On the Adaptive Nadaraya-Watson Kernel Estimator for the Discontinuity in the Presence of Jump Size

Muhammad HANIF1, Shabnam SHAHZADI2, Usman SHAHZAD3, Nursel KOYUNCU∗4

1,2,3 PMAS-Arid Agriculture University, Department of Mathematics and Statistics, Rawalpindi, PAKISTAN
4 Hacettepe University, Faculty of Science, Department of Statistics, Ankara, TURKEY

(Alınış / Received: 01.08.2017, Kabul / Accepted: 30.04.2018, Online Yayınlanma / Published Online: 12.06.2018)

Abstract: In this paper, we studied an Adaptive Nadaraya Watson kernel estimator to check the bias effect on both side of the discontinuity in the presence of jump size for regression discontinuity model. We have proposed the modified Adaptive Nadaraya Watson kernel estimator and derived its normality and variance. We have also compared with the asymptotic normality of the Mean Integrated Square Error (MISE) of Adaptive Nadaraya Watson kernel estimator and Nadaraya Watson kernel estimator. The results obtained from the simulation study have showed that Adaptive Nadaraya Watson estimator has better performance than the Nadaraya Watson Kernel estimator.

Keywords
Adaptive Nadaraya Watson, Jump size, Mean Integrated Square Error, Rate of convergence.

1. Introduction

Nonparametrics: For last few years, nonparametric regression has become important tool for data smoothing. Most commonly used estimates of nonparametric regression functions including kernel estimates based on smoothness of the regression functions. In many applications, the function is to be estimated as discontinuities or threshold points. Gao, et al.[1], for example, when studying the impact of advertising, the time at which this action takes place impact could effectively be modeled by the location of a jump point and the magnitude of the effect of this action is measured by the jump size. If we ignore the jump point we make serious error in order to draw the inference about the processes under study. Similarly also see Yin [2] for estimating locations of discontinuity points of the regression function.

Regression discontinuity design becomes one of the useful designs when there is threshold point in the treatment or in the probability assignment. The treatment assignment under the weak smoothness becomes random near the threshold point. Regression Discontinuity model is mostly used only for the information that is very close to threshold point. In Regression Discontinuity model lacks of smoothness is not only the problem but the size of discontinuity is also important. It is useful to estimate the conditional expectation of the boundary points to check the difference of results given by boundary estimation (see Porter [3]).

Demir and Toktamis [4] considered Adaptive Nadaraya Watson kernel estimator to estimate the regression function. Our paper is also composed of Adaptive Nadaraya Watson as a jump size instead of Nadaraya Watson. Because the main drawback of the Nadaraya Watson estimator used as jump size in regression discontinuity model has poor asymptotic bias behavior whereas Adaptive Nadaraya Watson overcomes this problem.

Variance estimation of regression discontinuity was constructed by Hardle [5]. Pagan and Ullah [6] also gave the consistent estimation of variance for the left hand limit i.e. \( \sigma^2 - \) and right hand limit i.e. \( \sigma^2 + \) of the derivatives. Similarly Silverman [7] gave density of the discontinuity of the jump size, \( f_0(\tilde{u}) \) which can be estimated consistently...
by kernel density estimation.

In this paper, we have proposed Adaptive Nadaraya Watson kernel estimator for the estimation of \( \sigma_1^2 \), \( \sigma_2^2 \) for jump size in regression discontinuity model. Further, we have derived the asymptotic properties of \( \sigma_1^2 \), \( \sigma_2^2 \). Marginal and hence optimal rate of convergence is derived theoretically to estimate the treatment effect in regression discontinuity model by using Stone’s [8] definition. The properties such as MSE, MISE and bias effect on the left and right hand side of the discontinuity of regression discontinuity model are also derived. The performance of proposed regression discontinuity model is compared with the existing once numerically.

The rest of the paper is organized as follows: Section 2 includes Nadaraya Watson kernel estimator as a jump size in the regression discontinuity model and technical lemmas are discussed. In section 3, proposed work based on Adaptive Nadaraya Watson kernel estimator and its assumptions are discussed. Simulation study is performed in Section 4. Real life data is used to see performance of proposed estimator in Section 5. The derivations of the Section 3 and Section 4 are available in Appendix .

2. Nadaraya Watson Estimator and Technical Lemmas

As the previous work is based on Nadaraya Watson kernel estimator used as a jump size in regression discontinuity model. The connection between the regression discontinuity model and the treatment effect have been discussed in Hahn et al.[9]. Trochim [10] distinguished among the two dissimilar connections between the Regression Discontinuous model, that depends upon the treatment assignment connected with the observed variable. The treatment assignment given by the indicator variables \( d\epsilon(0,1) \) are defined as;

\[
\begin{cases}
1 \text{ if } u > \bar{u} \\
0 \text{ if } u < \bar{u}.
\end{cases}
\]

where \( u \) is observed variable with the known threshold point \( \bar{u} \). Let \( y_0 \) and \( y_1 \) be the potential outcomes parallel to the two treatment assignments, and as usual \( Y = d\gamma_1 + (1 - d)\gamma_0 \) is the observed outcome. By using the smoothness assumption that \( E(\frac{Y}{u}) \) is continuous at \( \bar{u} \) for \( f = 0,1 \). So the expected casual effect of the treatment effect can be identified at the discontinuity point.

\[
\alpha = \lim_{u \rightarrow \bar{u}} E(y/u) - \lim_{u \rightarrow \bar{u}} E(Y/u)
\]

\[
\hat{\alpha}(\bar{u}) = \frac{1}{n}\sum_{i=1}^{n} K_h(\bar{u} - U_i)y_i - \frac{1}{n}\sum_{i=1}^{n} K_h(\bar{u} - U_i)\gamma_i(1 - d_i) - \frac{1}{n}\sum_{j=1}^{n} K_h(\bar{u} - U_j)y_j(1 - d_j) - \frac{1}{n}\sum_{j=1}^{n} K_h(\bar{u} - U_j)\gamma_j(1 - d_j)
\]

Following Lamas are useful in obtaining the main results of our paper (see Porter [3]).

**Lemma A1.** Suppose the kernel \( k \) is bounded, symmetric, zero outside of a bounded set \( M \) and Lipschitz, and \( f_0 \) is continuous on \( N_1 \).

If \( \frac{nh}{\ln n} \rightarrow \infty \), then \( \sup_{x \in N_0} |\bar{f}_+(x) - Ef_+(x)| = O_P(\sqrt{\frac{\ln n}{nh}}) \), \( \sup_{x \in N_0} |f_+(x) - Ef_+(x)| = O_P(\sqrt{\frac{\ln n}{nh}}) \), and \( \sup_{x \in N_0} |\hat{f}(x) - Ef(x)| = O_P(\sqrt{\frac{\ln n}{nh}}) \).

**Lemma A2.** Suppose the kernel \( k \) is bounded, symmetric, zero outside of a bounded set \( M \) and Lipschitz, \( f_0 \) is a continuous on \( N_1 \), \( \sup_{x \in N_0} |f_+(x) - Ef_+(x)| < \infty \) for some \( \xi > 0 \) and \( \sup_{x \in N_0} |f(x) - Ef(x)| = O_P(\sqrt{\frac{\ln n}{nh}}) \).

**Lemma A3.** Suppose the kernel \( k \) is bounded, symmetric, zero outside of a bounded set \( [-M, M] \). On \( N_1 \), \( f_0 \) is continuously differentiable for \( x \neq \bar{x} \), \( x \in N_1 \), and \( m \) is continuous at \( \bar{x} \) with finite right and left hand derivatives. Then \( \sup_{u \in \mathcal{E}^{[\bar{x}, \bar{x} + M]}} |\hat{f}(x) - r(x)| = O(h) \).

3. Proposed Estimator and Assumptions

The main purpose of our study is to minimize the bias effect on the left and right side of the discontinuity by using Adaptive Nadaraya Watson Kernel estimator as a jump size. We also obtained the optimal rate of convergence through simulation study.

Consider the random-design regression model given by;

\[
Y_i = m(u_i) + \varepsilon_i
\]

where \( m \) is an unknown regression function with compact interval \([0,1]\) and \( \varepsilon_i \) is the observation error which is independent identically distributed with mean 0 and variance \( \sigma^2 \). In the discontinuity model we have a cut off point existed for \( m \) function whereas cut off point (\( \bar{u} \)) is 0 < \( \bar{u} < 1 \). Usually regression function is defined as;

\[ y = m(u) + \alpha u + \epsilon \]

and

\[ 0 < u < 1 \]

Here, \( m(u) \) represents a continuous function defined on \([0,1]\). \( \alpha \)=jump size and can be defined as

\[ \alpha(\bar{u}) = m_2(\bar{u}) - m_1(\bar{u}) \]

where \( \bar{u} \) is jump size at cut off point. Basically the jump size at possible cut off point \( \bar{u} \) is \( m_1(\bar{u}) = \lim_{u \rightarrow \bar{u}} m(u) \) is at the right of the discontinuity curve and \( m_2(\bar{u}) = \lim_{u \rightarrow \bar{u}} m(u) \) is at the left side of the discontinuity curve.

\[ \alpha = E[|y_1 - y_0|/u] \]

\[ = E[y/\bar{u}] - E[y/0|/u] \]

\[ = \lim_{u \rightarrow \bar{u}} E[y_1/\bar{u}] - \lim_{u \rightarrow 0|/u} E[y_0/\bar{u}] \]

\[ \lim_{u \rightarrow \bar{u}} E[y/\bar{u}] - \lim_{u \rightarrow \bar{u}} E[y/\bar{u}] \]

where

\[ \alpha(\bar{u}) = \frac{1}{n} \sum_{i=1}^{n} K_h_+(-U_i)\gamma_i(1 - d_i) - \frac{1}{n} \sum_{i=1}^{n} K_h_+(-U_i)\gamma_i(1 - d_i) \]

\[ + \frac{1}{n} \sum_{j=1}^{n} K_h_+(-U_j)y_j(1 - d_j) - \frac{1}{n} \sum_{j=1}^{n} K_h_+(-U_j)y_j(1 - d_j) \]

\[ + \frac{1}{n} \sum_{j=1}^{n} K_h_+(-U_j)\gamma_j(1 - d_j) - \frac{1}{n} \sum_{j=1}^{n} K_h_+(-U_j)\gamma_j(1 - d_j) \]
Assumption 1

\[ \lambda_i = \left[ \frac{\hat{f}(u)}{\bar{u}} \right]^{0.5} \]

and \( d = 1(u_i \geq \bar{u}) \). \( I(A) \) is an indicator for event \( A \). \( u_i \) represents a random variable and \( d \) is an indicator whereas \( h \lambda_i \) shows adaptive bandwidth that is used to control the size of local neighborhood on average. First term in the \( \hat{\alpha}(\bar{u}) \) is a weighted average depends on the distance from the discontinuity \((\bar{u} - U_i)\) whereas \( \alpha \) is the jump size of the discontinuity model.

The cut-off point \( \bar{u} \) provides a chance to observe the average difference in the potential outcomes from the points on either side of the discontinuity. The main point of our estimation is that in all cases, the casual effect is found out from zero, \( m \) unequal left and right side derivatives of \( p \), compact interval \( \lambda \) and \( \lambda \geq 1 \) minus discontinuity. So, \( \hat{\lambda}_i = \left[ \frac{\hat{f}(U_i)}{\bar{u}} \right] - \alpha \)

and \( \alpha = 0.5 \)

This idea is formalized in the following theorem, which is based on the limiting distribution of the adaptive Nadaraya Watson estimator at a boundary point.

**Theorem 1.** By using assumption 1(a), 2(a) and assumption 3 holds \( l_q \) with any positive integer and \( l_r \) with any negative integer. If \( h \lambda_i \to 0, nh \lambda_i \to \infty \) and \( h \lambda_i \sqrt{n} \to p \) where \( 0 \leq p \leq \infty \)

\[ \sqrt{nh} \hat{\lambda}_i(\bar{u} - \alpha) \xrightarrow{d} (p2k_1(0)(m^+(\bar{u}) + m^-(\bar{u})),4d_0 \frac{\sigma^2(\bar{u}) + \sigma^2(\bar{u})}{f_0(\bar{u})} \]

**Remarks.**

Observations used in the first and second terms of the difference defining the \( \hat{\alpha} \) are independent. For Adaptive Nadaraya Watson estimator at its interior point shows that its asymptotic bias is written in a form that underscores its dependence on the rate with adaptive bandwidth approaches to zero. When \( p = 0 \) we are "under smoothness" and the asymptotic bias is zero. When \( h \lambda_i \sim (n^{-4/5}) \) and, \( \hat{\alpha} \) achieves its fastest rate of convergence at \( (n^{-4/5}) \) than the asymptotic bias term is considered. From this, we see that the bias of \( \hat{\alpha} \) is of order \( o(h \lambda_i) \). Whereas, from the theorem we find that higher-order bias-reducing kernels do not affect the order of asymptotic bias. Also, left and right hand derivatives of the cut-point are equal or not equal do not affect the order of the asymptotic bias.

**Theorem 2.** Suppose Assumption 1(a), 2(a), and 3 with \( l_q \) and \( l_r \) are positive integers and \( \xi \geq 2 \). If \( \frac{n \lambda_i}{\lambda_i} \to \infty \) and \( \hat{\alpha} \to p \), then \( \hat{\delta}^2(\bar{u}) \to p \) \( \delta^2(\bar{u}) \) and \( \hat{\delta}^2(\bar{u}) \to p \) \( \delta^2(\bar{u}) \).

4. Simulation Study

We have performed practical strategy for Adaptive Nadaraya Watson and Nadaraya Watson kernel estimator to provide simulation evidence for finite sample performance. Hence our objective is to estimate the discontinuity at particular point, for that we use unbiased cross validation which was proposed by Hall and Schucany [11] for the density estimation. Simulation study is used to compare the performance of the Nadaraya Watson kernel estimator and Adaptive Nadaraya Watson kernel estimators. For the simulation study we consider the regression discontinuity function.

\[ y = m(u) + \alpha d + \epsilon \]
5. Real Life Data

Data is taken from Gross Domestic Product, Current Prices. Values are based upon GDP in national currency converted to U.S. dollars using market exchange rates (yearly average). Exchanges rate projections are provided by country economics for the group of other emerging market and developing countries. Exchanges rates of advanced economies are established in the WEO assumptions for each WEO exercise. The data is summarized in Table3. (Source of Data:International Monetary Fund, World Economic Outlook Database, April 2015)

6. Conclusion

In this study, we concluded that the results obtained from the Adaptive Nadaraya Watson Kernel estimator gives better result than the Nadaraya Watson estimator used in discontinuity model. We showed it by finding Mean Square Error and Rate of convergence. From Table 4 and 5 we can see that the MSE of proposed estimator is less than the MSE of existing estimator. We can also check the rate of convergence of proposed Adaptive Nadaraya Watson estimator faster than existing Nadaraya Watson estimator.

References


Appendix

Proof. To estimate the Adaptive Nadaraya Watson estimator we express \( m(\bar{u}) \) in term of probability density function pdf \( f(u,y) \). As we have

\[
f(x) = \frac{1}{nh_{\lambda_i}^5} K\left(\frac{u-U_i}{h_{\lambda_i}^5}\right)
\]

whereas;

\[
\lambda_i^5 = \left[\frac{f(u)}{\bar{x}}\right]^{0.5}
\]

Where \( \alpha \) is sensitivity parameter which varies between (0, 1) or we can write it as \( 0 < \alpha < 1 \). Here we take \( \alpha = 0.5 \) As, we have

\[
m(\bar{u}) = E[Y|U=\bar{u}] = \int_{-\infty}^{+\infty} y f(y|u) dy = \frac{\int_{-\infty}^{+\infty} y f(u,y) dy}{\int_{-\infty}^{+\infty} f(u) du}
\]

\[
\hat{m}(\bar{u}) = \frac{\int_{-\infty}^{+\infty} y \hat{f}(u,y) dy}{\int_{-\infty}^{+\infty} \hat{f}(u) du}
\]

\[
= \frac{1}{f(u)} \int_{-\infty}^{+\infty} y \hat{f}(u,y) dy
\]

\[
= \frac{1}{f(u)} \int_{-\infty}^{+\infty} y \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{nh_{\lambda_i}^5} K\left(\frac{u-U_i}{h_{\lambda_i}^5}\right) K\left(\frac{y-Y_i}{h_{\lambda_i}^5}\right) dy
\]

\[
= \frac{1}{f(u)} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{nh_{\lambda_i}^5} k\left(\frac{u-U_i}{h_{\lambda_i}^5}\right) \int_{-\infty}^{+\infty} \frac{y}{nh_{\lambda_i}^5} \int_{-\infty}^{+\infty} \frac{y}{nh_{\lambda_i}^5} dy
\]

By using the transformation \( t = \frac{y-Y_i}{nh_{\lambda_i}^5} \) which is given as;

\[
\hat{m}(\bar{u}) = \frac{1}{f(x)} \sum_{i=1}^{n} Y_i k\left(\frac{u-U_i}{h_{\lambda_i}^5}\right) \int_{-\infty}^{+\infty} \frac{y}{nh_{\lambda_i}^5} \int_{-\infty}^{+\infty} \frac{y}{nh_{\lambda_i}^5} dy
\]
By using the property of kernel that is;
\[ \int_{-\infty}^{\infty} k(v)dv = 1, \int_{-\infty}^{\infty} vk(v)dv = 0 \]

Hence
\[ \hat{m}(\bar{u}) = \frac{1}{nh} \sum_{i=1}^{n} k(\frac{u - U_i}{h\hat{\lambda}^*})[h\hat{\lambda}^* + Y_i] \]

After simplification we have;
\[ = \frac{1}{f(u)} \sum_{i=1}^{n} \frac{1}{n} k(\frac{u - U_i}{h\hat{\lambda}^*})Y_i \]

By replacing in equation (5.1) we have;
\[ \hat{\lambda}^* = \left[ \frac{\hat{f}(u)}{\hat{\lambda}} \right]^{-0.5} \]
\[ \hat{m}(\bar{u}) = \frac{\sum_{i=1}^{n} \frac{1}{nh} k(\frac{u - U_i}{h\hat{\lambda}^*})}{\sum_{i=1}^{n} \frac{1}{nh} k(\frac{u - U_i}{h\hat{\lambda}^*})} \]

Hence, we obtain the Adaptive Nadaraya Watson kernel regression estimator as: By simplify we have;
\[ \hat{m}(\bar{x}) = \frac{\sum_{i=1}^{n} \frac{1}{nh} k(\frac{x - U_i}{h\hat{\lambda}^*})}{\sum_{i=1}^{n} \frac{1}{nh} k(\frac{x - U_i}{h\hat{\lambda}^*})} \]

Replacing it in equation (5.3) we have estimated jump size is
\[ \hat{\alpha}(\bar{u}) = \frac{1}{nh\hat{\lambda}^*} \sum_{i=1}^{n} K_{\hat{\lambda}^*} \left( \bar{u} - U_i \right)Y_i \]
\[ \frac{1}{nh\hat{\lambda}^*} \sum_{j=1}^{n} K_{\hat{\lambda}^*} \left( \bar{u} - U_j \right)d_j \]

Proof of Theorem 1.
Let g denote a positive generic constant and as we know M is a compact interval. Let suppose M_0 also be compact interval such that \( u \epsilon int(M_0) \) and \( M_0 \subset int(M) \) and by using assumption 1(a), suppose we support the kernel k is [-M,M]. Hence the observation just to the right of the discontinuity are more likely to be greater than the intercept \( m(\bar{u}) + \alpha \) giving upward biased. similarly, an average of observations just to the left of the discontinuity would provide a downward biased estimate \( m(\bar{u}) \).

We have
\[ \hat{\alpha}(\bar{u}) = (m(\bar{u}) + \alpha) \]
\[ \frac{1}{n} \sum_{i=1}^{n} \frac{1}{nh\hat{\lambda}^*} k(\frac{u - U_i}{h\hat{\lambda}^*}) \left( \bar{u} - U_i \right)Y_i \]
\[ \frac{1}{n} \sum_{j=1}^{n} \frac{1}{nh\hat{\lambda}^*} k(\frac{u - U_j}{h\hat{\lambda}^*}) \left( \bar{u} - U_j \right)d_j \]

By replacing
\[ Y_i = m(u_i) + e_i \]

And rearranging the above equation, we have
\[ \hat{\alpha} - \alpha = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{nh\hat{\lambda}^*} k(\frac{u - U_i}{h\hat{\lambda}^*}) \left( \bar{u} - U_i \right)Y_i \]
\[ \frac{1}{n} \sum_{j=1}^{n} \frac{1}{nh\hat{\lambda}^*} k(\frac{u - U_j}{h\hat{\lambda}^*}) \left( \bar{u} - U_j \right)d_j \]

Multiple both sides with \( \sqrt{n\hat{\lambda}^*} \), we have \( n\hat{\lambda}^* (\hat{\alpha} - \alpha) \);
\[
\frac{1}{n} \sum_{i=1}^{n} \frac{1}{h \lambda_j} k\left(\frac{\bar{d}_i - u}{h \lambda_j}\right) (1 - d_j) [m(\bar{u}_i) - m(\bar{u}) + e_i].
\]

Now, by taking the denominator of the first term, we have
\[
\frac{1}{n} \sum_{j=1}^{n} \frac{1}{h \lambda_j} k\left(\frac{\bar{d}_i - u_j}{h \lambda_j}\right) (1 - d_j).
\]

Hence, we know that
\[
\frac{1}{n} \sum_{j=1}^{n} \frac{1}{h \lambda_j} k\left(\frac{\bar{u}_i - u_j}{h \lambda_j}\right) (1 - d_j).
\]

Similarly, we do with second term of equation (5.3), by taking their denominator;
\[
\frac{1}{n} \sum_{j=1}^{n} \frac{1}{h \lambda_j} k\left(\frac{\bar{u}_i - u_j}{h \lambda_j}\right) (1 - d_j).
\]

Taking variance on both sides, we have
\[
\text{var}\left[\frac{1}{n} \sum_{j=1}^{n} \frac{1}{h \lambda_j} k\left(\frac{\bar{d}_i - u_j}{h \lambda_j}\right) (1 - d_j)\right] \leq \frac{1}{(nh \lambda_j)^2} E[k^2\left(\frac{u_i - \bar{u}}{h \lambda_j}\right)]
\]

Hence, we have
\[
= \frac{1}{(nh \lambda_j)^2} \int_{0}^{M} k^2(v) f_0(\bar{u} + h \lambda_j v) dv
\]

By simplification, we have
\[
= O\left(\frac{1}{nh \lambda_j}\right) = o(1)
\]

Then, by using Chebyshev’s Inequality the general formula is \(1 - 1/SD^2\).

\[
\frac{1}{n} \sum_{j=1}^{n} \frac{1}{h \lambda_j} k\left(\frac{\bar{u}_i - u_j}{h \lambda_j}\right) (1 - d_j) = E\left[\frac{1}{n} \sum_{j=1}^{n} \frac{1}{h \lambda_j} k\left(\frac{\bar{u}_i - u_j}{h \lambda_j}\right) (1 - d_j)\right] + \alpha_p(1)
\]

By using the transformation, we have;
\[
= f_0(\bar{u}) \int_{0}^{M} k(v) dv + \alpha_p(1)
\]

\[
= f_0(\bar{u}) + \alpha_p(1)
\]

\[
= \frac{1}{n} \sum_{j=1}^{n} \frac{1}{h \lambda_j} k\left(\frac{u_i - \bar{u}}{h \lambda_j}\right) (1 - d_j) + \alpha_p(1)
\]

Hence, we know that
\[
\frac{1}{n} \sum_{j=1}^{n} \frac{1}{h \lambda_j} k\left(\frac{\bar{u}_i - u_j}{h \lambda_j}\right) (1 - d_j) = f_0(\bar{u}) + \alpha_p(1)
\]

Then, by using Chebyshev’s Inequality the general formula is \(1 - 1/SD^2\).
Table 5. Adaptive Nadaraya-Watson Estimator With Epanechnikov As Density Function (Proposed Model)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Bandwidth</th>
<th>Adaptive Bandwidth</th>
<th>Jump Size</th>
<th>MSE</th>
<th>Rate convergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>0.1104331</td>
<td>0.0186666</td>
<td>-0.5111896</td>
<td>0.856058</td>
<td>0.01477945</td>
</tr>
</tbody>
</table>

By using the transformation

\[ f_0(\hat{u}) \int_0^M k(v)dv + o_p(1) \]

\[ = f_0(\hat{u}) \int_0^M k(v)dv + o_p(1) \]

Hence, we know that

\[ \frac{1}{n} \sum_{j=1}^n \frac{1}{h\lambda_j^*} k\left(\frac{u_i - \hat{u}}{h\lambda_j^*}\right)(1 - d_j) \overset{p}{\to} \frac{f_0(\hat{u})}{2} \]

Hence, when we have large number we use Liapunov’s condition;

\[ \sum_i E\left(\frac{1}{n} \sum_{j=1}^n \frac{1}{h\lambda_j^*} k\left(\frac{u_i - \hat{u}}{h\lambda_j^*}\right)d_\xi_i \right)^2 = \frac{1}{(nh\lambda_j^*)^2} E\left(\frac{1}{n} \sum_{j=1}^n \frac{1}{h\lambda_j^*} k\left(\frac{u_i - \hat{u}}{h\lambda_j^*}\right)^2 |^2 | d_\xi_i \right)^2 | u \]

\[ \leq \frac{1}{nh\lambda_j^*} \left[ sup_M E(\left| d_\xi_i \right|^2) u \right] \int_0^M \left| k(v) \right|^2 f_0(\hat{u} + h\lambda_j^* v)dv \]

\[ = o(1) \]

By using central limit theorem (CLT) to find out the asymptotic variance;

\[ \sum_j \text{var} \left( \frac{1}{\sqrt{nh\lambda_j^*}} \sum_{i=1}^n k\left(\frac{u_i - \hat{u}}{h\lambda_j^*}\right)d_\xi_i \right) = E\left[ \frac{1}{h\lambda_j^*} k\left(\frac{u_i - \hat{u}}{h\lambda_j^*}\right)d_\xi_i \delta_j^2(u) \right] \]

\[ = \int_0^M k^2(v) \delta_j^2(\hat{u} + h\lambda_j^* v)dv \]

Similarly, we have

\[ = \delta_j^2(\hat{u}) f_0(\hat{u}) \int_0^M k^2(v)dv + o(1) \]

now by using Liapunov’s CLT and after simplification, we have

\[ \frac{1}{nh\lambda_j^*} \sum_{i=1}^n k\left(\frac{u_i - \hat{u}}{h\lambda_j^*}\right)d_\xi_i \]

\[ = \frac{1}{nh\lambda_j^*} \sum_{i=1}^n k\left(\frac{u_i - \hat{u}}{h\lambda_j^*}\right)(1 - d_j) \overset{d}{\to} N(0, 4\delta_j^2(\hat{u}) + \delta_j^2(\hat{u}) f_0(\hat{u}) \int_0^M k^2(v)dv) \]

Finally, we consider the bias of the estimator;

\[ \text{var} \left( \frac{1}{\sqrt{nh\lambda_j^*}} \sum_{i=1}^n k\left(\frac{u_i - \hat{u}}{h\lambda_j^*}\right)d_\xi_i \left| m(u) - m(\hat{u}) \right) \right] \]

\[ \leq \frac{1}{h\lambda_j^*} Ek\left(\frac{u_i - \hat{u}}{h\lambda_j^*}\right)d_\xi_i \left| m(u) - m(\hat{u}) \right)^2 \]

\[ \leq \left[ sup_M \mathbb{E}(\left| m(\hat{u}) + rh\lambda_j^* \right) - m(\hat{u}) \right)^2 \]
of discontinuity. Whereas, $\hat{m}(y_i) = \frac{1}{n} \sum_{j=1}^{n} k\left(\frac{y_i - y_j}{h}\right)$ is the consistent adaptive nadaraya Watson kernel estimator of $m(y_i)$.

Such that $q_ -$ is $l_p$ times continuously differentiable and bounded away from zero. And

$$v_j = y_i + \alpha d$$

We have,

$$\ell_i = y_i - \hat{m}(y_i) - d_i \alpha$$

$$\hat{m}(y) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h^2} k\left(\frac{y_i - y}{h}\right)(y_i - d_i \alpha)$$

Whereas, we have

$$\delta^2_+(\hat{y}) = \lim_{y \to y^+} \var(y/y = y)$$

and

$$\delta^2_-(\hat{y}) = \lim_{y \to y^-} \var(y/y = y)$$

$\delta^2(y) = \var(y/y_i)$ is uniformly bounded near $y_0$ and the limits $\delta^2_+(y_0), \delta^2_-(y_0)$ exist finite we have;

$$\hat{\delta}(y) = \frac{1}{nh_\alpha^2} \sum_{i=1}^{n} k\left(\frac{y_i - y}{h_\alpha}\right) \ell_i$$

and

$$s(y) = m(y)q_0(y)$$

Suppose $\alpha$ (jump size) is a consistent estimator for $\alpha$. Then by defining their left and right hand side of variance estimation by

$$\delta^2_+(\hat{y}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h^2} k\left(\frac{y_i - y}{h}\right) (y_i - d_i \alpha)^2$$

and

$$\delta^2_-(\hat{y}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h^2} k\left(\frac{y_i - y}{h}\right) ((1 - d_i) \alpha)^2$$

Let we have

$$\delta^2_+(\hat{y}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h^2} k\left(\frac{y_i - y}{h}\right) (y_i - d_i \alpha)^2$$

Hence, we consider the numerator of the equation (5.8)

$$- \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_\alpha^2} k\left(\frac{y_i - y}{h_\alpha}\right)\ell_i$$

Divide equation (5.9) in part $(a + b + c)$ and solve one by one;

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_\alpha^2} k\left(\frac{y_i - y}{h_\alpha}\right) d_i \ell_i^2 + \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_\alpha^2} k\left(\frac{y_i - y}{h_\alpha}\right) d_i \hat{\delta}(y) - \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_\alpha^2} k\left(\frac{y_i - y}{h_\alpha}\right) d_i \delta_i \ell_i$$

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_\alpha^2} k\left(\frac{y_i - y}{h_\alpha}\right)\ell_i (m(y)) - \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_\alpha^2} k\left(\frac{y_i - y}{h_\alpha}\right) d_i \hat{\delta}(y)$$

$$- \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_\alpha^2} k\left(\frac{y_i - y}{h_\alpha}\right) d_i \hat{\delta}(y) - \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_\alpha^2} k\left(\frac{y_i - y}{h_\alpha}\right) d_i \delta_i \ell_i$$

$$\hat{\alpha}(\hat{y}) (m(y)) - \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_\alpha^2} k\left(\frac{y_i - y}{h_\alpha}\right) d_i \delta_i \ell_i$$

$$= \alpha_p \left( \frac{1}{nh_\alpha^2} \right)$$

now by taking part we have

$$\sup_{y \in E_0} |\hat{\delta}_+(d_{q_+})| \leq \sup_{y \in E_0} |\hat{\delta}_+(d_{q_+}) - E(\hat{\delta}_+(d_{q_+}))| + \sup_{y \in E_0} |d_{q_+} - q_0(y)|$$

$$= \alpha_p \left( \frac{1}{nh_\alpha^2} \right)$$

Similarly by taking part, we have

$$\sup_{y \in \bar{E}_0} |\hat{\delta}_-(d_{q_+}) - q_0(y)| \leq \sup_{y \in \bar{E}_0} |\hat{\delta}_-(d_{q_+}) - q_0(y)| + |q_0(y) - q_0(y)|$$

$$= \alpha_p \left( \frac{1}{nh_\alpha^2} \right)$$

By using Lemma A1, we have

$$\sup_{y \in \bar{E}_0} |\hat{\delta}_-(d_{q_+}) - q_0(y)| \leq \sup_{y \in \bar{E}_0} |\hat{\delta}_-(d_{q_+}) - q_0(y)| + |q_0(y) - q_0(y)|$$

$$= \alpha_p \left( \frac{1}{nh_\alpha^2} \right)$$

By using Lemma A3, we have

$$\sup_{y \in \bar{E}_0} |\hat{\delta}_-(d_{q_+}) - s(y)| \leq \sup_{y \in \bar{E}_0} |\hat{\delta}_-(d_{q_+}) - s(y)| + |s(y) - s(y)|$$

$$= \alpha_p \left( \frac{1}{nh_\alpha^2} \right)$$

$$\sup_{y \in \bar{E}_0} |\hat{\delta}_-(d_{q_+}) - q_0(y)| \leq \sup_{y \in \bar{E}_0} \left|\frac{m(y)}{q_0(y)} \right| |\hat{\delta}_-(d_{q_+}) - q_0(y)| + |q_0(y) - q_0(y)|$$

$$= \alpha_p \left( \frac{1}{nh_\alpha^2} \right)$$
Whereas we have
\[
\text{var}\left(\frac{1}{n \lambda h^2} k\left(\frac{\tilde{y} - y}{h \lambda^2}\right)d\tilde{y}ight) \leq \frac{1}{(nh \lambda^2)^2} E\left[k^2\left(\frac{\tilde{y} - y}{h \lambda^2}\right)d\delta^2(y)\right]
\]
By using approximation, we have
\[
= \frac{1}{nh \lambda^2} \int_0^M k^2(v)\delta(\tilde{y} + h \lambda^2 v)f_0(\tilde{y} + h \lambda^2 v)dv
= o\left(\frac{1}{nh \lambda^2}\right)
\]
As already we have prove they asymptotic properties of the estimator, we use the asymptotic variance that is;
\[
\text{var}\left(\frac{1}{n \lambda h^2} k\left(\frac{\tilde{y} - y}{h \lambda^2}\right)d\tilde{y}\right) = o\left(\frac{1}{nh \lambda^2}\right)
\]
Hence, by using the results of chebyshev’s inequality for the uniform convergence, and also that \((\tilde{\alpha} - \alpha) = o_p\left(\frac{1}{nh \lambda^2}\right)\) we have the last term of the equation (5.7) equals to \(o_p\left(\frac{1}{nh \lambda^2}\right)\) Hence, by applying the application of chebyshev’s inequality we prove that; we have the last term of the equation (5.8) equals to \(o_p\left(\frac{1}{nh \lambda^2}\right)\) Hence, by applying the application of chebyshev’s inequality we prove that;
\[
\delta^2(\tilde{y}) \rightarrow \delta^2(\tilde{y})
\]
Similarly, we do for
\[
\delta^2(\tilde{y}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h \lambda^2} k\left(\frac{\tilde{y} - y}{h \lambda^2}\right)(1 - d_i)\tilde{e}_i^2 \left(\frac{\tilde{y}}{q_0(\tilde{y})}\right)
\]  
(5.10)
Hence, we consider the numerator of the equation (5.10)
\[
\frac{1}{n} \sum_{i=1}^{n} \frac{1}{h \lambda^2} k\left(\frac{\tilde{y} - y}{h \lambda^2}\right)(1 - d_i)\tilde{e}_i^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h \lambda^2} k\left(\frac{\tilde{y} - y}{h \lambda^2}\right)(1 - d_i)\tilde{e}_i^2 + \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h \lambda^2} k\left(\frac{\tilde{y} - y}{h \lambda^2}\right)(1 - d_i)[(m(y)) - \frac{1}{\hat{q}_0(\tilde{y})}\left(\frac{\lambda}{nh \lambda^2}\right)^2(\tilde{\alpha} - \alpha)^2]
+ \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h \lambda^2} k\left(\frac{\tilde{y} - y}{h \lambda^2}\right)(1 - d_i)\hat{e}_i[(m(y)) - \frac{1}{\hat{q}_0(\tilde{y})}\left(\frac{\lambda}{nh \lambda^2}\right)^2(\tilde{\alpha} - \alpha)]
- \frac{1}{\hat{q}_0(\tilde{y})}\left(\frac{\lambda}{nh \lambda^2}\right)^2(\tilde{\alpha} - \alpha)(m(y)) - \frac{1}{\hat{q}_0(\tilde{y})}\left(\frac{\lambda}{nh \lambda^2}\right)^2(\tilde{\alpha} - \alpha)
\]  
(5.11)
Divide equation (5.11) in parts and solve one by one;
\[
\frac{1}{n} \sum_{i=1}^{n} \frac{1}{h \lambda^2} k\left(\frac{\tilde{y} - y}{h \lambda^2}\right)(1 - d_i)\tilde{e}_i^2 + \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h \lambda^2} k\left(\frac{\tilde{y} - y}{h \lambda^2}\right)(1 - d_i)[(m(y)) - \frac{1}{\hat{q}_0(\tilde{y})}\left(\frac{\lambda}{nh \lambda^2}\right)^2(\tilde{\alpha} - \alpha)^2]
\]  
By solving using Lemmas $A_1, A_2$ we have
\[
sup_{\lambda \leq y + nh \lambda^2} d\hat{q}_0(y) - E\hat{q}(y) + sup_{\lambda \leq y + nh \lambda^2} d\tilde{q}_0(y) = o_p\left(\frac{1}{nh \lambda^2}\right)
\]
Know by taking part we have
\[
\frac{1}{n} \sum_{i=1}^{n} \frac{1}{h \lambda^2} k\left(\frac{\tilde{y} - y}{h \lambda^2}\right)(1 - d_i)\hat{e}_i[(m(y)) - \frac{1}{\hat{q}_0(\tilde{y})}\left(\frac{\lambda}{nh \lambda^2}\right)^2(\tilde{\alpha} - \alpha)]
\]
Similarly by taking part, we have
\[
\frac{1}{n} \sum_{i=1}^{n} \frac{1}{h \lambda^2} k\left(\frac{\tilde{y} - y}{h \lambda^2}\right)(1 - d_i)\tilde{e}_i[(m(y)) - \frac{1}{\hat{q}_0(\tilde{y})}\left(\frac{\lambda}{nh \lambda^2}\right)^2(\tilde{\alpha} - \alpha)]
\]
using Lemma $A_3$, we have
\[
sup_{\lambda \leq y + nh \lambda^2} |\hat{q}(y) - \tilde{q}(y)| \leq sup_{\lambda \leq y + nh \lambda^2} |\hat{q}(y) - \tilde{q}(y)|
\]  
By using Lemma $A_3$
\[
\sup_{\lambda \leq y + nh \lambda^2} |\hat{q}(y) - \tilde{q}(y)| = o_p\left(\frac{1}{nh \lambda^2}\right)
\]  
Whereas we have
\[
\text{var}\left(\frac{1}{n \lambda h^2} k\left(\frac{\tilde{y} - y}{h \lambda^2}\right)(1 - d_i)\tilde{e}_i\right) \leq \frac{1}{(nh \lambda^2)^2} E\left[k^2\left(\frac{\tilde{y} - y}{h \lambda^2}\right)(1 - d_i)\delta^2(y)\right]
\]
By using approximation, we have
\[
= \frac{1}{nh \lambda^2} \int_0^M k^2(v)\delta(\tilde{y} + h \lambda^2 v)f_0(\tilde{y} + h \lambda^2 v)dv
\]
As already we have prove they asymptotic properties of the estimator, we use the asymptotic variance that is:

\[ \text{var} \left( \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h \lambda_i} k \left( \frac{\bar{y} - y_i}{h \lambda_i} \right) (1 - d_i) \right) = o \left( \frac{1}{nh \lambda_i} \right) \]

Hence, by using the results of Chebyshev’s inequality for the uniform convergence, and also that

\( (\hat{\alpha} - \alpha) = o_p \left( \frac{1}{nh \lambda_i} \right) \)

we have the last term of the equation (5.11) equals to

\[ o_p \left( \frac{1}{nh \lambda_i} \right) \]

Hence, by applying the application of Chebyshev’s inequality we prove that; we have the last term of the equation (5.11) equals to

\[ \delta^2 (\hat{\gamma}) \stackrel{D}{\to} \delta^2 (\gamma) \]

This result requires a more stringent moment for the limiting distribution to estimate. variance estimator and the above consistency theorem don’t required continuity in the derivatives of \( m \) at the discontinuity.