Journal of Techniques, ISSN: 2708-8383, Vol. 5, No. 1, 2023, Pages 203-207
DOI: https://doi.org/10.51173/jt.v5i1.1080

JOURNAL OF TECHNIQUES

Journal homepage: http://journal.mtu.edu.iq

RESEARCH ARTICLE - MANAGEMENT

Comparison Between the Kernel Functions Used in Estimating the Fuzzy Regression Discontinuous Model

Mohammad Jasim Mohammad [1], Sajad Sammer Abd Al-Razzaq [1*]

[1] College of Administration & Economics, University of Baghdad, Baghdad, Iraq

[*] Corresponding author E-mail: Sajad.samir1201a@coadec.uobaghdad.edu.iq

Article Info.

Article history:

Received
21 November 2022

Accepted
28 December 2022

Publishing
04 April 2023

Abstract

Some experiments need to know the extent of their usefulness to continue providing them or not. This is done through the fuzzy regression discontinuous model, where the Epanechnikov Kernel and Triangular Kernel were used to estimate the model by generating data from the Monte Carlo experiment and comparing the results obtained. It was found that the. Epanechnikov Kernel has a least mean squared error.

Publisher : Middle Technical University

Keywords: Fuzzy Regression Discontinuity; Robust Local Polynomial Regression Estimators; Method Mean Squared Error Approximation and Optimal Bandwidth; Monte Carlo Experiment.

1. Introduction

Many educational and economic programs are implemented in our lives that need to be evaluated and studied the usefulness of them to continue offering them or stop them, and the feature of taking the individual to treatment made it difficult to use random experiments in evaluating these programs, so we resort to regression discontinuous designs, and in this paper, we will deal with Fuzzy regression discontinuity(FRD) Specifically, We will estimate the fuzzy regression discontinuous model and compare the kernel functions used in the robust local polynomial method and use the mean square error (MSE) as a standard for comparison.

Used Angrist and Rokkanen fuzzy regression discontinued to study the effects of Boston exam schools for applicants who apply to an area below the cut-off point (acceptance point). These estimates indicate that the causal effects of school enrollment on the exam for applicants in ninth grade who have values that are very far from the cut-off point (acceptance point) differ slightly from those for applicants with values that put them on the fringe of the cut-off point (acceptance point) [1].

Sebastian Calonico et al. (2017) studied the effects of bias correction on confidence intervals in the context of endodontic function density and local polynomial regression estimation. They derived the optimal error bandwidth and discussed the limitations of bandwidth. It was shown that the optimal mean square error bandwidth of the original point estimator provided the lowest mean square error. And they reached important results in the experimental work because they indicate that the bias-corrected confidence intervals and the appropriate standard error have smaller coverage errors and are less sensitive to adjusting the parameter choices. They reached all these results using simulation [2].
Data at different sizes (75, 100, 125, 150) were simulated, and the fuzzy regression discontinuous model was estimated using different bandwidth estimation methods.

2. Materials and Methods

2.1. Regression Discontinuity Designs

The basic idea of regression discontinuous designs (RDD) is that the assignment to a treatment is determined in whole or in part by the value of the explanatory variable $X_i$on on either side of the cut-off point (threshold). It is determined by the explanatory variable because it is not affected by treatment. The result is assumed to be a linear function of the variable x except at the cut-off because it will be discontinued at this point, and therefore any discontinuity in the conditional distribution (or a property of this conditional distribution such as the conditional expectation) of the result is interpreted as a function This explanatory variable at the cut-off value is interpreted as evidence of the causal effect of treatment.

Nomenclature & Symbols			
RDD	regression discontinuous designs	FRDD	fuzzy regression discontinuous model
MSE	mean square error	FRD	fuzzy regression discontinuous

The explanatory variable affects the response variable, so it is measured and its results controlled to determine the causal effect. The objective of the regression discontinuous designs is to study the effect of the treatment variable (W_i) on the outcome variable (W_i). The regression discontinuous designs are of two types:

1. Sharp design: Only people below the threshold will receive treatment and no people above the threshold will receive treatment.
2. Fuzzy design: It is of two types:

a. The first type of fuzzy design, in which some of the treatment group did not attend to receive treatment in a randomized trial, and this case is called [3].
b. The second type of fuzzy design, in which some of the members of the treatment group do not attend to receive treatment and some of the members who are not subject to the treatment process (the comparison group) are banned from treatment and they are called (transition process) [4].

We can write the fuzzy regression discontinuous model (FRDD) as follows: [5]

$$\mathrm{Y_i} = \alpha_0 + \tau W_i + \alpha_1 k_i + \varepsilon_i \tag{1}$$

$\mathrm{Y_i}$: is the response variable (post-trial).
α_0: is the constant limit parameter.
τ: It is a parameter of the causal effect of the treatment.
W_i: It is the classification variable (ordinal variable) and it is the explanatory variable in this model.
α_1: is the parameter of the explanatory variable $\mathrm{X_i}$.
k_i: is $(X_i - c)$.
ε_i: It is the error term of the model, which is represented randomly and distributed naturally with a mean equal to zero and a variance σ^2.

3. Methods of Estimation

3.1. Robust Local Polynomial Regression Estimators

It is one of the nonparametric estimation methods for estimating the bandwidth based on the weighted least squares method, Local polynomial estimation methods are preferred over global polynomial methods to avoid many methodological problems caused by the use of the global polynomial method such as irregular behavior near boundary points, irrational weight, and overfitting [6].

We will estimate the bandwidth using the coverage error optimality method Mean Squared Error Approximation and Optimal Bandwidth [7].

$$\hat{h}_{MSE}(\hat{\varphi}\mathrm{h}) = \hat{C}_{MSE}\, n^{-\frac{1}{2p+3}} \tag{2}$$

$$\hat{C}_{MSE} = \left(\frac{\hat{V}_{MSE}}{2(p+1)\hat{B}_{MSE}^2}\right)^{\frac{1}{2p+3}} \tag{3}$$

Where B_{MSE} is the bias and V_{MSE} is the variance.
And to estimate the variance

$$\hat{V} = \hat{\sigma}_{Yr}^2 + \hat{\sigma}_{Yl}^2 + \hat{\tau}^2\left(\hat{\sigma}_{Wr}^2 + \hat{\sigma}_{Wl}^2\right) - 2\hat{\tau}\left(\hat{\sigma}_{YWr} + \hat{\sigma}_{YWl}\right) \tag{4}$$

$$\hat{\sigma}_{Yr}^2 = \frac{1}{n_{1r}-1}\sum_{i:c\le X_i\le c+h_1}(Y_i - \bar{Y}_r)^2 \tag{5}$$

$$\hat{\sigma}_{Yl}^2 = \frac{1}{n_{1l}-1}\sum_{i:c-h_1\le X_i<c}(Y_i - \bar{Y}_l)^2 \tag{6}$$

$$\hat{\sigma}_{Wr}^2 = \frac{1}{n_{1r}-1}\sum_{i:c\le X_i\le c+h_1}(W_i - \bar{W}_r)^2 \tag{7}$$

$$\hat{\sigma}_{Wl}^2 = \frac{1}{n_{1l}-1}\sum_{i:c-h_1\le X_i<c}(W_i - \bar{W}_l)^2 \tag{8}$$

$$\hat{\sigma}_{YWr} = \frac{1}{n_{1r}-1}\sum_{i:c\le X_i\le c+h_1}(Y_i - \bar{Y}_r)(W_i - \bar{W}_r) \tag{9}$$

$$\hat{\sigma}_{YWl} = \frac{1}{n_{1l}-1}\sum_{i:c-h_1\le X_i<c}(Y_i - \bar{Y}_l)(W_i - \bar{W}_l) \tag{10}$$

$$n_{1r} = \sum_{i=1}^{n} 1_{c\le X_i\le c+h_1} \tag{11}$$

$$n_{1l} = \sum_{i=1}^{n} 1_{c-h_1\le X_i<c} \tag{12}$$

$$\bar{Y}_r = \frac{1}{n_{1r}}\sum_{i:c\le X_i\le c+h_1} Y_i \tag{13}$$

$$\bar{Y}_l = \frac{1}{n_{1l}}\sum_{i:c-h_1\le X_i<c} Y_i \tag{14}$$

$$\bar{W}_r = \frac{1}{n_{1r}}\sum_{i:c\le X_i\le c+h_1} W_i \tag{15}$$

$$\bar{W}_l = \frac{1}{n_{1l}}\sum_{i:c-h_1\le X_i<c} W_i \tag{16}$$

We will use Silverman's rule to calculate the experimental bandwidth to calculate the intensity and contrast at c [8].

$$h_1 = \left(\frac{4S_X^{(2p+3)}}{3n}\right)^{\frac{1}{(2p+3)}} \quad (17)$$

Where S_x represents the variance of x at the cut-off point.

If we assume that the degree (p = 1)

$$h_1 = \left(\frac{4S_X^5}{3n}\right)^{\frac{1}{5}} \quad (18)$$

$$h_1 \approx 1.06 S_X n^{-\frac{1}{5}} \quad (19)$$

This depends on the normal kernel function and the normal density function. We will modify it to the uniform kernel function of [-1,1] and a normal density function, and we will get

$$h_1 = 1.84 S_X n^{-\frac{1}{5}} \quad (20)$$

$$S_X^2 = \frac{\sum(x_i-\bar{X})^2}{n-1} \quad (21)$$

And to estimate the bias

$$\hat{B} = \hat{f}(c)\left(\left(\hat{m}_{Yr}^{(2)} - \hat{m}_{Yl}^{(2)}\right) - \hat{\tau}\left(\hat{m}_{Wr}^{(2)} - \hat{m}_{Wl}^{(2)}\right)\right) \quad (22)$$

$$\hat{f}(c) = \frac{n_{1l}+n_{1r}}{2nh_1} \quad (23)$$

o estimate $\hat{m}_{Yr}^{(2)}$, $\hat{m}_{Yl}^{(2)}$,$\hat{m}_{Wr}^{(2)}$, $\hat{m}_{Wl}^{(2)}$ we will write a quadratic polynomial regression equation for Y_i and W_i, And the equation for the regression of Y_i to the left side when X_i<c will be as follows:

$$Y_{il} = \theta_{0l} + \theta_{1l} 1_{X_i<c} + \theta_{2l}(X_i - c) + \theta_{3l}(X_i - c)^2 + \varepsilon_i \quad (24)$$

And the equation for the regression of Y_i to the right side when X_i≥c is as follows:

$$Y_{ir} = \theta_{0r} + \theta_{1r} 1_{X_i \geq c} + \theta_{2r}(X_i - c) + \theta_{3r}(X_i - c)^2 + \varepsilon_i \quad (25)$$

And the regression equation W_i for the left side when X_i<c will be as follows:

$$W_{il} = \delta_{0l} + \delta_{1l} 1_{X_i<c} + \delta_{2l}(X_i - c) + \delta_{3l}(X_i - c)^2 + \varepsilon_i \quad (26)$$

And the regression equation W_i for the right side when X_i≥c is as follows:

$$W_{ir} = \delta_{0r} + \delta_{1r} 1_{X_i \geq c} + \delta_{2r}(X_i - c) + \delta_{3r}(X_i - c)^2 + \varepsilon_i \quad (27)$$

And to get $\hat{m}_{Yr}^{(2)}$ we will find the second derivative of the second moment of equation (25):

$$\hat{m}_{Yr}^{(2)} = 2\hat{\theta}_{3r} \quad (28)$$

And to get $\hat{m}_{Yl}^{(2)}$we will find the second derivative of the second moment of equation (24):

$$\hat{m}_{Yl}^{(2)} = 2\hat{\theta}_{3l} \quad (29)$$

And to get $\hat{m}_{Wr}^{(2)}$we will find the second derivative of the second moment of equation (27):

$$\hat{m}_{Wr}^{(2)} = 2\hat{\delta}_{3r} \quad (30)$$

And to get $\hat{m}_{Wl}^{(2)}$we will find the second derivative of the second moment of equation (26):

$$\hat{m}_{Wl}^{(2)} = 2\hat{\delta}_{3l} \quad (31)$$

After estimating the bandwidth, we will estimate the rest of the parameters through the weighted least squares method, as follows:

$$(\hat{\alpha}_{0l}, \hat{\alpha}_{1l}) = \min_{\alpha_1} \sum_{i=1}^{n} 1(X_i < c)\left(Y_i - \alpha_{0l} - \alpha_{1l}(X_i - c)\right)^2 K_h\left(\frac{X_i-c}{h}\right) \quad (32)$$

$$(\hat{\alpha}_{0r}, \hat{\alpha}_{1r}) = \min_{\alpha_1} \sum_{i=1}^{n} 1(X_i \geq c)\left(Y_i - \alpha_{0r} - \alpha_{1r}(X_i - c)\right)^2 K_h\left(\frac{X_i-c}{h}\right) \quad (33)$$

$$\left(\hat{\beta}_{0l}, \hat{\beta}_{1l}\right) = \min_{\beta_1} \sum_{i=1}^{n} 1(X_i < c)\left(W_i - \beta_{0l} - \beta_{1l}(X_i - c)\right)^2 K_h\left(\frac{X_i-c}{h}\right) \quad (34)$$

$$\left(\hat{\beta}_{0r}, \hat{\beta}_{1r}\right) = \min_{\beta_1} \sum_{i=1}^{n} 1(X_i \geq c)\left(W_i - \beta_{0r} - \beta_{1r}(X_i - c)\right)^2 K_h\left(\frac{X_i-c}{h}\right) \quad (35)$$

K_h (.): It is a kernel function. [9]

we can estimate the average causal effect as follows:

$$\hat{\tau} = \frac{\hat{\tau}_Y}{\hat{\tau}_W} \quad (36)$$

$$\hat{\tau}_Y = \hat{\mu}_{Yr} - \hat{\mu}_{Yl} \quad (37)$$

$$\hat{\tau}_W = \hat{\mu}_{Wr} - \hat{\mu}_{Wl} \quad (38)$$

Two kernel functions will be used:[10]

1. Epanechnikov kernel

$$k(u)\begin{cases}\frac{3}{4}(1-u^2) & |u|\le 1\\ 0 & o.w\end{cases} \tag{39}$$

2. Triangular kernel

$$k(u)=\begin{cases}(1-|u|) & |u|\le 1\\ 0 & o.w\end{cases} \tag{40}$$

4. Discussion of Results

4.1. Simulation

The Monte Carlo experiment will be used to generate the data as follows:

$$y_i = w_i + \varepsilon_{y_i} \quad , i = 1, \dots, n \tag{41}$$

The treatment is defined as follows:

$$w_i = 1\{\varepsilon_{x_i} \le 0\} \times 1\{x_i < 0\} + 1\{\varepsilon_{x_i} \le 0\} \times 1\{x_i \ge 0\} \tag{42}$$

The errors ε_{y_i} and ε_{x_i} are distributed in the common normal distribution as follows:

$$\begin{pmatrix}\varepsilon_{y_i}\\ \varepsilon_{x_i}\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho_1\\ \rho_2 & 1\end{pmatrix}\right) \tag{43}$$

The independent variable has a normal distribution as follows:

$$x_i \sim N(0,1) \tag{44}$$

And that ρ_1 and ρ_2 are hypothetical values, and if we assume that the covariance and covariance matrix of errors is the same, we will assume the values as follows in Table 1.

Table 1. Values (ρ_1,ρ_2)

$\boldsymbol{\rho_1}$	$\boldsymbol{\rho_2}$
0.75	**0.75**
0.5	**0.75**
0.75	**0.99**
0.99	**0.5**

We will repeat the generation 1000 times and the results will be calculated [11].
The program (R- program) was used in writing the program for data analysis, Table 2.

Table 2. Mean Saquer Error of Model

N	$\boldsymbol{\rho_1 = \rho_2 = 0.75}$		$\boldsymbol{\rho_1 = 0.5, \rho_2 = 0.75}$		$\boldsymbol{\rho_1 = 0.75, \rho_2 = 0.99}$		$\boldsymbol{\rho_1 = 0.99, \rho_2 = 0.5}$	
	Triangular	**Epanechnikov**	**Triangular**	**Epanechnikov**	**Triangular**	**Epanechnikov**	**Triangular**	**Epanechnikov**
75	0.6027421	0.5909573	0.7869537	0.7707444	0.6027421	0.5909573	0.3539588	0.3472068
100	0.6160888	0.6062357	0.8049311	0.7932837	0.6160888	0.6062357	0.3627386	0.3563045
125	0.6198637	0.6155493	0.816737	0.8095917	0.6198637	0.6155493	0.3630471	0.3594187
150	0.6244743	0.6204156	0.8187302	0.8129859	0.6244743	0.6204156	0.3657737	0.3625651

5. Conclusion

We notice from the simulation results that the Epanechnikov Kernel is better than the Triangular Kernel because it has the least mean squares of error. An increase in the mean squares of error occurs when the sample size increases because of the increase in values outside the bandwidth.

Acknowledgement

I extend my thanks and appreciation to the teacher at the University of Baghdad in the College of Administration & Economics in the Department of Statistics, to all the teaching staff in the Statistics Department at the University of Baghdad, and to everyone who contributed to this research.

Reference

[1] Angrist and Rokkanen. (2015).” Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away From the Cutoff”. Journal of the American Statistical Association.Vol.110,No.512,PP 1331-1344.

[2] Cattaneo, M. D., Jansson, M., Ma, X.(2017).” Simple Local Polynomial Density Estimators”. Journal of the American Statistical Association. 115(531) PP1449-1455.

[3] Bloom, B. S. (1984). “The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring”. Educational Researcher. Vol.13, No.6, PP. 4-16.

[4] Bloom et al.(1997).” The Benefits and Costs of JTPA Title II-A Programs: Key Findings from the National Job Training Partnership Act Study”. The Journal of Human Resources. Vol.32, No.3, PP. 549-576.

[5] Jacob, R. et al.(2012).” A Practical Guide to Regression Discontinuity”. MDRC.

[6] Fan, J. and Gijbels, I. (1996).” Local Polynomial Modelling and Its Applications”. Chapman and Hall, London. Introductory Econometrics (4th Edition) A Modern Approach, by Jeffrey M. Wooldridge Hardcover, 888 Pages, Published 2008 by South-Western.

[7] Calonico, S., M. D. Cattaneo, and M. H. Farrell. (2017). “On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference”. Journal of the American Statistical Association. Vol.113, No.522, PP. 767-779.

[8] Silverman, B.W. (1986).” Density Estimation for Statistics and Data Analysis”.Chapman & Hall, London. 30(7) PP 876-877.

[9] Gelman and Imbens. (2019)." Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs”. Journal of Business & Economic Statistics. Vol.37, No3, PP. 447-456.

[10] Epanechnikov, V.A. (1969). “Non-parametric estimation of a multivariate probability density”. Theory of Probability and its Applications. Vol.14, No.1, PP. 153-158.

[11] Rong Ma et al.(2016).” Robust Inference in Fuzzy Regression Discontinuity with Multiple Forcing Variables”. Central University of Finance and Economics.