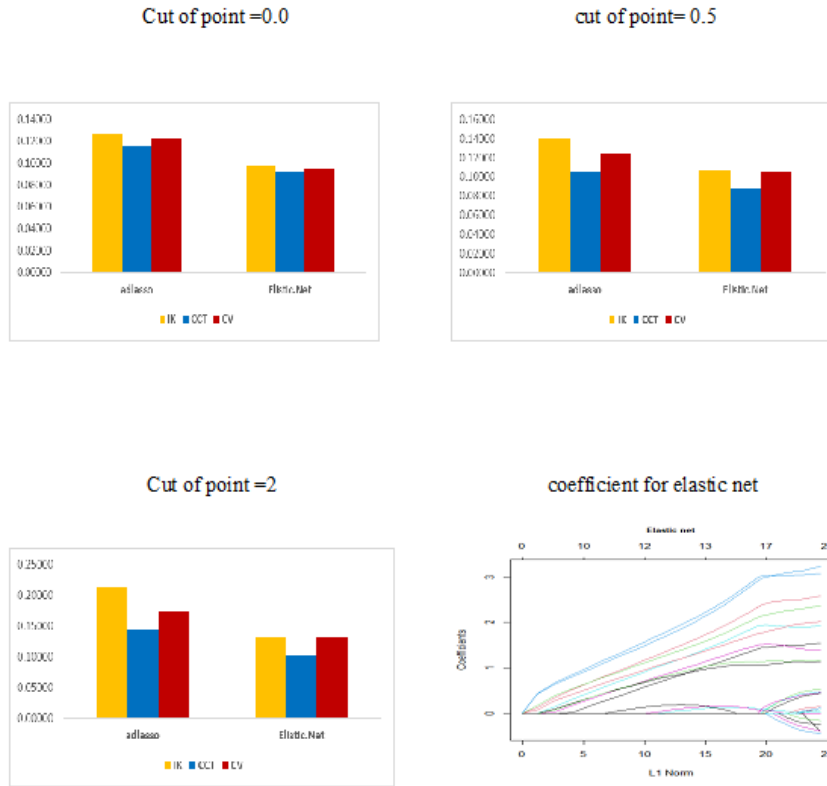
Figure 3: MSE values for methods of study for n=100, p= 15, s=5 and ρ =0.75

Figure (3) shows the preference of our suggested method over the adaptive lasso method for RDD model. According to the (MSE) values, we note that the number of important variables appeared far from zero, as is assumed by the simulation through and the figure that shows the features in the method Elastic net.

Example 3: Samples size (n=100) , number of variables (p=50) ,(s=20),(p-s=30) and $\rho = 0.75$. where

$$\beta = (\underbrace{0.5}_{S}, 1, 1.5.\ 2,, 3, \underbrace{0,0, ...,0}_{P-S}$$

Table 3: MSE values for methods of study for n=100, p= 50, s=20 and $\rho =0.75$

Cut of point	Methods	bandwidth		
		IK	CCT	CV
0.0	Ad Lasso	0.29835	0.28420	0.30183
	Elistic.Net	0.20034	0.18285	0.20343
0.5	Ad Lasso	0.32432	0.31245	0.31917
	Elistic.Net	0.25836	0.25021	0.25797
2.0	Ad Lasso	0.37813	0.32079	0.34137
	Elistic.Net	0.25484	0.21182	0.23132

In the example 3, From Table 3, with n= 100, p= 50, s=20 and $\rho =0.75$, We notice the superiority of our suggested method (Elastic Net) over the adaptive lasso. and that is through MSE, we also note that the best method of the bandwidth is CCT for all cut-off points.
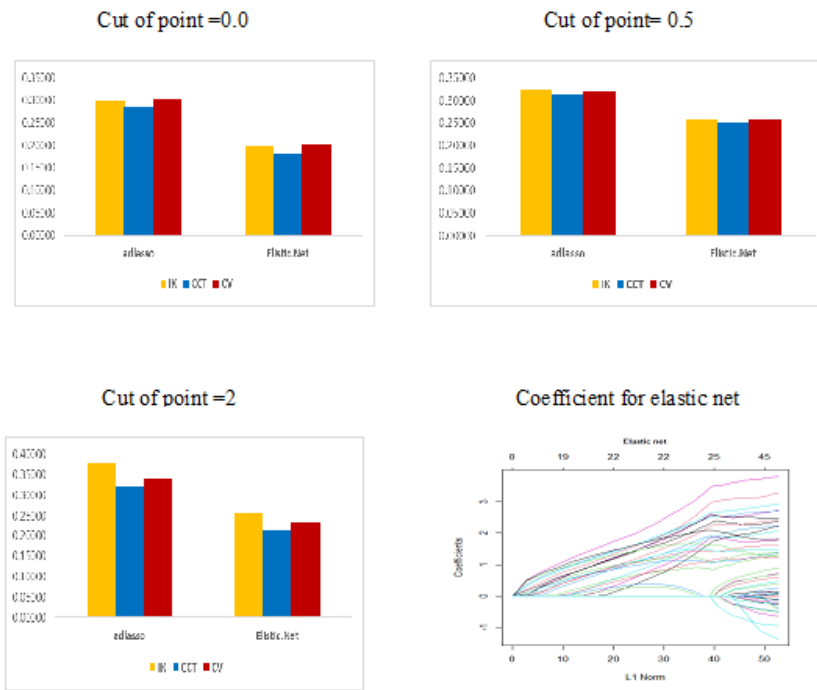
Figure 4: MSE values for methods of study for n=100, p= 15, s=5 and ρ =0.75

Figure (4) shows the preference of our suggested method over the adaptive lasso method in RDD model. According to the (MSE) criterion, we note that the number of important variables appeared far from zero, as is assumed by the simulation through and the figure that shows the features in the method Elastic net.

6.2 Real Data:

The medical data were collected from the Department of Nephrology (Hemodialysis) in Al-Diwaniyah Teaching Hospital. The most important factors that were believed to cause renal failure in the people were selected. This is done with the help of some specialized doctors in the hospital. The data that collected on the cases of kidney failure before and after the patients who underwent the dialysis process, as for the patients who did not undergo the process because of their dependence on medical drugs, their kidney condition was recorded once before the dialysis process. The table below illustrates the most important factors believed to be the cause of kidney failure;

Table (4) Description of variables

y_i	Response variable for patients with kidney disease before and after dialysis
x_1	Presence of kidney stones
x_2	An increase in blood pressure .
x_3	Diabetes.
x_4	patient's age
x_5	Use of analgesic drugs
x_6	Smoking (no smoking)
x_7	Genetics
x_8	Drought
x_9	heart disease
x_{10}	Body fat percentage(BFP)

5.3 Results and discussion:

In this section we applied the the elstic net for variable selection to choose the most important variable that cause the kidney failure. Table 5 shows the coefficient for hae

variable selection model. It clear to see that the variavles x5, x6 ,x7, x8 and x9 are exclouded from the model, were the coefficient for these variable equal to zero.

Table(5) represents the estimated parameters values

| Variable	Estimate
(Intercept)	3.1361
T_r	9.4405
X_l	6.0584
X_r	3.1701
S_{x1}	2.0231
S_{x2}	2.2747
S_{x3}	1.8803
S_{x4}	2.5055
S_{x5}	0.000
S_{x6}	0.000
S_{x7}	0.000
S_{x8}	0.000
S_{x9}	0.000
S_{x10}	2.0841

Table (6) shows a summary of the estimated model statistics are adopted CCT bandwidth.

Variable	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.1361	0.4565	6.870	4.86e-08 ***
T_r	9.4405	0.6900	13.682 7	7.83e-16 ***
X_l	6.0584	7.2614	0.834	0.410
X_r	3.1701	6.9469	0.456	0.651
S_{x1}	2.0231	0.3036	6.664	9.08e-08 ***
S_{x2}	2.2747	0.3075	7.398	9.93e-09 ***
S_{x3}	1.8803	0.2259	8.325	6.54e-10 ***
S_{x4}	2.5055	0.2962	8.459	9.93e-09 ***
S_{x10}	0.51240	0.22024	2.327	0.032605 *

The table above shows estimates of the coefficients for the variables, including them.
T_r : Treatment indicator is whether the patient has received treatment, X_l : Represents (critical point, risk ratio) and represents the degree of the experimental unit from the degrees of the forcing variable, which represents the explanatory variable in the RDD , where it represents the percentage of urea in the blood before the (dialysis) process.)

and X_r : It represents the extent of the effect of treatment on the proportion of urea in the blood. The variables $(s_{x1}\dots s_{x10})$ represent the covariates that have a relationship with the variable (X_1). which parameters were estimated using elastic net method and the figure(5) that shows the cut-off point at score 47. Which represents the critical degree for patients, kidney failure

Figure 5 Treatment effect when the cut-off point=47

RDD CCTE
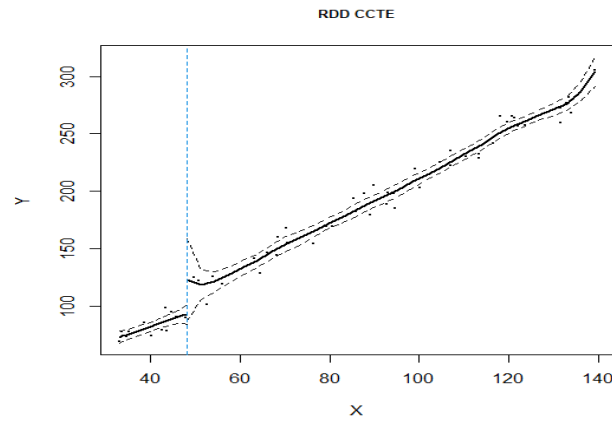
7. Conclusions:

Model(RDD) is used in many economic, social, medical, and other applications, and when this model is combined with one of selecting variable method, its performance and results are acceptable and satisfactory. One of these methods is the Elastic net, which gave good results through (MSE) in the simulation study as well as in the real data. It was reached through real data on some important factors that lead to kidney failure, including, presence of kidney stones, an increase in blood pressure, diabetes, patient's age, use of analgesic drugs, smoking, genetic cause, and Body fat percentage(BFP). We recommend employing some other variable selection methods such as group Lasso, SCAD, PACS, and others with the (RDD).

References:

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

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. Econometrica, 82(6), 2295-2326.

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2018). A Practical Introduction to Regression Discontinuity Designs: Volume II. Cambridge Elements: Quantitative and Computational Methods for Social Science, II, 113..

Hill, R. Carter and Kang-sun Lee. (2001). Performance of Bandwidth Selection Rules for the Local Linear Regression. Departmental Working Papers 2001-10, Department of Economics, Louisiana State University.

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

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. Journal of economic literature, 48(2), 281-355.

Ludwig, J., & Miller, D. L. (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. The Quarterly journal of economics, 122(1), 159-208.

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

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

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

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

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

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