The University of Manchester

MANCHESTER 1824

Economics Discussion Paper Series EDP-1016

An application of local linear regression with asymmetric kernels to regression discontinuity designs

Eduardo Fé

August 2010

Economics
School of Social Sciences
The University of Manchester
Manchester M13 9PL

An application of Local Linear Regression with Asymmetric Kernels to Regression Discontinuity designs

Eduardo Fé

30^{th} June 2010

Abstract

In this article Local Linear Regression with asymmetric kernels is applied to the problem of estimating the average local treatment effect in a Regression Discontinuity design. In these designs identification of the treatment effect is possible in a neighborhood of a cut-off point in the assignment variable. In order to avoid model misspecification nonparametric regression methods have been suggested. However, since identification happens at the cut-off point, nonparametric estimators with good boundary properties are necessary. Nonparametric methods are typically used with symmetric kernels. At the boundary, these kernels are known to have part of their window width devoid of data. This leads to small sample biases and (in the case of the Nadaraya-Watson estimator) to a loss of efficiency. The use of Local Linear Regression corrects the asymptotic loss of efficiency but it does not eliminate the finite sample bias due to the use of a symmetric kernel. It has been pointed out that asymmetric kernels can be used with Local Linear Regression to circumvent both problems. These kernels can be chosen to match the support of the regression function, thus distributing all their weight only in the domain of the function. As a consequence the effective sample size used by the nonparametric estimator is increased and the boundary bias problem is avoided even with finite samples. This article establishes the asymptotic normality of the estimator of the local average treatment effect when local linear regression with asymmetric kernels is used in applications, and the small sample properties of the ensuing statistics are studied in a Monte Carlo Experiment.

Keywords: Regression Discontinuity, Asymmetric Kernels, Local Linear Regression.

JEL Classification: *C13, C14, C21.*

1 Introduction

This article explores the use of local linear regression with asymmetric kernels for the estimation of treatment effects in a Regression Discontinuity (RD) design. Regression Discontinuity (Thistlethwaite and Campbell (1960)) is a quasi-experimental program evaluation technique applicable when the probability distribution describing allocation into treatment is suspected to have a discontinuity at a known cut-off level of certain continuous *running* or *assignment* variable. When a program of these characteristics is perfectly implemented, researchers can observe to a great extent the selection process. Then, under certain weak uniformity conditions, a comparison of the conditional means of outcomes of treatment between recipients and non-recipients is sufficient to identify the average effect of the intervention within a small neighborhood of the cut-off point. What makes the method specially attractive is that local identification is possible even when treatment effects vary across individuals or there is selection into treatment due to, for instance, anticipated gains. Thus RD has become very popular among practitioners, and one finds numerous applications including van der Klaauw (1996), Angrist and Lavy (1999)), Ludwig and Miller (2007) or Almond et al. (2010) to mention but a representative few.

In RD designs, the cut-off point splits the region of estimation in two (or more) intervals with known bounds and it is at these bounds where the treatment effect is identifiable. Therefore, valid inference of the causal effect requires, firstly, estimators with good boundary properties and, secondly, the correct specification of the conditional mean of the outcome and assignment variables in a neighborhood of the cut-off points. To reduce the likelihood of model misspecification, Hahn et al. (1999) and Porter (2003) proposed the use of nonparametric regression methods. Among these estimators, Local Linear Regression (LLR) with symmetric kernels has been widely used. This estimator is consistent, asymptotically normal distributed and it circumvents the so called boundary bias problem: its bias vanishes at a similar rate at the boundary and interior of the regions of estimation. In contrast, the bias of the popular Nadaraya-Watson estimator disappears at a slower rate at the boundary and so larger sample sizes are required in this region to obtain a given level of accuracy (Fan (1993), Wand and Jones (1994) or Ruppert and Wand (1994)).

The boundary bias thus described is an asymptotic problem. In finite samples, at the boundary, symmetric kernels typically used in LLR still have part of the kernel window devoid of data, leading to the introduction of a small sample bias. Using kernels with a bounded support (such as uniform, triangular or Epanechnikov kernels) could, in principle, ameliorate the problem. However Seifert and Gasser (1996) have shown that the LLR estimator with a compact kernel has unbounded unconditional variance in finite samples, unlike LLR based on Gaussian kernels, which suggest this kernels can result in fairly imprecise estimates. In a series of papers, Brown and Chen (1999) and Chen (2000, 2001) study kernel selection when estimating conditional moments with bounded support. They show that using gamma and beta densities as kernels in LLR has a number of advantages. These kernels match the support of the regression function. This has two implication. Firstly, it implies that no weight is allocated beyond the boundaries of the support (thus eliminating small sample biases). Secondly the whole sample is effectively used by the estimator, thus increasing the stability statistic. Unlike LLR with kernels of bounded support, the LLR estimator with Beta and Gamma kernels has a bounded finite sample variance. Furthermore, at the boundary, the rate of convergence to the true regression equals to that exhibited by standard LLR. These gamma and beta kernels are reparameterised so that their shape and scale depends on a smoothing parameter as well as the point at which estimation takes place. As a result these kernels adapt the amount of smoothing to the location of estimation. Chen (2001) shows through simulations that, at the boundary, these new methods have biases at least comparable to those obtained with a Gaussian kernels while the variance is substantially smaller across the domain of the curve and so the new estimators are preferred in accordance to a mean square error criterion. Given that in RD designs identification of treatment effect is possible at the boundary, the theoretical results in Chen (2001) suggest that gamma and beta kernels can contribute to more accurate estimation of the local treatment effect, and therefore this article explores the magnitude of that contribution.

The structure of the paper is as follows. In section 2 we obtain the asymptotic distribution of the modified estimator of the treatment effect. This section extends the results in Chen (2001), by establishing the asymptotic normality of the LLR with Gamma and Beta kernels. Given the type of problem under study, our results specialize to the case of estimation at the boundary. However our results can be modified in a straightforward fashion to obtain the asymptotic normality of the estimator anywhere else in the domain or the regression function[1] The asymptotic Mean Square Error of the estimator is used to obtain the optimal value of the bandwidth parameter, and we show that its rate of convergence to 0 is identical to that derived from LLR with symmetric kernels. Section 3 collects the results of a Monte Carlo experiment devised to compare the performance of the new estimators with previous statistics considered

[1]However we anticipate that the rate of convergence of the estimator in the interior is likely to differ from that obtained at the boundary. See Chen (2001).

in the literature. The results suggest that the estimator of treatment effects based on LLR with gamma and beta kernels are moderately robust to the value of the smoothing parameter, and they tend to dominate previous estimators in terms of Mean Square Error. Section 4 concludes.

2 Nonparametric Estimation of a Regression Discontinuity with Asymmetric Kernels.

Consider a researcher who is interested in measuring the effect that an intervention has on an outcome variable $y_i \in \mathbb{R}$. The intervention has been designed in such a way that the probability of allocation into treatment depends on the levels of the *running* variable z. The researcher has a sample of $i = 1, \ldots, n$ independent observations $(y_i, z_i, x_i)_{i=1}^n$, where x_i is a binary indicator of allocation into treatment such that $x_i = 1$ if the individual has received treatment, while $x_i = 0$ otherwise. What characterises a RD design is that the indicator x_i is a random variable whose probability distribution, $P(x_i = 1|z_i) = E(x_i|z_i)$, is discontinuous at z_o, a known threshold or cut-off point in the range of z. A particular case of this setting is the so called Sharp RD design, where the probability distribution of x_i is degenerated at z_o so that $P(x_i = 1|z_i)$ jumps from 0 to 1 at the threshold. Although we do not make explicit reference to this case, our results can be extrapolated in a natural manner.

For each individual in the sample there are two potential outcomes: y_{i1} if the individual receives treatment and y_{i0} otherwise. Both outcomes are not observed simultaneously. Therefore, the model for the observed outcome is

$$y_i = y_{i0} + x_i(y_{i1} - y_{i0}) = \alpha_i + x_i\beta_i$$

where $\beta_i = (y_{i1} - y_{i0})$ captures the effect of the intervention and it is the object of interest.

Hahn et al. (1999, 2001) formalize the conditions under which the treatment effect is (locally) identifiable at the threshold z_o in a RD design. These are summarized on the following assumption.

Assumption 2.1 *(Hahn et al. (1999)). Let $m^+ = m^+(z_o) = \lim_{z\to z_o^+} E(y|z)$, $m^- = m^-(z_o) = \lim_{z\to z_o^-} E(y|z)$ and suppose that (i) $p^+ = p^+(z_o) = \lim_{z\to z_o^+} E(x|z)$, $p^- = p^+(z_o) = \lim_{z\to z_o^-} E(x|z)$ exist and $p^+ \neq p^-$, and assume that (ii) $E(\alpha_i|z = z_o)$ is* *continuous in z at z_o and $\beta_i = \beta$ for all i. Then,*

$$\beta = \frac{m^+ - m^-}{p^+ - p^-}. \tag{2.1}$$

If β_i varies across individuals and it also holds that (iii) $E(\beta_i|z = z_o)$ is continuous at z_o and (iv) x is independent of β_i given z near z_o, then

$$E(\beta_i|z = z_o) = \frac{m^+ - m^-}{p^+ - p^-}. \tag{2.2}$$

Finally, if conditions (i) and (ii) hold, and if (v) $(\beta_i, x(z))$ are jointly independent of z near z_o and (vi) there exists an $\varepsilon > 0$, such that $x(z_o + e) \geq x(z_o - e)$ for all $0 < e < \varepsilon$, then,

$$\lim_{e \to 0^+} E(\beta_i|x(z_o + e) - x(z_o - e) = 1] = \frac{m^+ - m^-}{p^+ - p^-}. \tag{2.3}$$

Under Assumption 2.1., the quantity $m^+ - m^-/p^+ - p^-$ locally identifies β_i in a variety of situations, including homogeneous and heterogeneous treatment effects. More interestingly, conditions i, ii and v ensure that identification of treatment effect is possible even when individuals self-select into treatment (due to, for example, anticipated gains from treatment).

Estimation of the treatment effect in this setting requires estimates of the limits m^+, m^-, p^+ and p^-. To avoid too strong parametric assumptions and reduce the chances of model misspecification, Hahn et al. (1999) suggested estimating these quantities via Local Linear Regression (LLR). Thus, the estimator of $m^+(z_o)$ is the value of a solving the weighted least squares problem

$$\arg\min_{a,b} n^{-1} \sum_{i=1}^{n} (y_i - a - b(z_i - z_o) K_h \left(\frac{z_i - z_o}{h} \right) \mathbb{I}(z_i > z_o) \tag{2.4}$$

where $K_h = h^{-1}K(.)$ is a kernel function distributing weights across the sample points, $h = h(n)$ is a bandwidth parameter regulating the width of the kernel and such that $h \to 0$ as $n \to \infty$. $\mathbb{I}(.)$ is an indicator function taking the value 1 when the condition inside the brackets is true. Therefore, unlike standard LLR, only sample points above (below) the cut-off point z_o are used in the estimation of m^+ and p^+ (m^- and p^-). The estimator of β_i is then,

$$\hat{\beta}_i = \frac{\hat{m}^+ - \hat{m}^-}{\hat{p}^+ - \hat{p}^-} \tag{2.5}$$

What makes LLR specially attractive in a RD framework is its boundary bias properties. The order of the bias of the LLR at points within one bandwidth of the boundary is equal to $O(h^2)$ while the order of this bias is $O(h)$ when using the better known Nadaraya-Watson estimator. As a result, the NW estimator requires larger amounts of data in order to attain a given level of precision. In a RD discontinuity identification of treatment effect is possible at the cut-off point, which generates two regions with a least one well defined upper/lower bound. Thus, estimators with good boundary properties seem particularly important.

At the core of the boundary bias problem is the fact that part of the window of the kernel weight is allocated beyond the boundary of the region of estimation. With finite samples, when estimating a regression function at a boundary point, typical symmetric kernels used in nonparametric regression (such as the Gaussian kernel) will allocate a portion of the window width in a region outside the support of the regression function. The problem can be mitigated using a kernel with compact support, such as Epanechnikov or Triangular kernels. However, as shown by Seifert and Gasser (1996) the LLR estimator with a compact kernel has unbounded unconditional variance in finite samples, unlike LLR with a Gaussian kernel. This suggest these kernels can lead to imprecise estimates.

Chen (2001) has shown that it is possible to find kernels that, matching the support of the regression function, distribute all their weight only on this region and, furthermore, when used in a LLR setting, these kernels produce estimators that retain the bounded finite sample variance of the LLR with a Gaussian kernel. The two kernels considered by Chen (2001) are

$$K^B_{z_o,b}(z_i) = \frac{z_i^{z_o/b}(1-z_i)^{(1-z_o)/b}}{B\left(\frac{z_o}{b}+1, \frac{1-z_o}{b}+1\right)} \tag{2.6}$$

$$K^G_{z_o,b}(z_i) = \frac{z_i^{z_o/b}e^{-z_i/b}}{b^{\frac{z_o}{b}+1}\Gamma\left(\frac{z_o}{b}+1\right)} \tag{2.7}$$

where $B(.,.)$ is the beta function and $b = b(n) > 0$ is a smoothing parameter satisfying $b \to 0$ as $n \to \infty$. Here $K^B(.)$ is a beta density with parameters $r = z_o/b + 1$ and $s = 1 - z_o/b + 1$, while $K^G(.)$ is a gamma density with parameters $r = b$ and $s = z_o/b + 1$. These parameterizations locate the mode of the density at z_o. Furthermore, the shape of these densities depends on the values of their parameters, and these parameters are functions of z_o and the smoothing parameter b, so that for fixed b, these kernels provide a kind of locally adaptive smoothing approach.

The domain of the gamma kernel is $[0, \infty)$ while the domain of the beta kernel is

$[0,1]$. In practice the regions generated by the cut-off point in a RD design will rarely equal to these intervals so these kernels would not match the support of the regression function at the outset. However, since the cut-off values are known to the researcher, it is always possible to map the different regions onto the intervals $[0,\infty)$ or $[0,1]$ (leaving y unaltered). Thus, for example, if the region of interest is $[a,b]$, (a, b finite), then $z-a/b-a$ is a mapping onto $[0,1]$. By matching the support of the regression function, LLR based on these kernels will use all the sample points in that region but no weight will be allocated outside the boundaries of the interval, thus increasing the effective sample size without incurring in a boundary bias problem in small samples. In practice, it seems clear that the Gamma kernel will be required when the (transformed) domain of the regression has only a lower bound, while the beta kernel would be more appropriate if upper and lower bounds are known to the domain of the conditional moment. However both procedures could be used in any application. More specifically, in the case of the beta kernel researchers can let the bounds coincide with the observation with a largest and smallest abscissa.

We present next the asymptotic properties of the ensuing estimator. Prior to that, we introduce some notation and a few assumptions. Following Chen (2001), we define the point z in the support of z_i to be a boundary point if:

1. $z/b \to \kappa$, when using gamma kernel
2. $z/b \to \kappa$ or $(1-z)/b \to \kappa$ when using a beta kernel.

for some $\kappa \geq 0$. The following properties of the asymmetric kernels above will be necessary to obtain the asymptotic distribution of the estimator of β_i.

Property 2.1 *(Chen (2001)). Let ξ be a random variable with density function (2.7) and let z be a point in the range of z_i. Define $p_l(z) = E(\xi - z)^l$. Then*

1. *If ξ has the density (2.7), $p_2(z) = bz + 2b^2$ and $p_l(z) = O(b^2)$ for $l \geq 3$. In particular, if z is a boundary point, $p_2(z) = b^2(2+\kappa)$. Furthermore, $K^G_{z,b}(z_i)^2 = A_b(z)K_{z\prime,b\prime}(z_i)$. If z_o is a boundary point, then $A_b(z) = b^{-1}\Gamma(2\kappa+1)/2^{2\kappa+1}\Gamma^2(\kappa+1) + o(b^{-1})$.*
2. *If ξ has the density (2.6), $p_2(z) = bx(1-x) + O(b^2)$ and $p_l(z) = O(b^2)$ for $l \geq 3$. In particular, if z is a boundary point, $p_2(z) = b^2(2+\kappa)$. Furthermore, at the boundary, $K^B_{z,b}(z_i)^2 = b^{-1}\Gamma(2\kappa+1)/2^{2\kappa+1}\Gamma^2(\kappa+1) + o(b^{-1})$*

The key aspect of the above properties is that the moments $p_2(.)$ and $K()^2$ coincide for both kernels at the boundary. These two moments ultimately characterize the asymptotic distribution of the LLR with asymmetric kernels (as will be seen in the Appendix), and therefore, the coincidence of these quantities at the boundary will lead to estimators with identical asymptotic distribution. The claim is proved below.

In order to introduce the main result of the paper, we require the satisfaction of the following assumptions.

Assumption 2.2 *Let $\mathcal{V} = [-M, M]$, $0 < M < \infty$, be a neighborhood of z_o. The following assumptions are true*

1. *The running variable, z, has marginal density $f(z)$, bounded, positive and twice continuously differentiable in $\mathcal{V}$.*

2. *The conditional expectations $m(z)$, $p(z)$, $\sigma^2(z) = var(y|z)$ and $\eta(z) = cov(y, x|z)$ are twice continuously differentiable at $\mathcal{V} \backslash z_o$. Their left limits and first and second derivatives exist and are uniformly bounded on $[z_o - M, z_o)$, and similarly, the right limits of $m(z)$, $p(z)$ and its first and second derivatives exist and are uniformly bounded on $(z_o, z_o + M]$.*

3. *$m(z_i)$, $\sigma^2(z_i)$ and $\eta(z_i)$ satisfy the expansion,*

$$\zeta(z_i) = g(z_i) - g^+(z) - g'^+(z) - \frac{1}{2}m''^+(z)(z_i - z)^2$$

where $\sup_{z<z_i<z+Mb} |\zeta(z_i)| = o(b^2)$, for $0 < M < \infty$ and where $g(.)$ stands for m, σ^2 or η.

4. *$E\left[(y - m(z))^{2+\delta}|z\right]$ and $E\left[(x - p(z))^{2+\delta}|z\right]$ are uniformly bounded on $\mathcal{V}$, for $\delta >$ 0.*

5. *$b \to 0$, $nb \to \infty$ as $n \to \infty$.*

The conditions above are standard in the literature (Hahn et al. (1999); Porter (2003)). Assumption 2.3 ensures the satisfaction of a suitable Lyapounov condition, and weaker than the condition found in Hahn et al. (1999). Unlike in other sources continuous differentiability of second order is imposed on η, σ^2 and m (through assumption 2.2.3 above) , in order to approximate these quantities on a neighborhood of the cut-off point. Hahn et al. (1999) do not require such restriction and their results use only the dominated convergence theorem.

Lemma 2.1 *Let $\hat{m}^+(z_o)$ be a LLR estimator of $m^+(z_o)$ with the gamma or beta kernels defined in equations (2.7) and (2.6), let the point z_o be the boundary point of the support of the regression function and assume that $b^2\sqrt{nb} \to \varrho > 0$. Then, under assumptions 2.1, 2.2. and property 2.1,*

$$\sqrt{nb}\left\{\hat{m}^+(z_o) - m^+(z_o)\right\} - \varrho m''^+(z_o) \sim N\left(0, \frac{\sigma^{2+}(z_o)}{2f(z_o)}\right). \tag{2.8}$$

The proof is given in the Appendix. The expression of the asymptotic bias and variance is particularly simple, and it does not depend on nuisance parameters. This is possible because we use specific kernel, and our results specialize to particular values of z (namely 0 and 1) that satisfy the boundary conditions. In accordance to the above lemma, one can easily derive the expression of the Mean Square Error of this estimator and from this, the optimal value of the smoothing parameter b, which is given by,

$$b^* = \left\{\frac{\sigma^{+2}(z_o)}{2f(z_o)m''^{+2}(z_o)}\right\}^{1/5} n^{-1/5} = Cn^{-1/5} \tag{2.9}$$

Thus the optimal smoothing parameter converges to 0 at the same rate than h in the standard nonparametric regression with symmetric kernels. However, the magnitude of the constant C will in general we distinct to that accompanying the theoretically optimal value of h, so that estimations based on equal h and b are not directly comparable. From the lemma we see that, as in the case of LLR with symmetric kernels, the bias of the estimator depends on the curvature of the (limit) of the regression function in a neighborhood of the boundary point. If $m(.)$ is linear in this neighborhood, the estimator will be unbiased and optimal bandwidth will tend to ∞; as the complexity of the design increases and the bias increases, the optimal value of the bandwidth gets progressively smaller. Finally, since b^* depends on unknown moments, the usual problems associated to bandwidth selection apply here as well (see, for example, Hart (1997)). Ludwig and Miller (2007) and Imbens and Lemieux (2008) discuss several data-driven methods devised with a RD design in mind, all of which are applicable here.

With assumptions 1 and 2 in place, it is now possible to establish the main result of the paper.

Theorem 2.1 *Let assumptions 2.1 and 2.2 hold, $b^2\sqrt{nb} \to \varrho$, $0 \leq \varrho < \infty$, and assume that $K_{x,b}(z_i)$ is either the gamma or beta density in equation (2.7). Then, at z_o*

$$\sqrt{nb}\left(\frac{\hat{m}^+ - \hat{m}^-}{\hat{p}^+ - \hat{p}^-} - \frac{m^+ - m^-}{p^+ - p^-}\right) \to \mathcal{N}\left(\lambda, \tau\right) \tag{2.10}$$

where,

$$\lambda = \varrho\left(\frac{1}{p^+ - p^-}(m''_+ - m''_-) - \frac{m^+ - m^-}{(p^+ - p^-)^2}(p''_+ - p''_-)\right) \tag{2.11}$$

and

$$\begin{aligned} \tau &= \frac{1}{2f}\left(\frac{1}{(p^+ - p^-)^2}(\sigma^2_+ - \sigma^2_-)\right. \\ &- 2\frac{m^+ - m^-}{(p^+ - p^-)^3}(\eta^+ - \eta^-) \\ &+ \left.\frac{(m^+ - m^-)^2}{(p^+ - p^-)^4}(p^+(1-p^+) - p^-(1-p^-))\right) \end{aligned} \tag{2.12}$$

where f and all the limits of the conditional expectations are evaluated at the cut-off point $z_o = 0$.

3 Monte Carlo.

This section collects the result of a Monte Carlo experiment devised to compare the performance of the estimator of β_i under different choices kernel. A fuzzy RD setting was designed where the probability of assignment to treatment was given by:

$$\mathbb{P}(x_i|z_i) = \begin{cases} (1 + \exp(-(z - \mu_\ell)/s_\ell))^{-1} & \text{for } z_i \leq z_o \\ (1 + \exp(-(z - \mu_h)/s_h))^{-1} & \text{for } z_i > z_o \end{cases} \tag{3.1}$$

where $(\mu_\ell, s_\ell) = (7, 1)$, $(\mu_h, s_h) = (4, 1)$ and z_i is the running variable and the cut-off value is $z_o = 5$. Three specifications were used to generate the outcome variable:

$$\text{DPG 1:} \quad y_i = \beta_0 + \beta_1 z_i + \beta_{2i} x_i + \varepsilon_i \tag{3.2}$$

$$\text{DPG 2:} \quad y_i = \frac{1}{1 + \exp(\beta_0 + \beta_1 z_i + \beta_{2i} x_i + \varepsilon_i))} \tag{3.3}$$

$$\text{DGP 3:} \quad y_i = \beta_0 (z_i^* - 5)^3 (1 - x_i) + \left(\beta_1 (z_i^* - 5)^3 + \beta_2\right) x_i + \varepsilon_i \tag{3.4}$$

where x_i is the indicator of treatment and z_i, z_i^* were the running variables defined as:

$$z_i \sim N(5, 1) \tag{3.5}$$

$$z_i^* \sim \begin{cases} Beta(1, 5) & \text{whenever } z_i^* \leq 5 \\ Beta(5, 1) & \text{whenever } z_i^* > 5 \end{cases} \tag{3.6}$$

The treatment effects in each model were estimated using the statistic described in the previous section, using four different kernels: Gaussian, Epanechnikov, Gamma and Beta.

Each regression function exhibits different levels of curvature in a neighborhood of the cut-off point. In DGP 1, $\beta_0 = \beta_1 = 1$ and $\varepsilon \sim N(0,1)$. The treatment effect is captured by $\beta_{2i} = 1 + \nu_2$ where $\nu_j \sim N(0, 0.25)$. Thus, DGP 1 exhibits heterogeneous treatment effects, with a mean value of 1. This model is linear, and therefore the regression function is estimated free of bias by the LLR (since $m''(z) = 0$ for all z). As a result, when estimating the limits m^+ and m^-, the value of the optimal bandwidth converges to ∞ as $m''(z) \to 0$.

In DGP 2, $\beta_0 = -6$, $\beta_1 = 1$ and $\beta_{2i} = 1 + \nu_3$, $\nu_j \sim N(0, 0.25)$, and $\varepsilon \sim N(0,1)$. Treatment effect is once again heterogeneous across individuals. The regression function combines two logistic functions, with a discontinuity at $z = 5$. The value of the discontinuity at $z = 5$ is 0.23106 (which corresponds with the average treatment effect). The value of the slope of the regression function changes only moderately in a neighborhood of the threshold, and so departure from a linear specification is not totally obvious in this neighborhood. Thus, when estimating m^+ and m^-, large bandwidths will tend to yield more accurate estimates.

Finally, in DGP 3, $\beta_0 = \beta_1 = \beta_2 = -2$ and and $\varepsilon \sim N(0,1)$. The treatment effect is capture by β_2. In this case we have homogeneous treatment effects across individuals. What characterises this model is the change of the sign of the second derivatives of the regression function at the threshold, so that the function switches from a convex to a concave mapping. Thus this model is relatively more complex to estimate than the previous two specification and the fast change of the second derivatives about z_o will require smaller bandwidths than in the previous two cases.

It has been commented that the estimator in (2.5) can locally identify the treatment effect even in the presence of selection into treatment. Therefore, to simulate selection into treatment we forced $x_i = 1$ whenever $z_i \leq 5$ and $\varepsilon_i < m$ where $m = 1.645$, the 90 quantile of the standard normal distribution.

Rather than using a single value of the bandwidth to estimate each model, arrays of values were considered instead. Estimation of the treatment effect in models 1 and 2 was based on $h, b = c = 0.3 + j * 0.1$, with $j = 3, \ldots, 40$, while for model 3 $h, b = c = 0.1 + j * 0.02$, with $j = 1, \ldots, 90$.

Samples of 500, 1000 and 2000 observations were drawn 10.000 times for each design and the treatment effect was estimated using each of the four different kernels. The mean squared bias, $MSE = R^{-1} \sum_{j=0}^{10.000} (\hat{\tau} - \tau)^2$ was calculated as a measure of overall performance. Figures 1 to 9 summarize the results of the simulation.

The curves depicted in figures 1 to 3 describe the variation of $\log(MSE)$ as a function of the bandwidth parameter when data are generated from the linear specification in DGP 1. The performance of all four estimators improves (at decreasing rates) as the magnitude of the smoothing parameter increases. It thus seems that the reduction in the bias in the numerator of the statistics (as the magnitude of the smoothing parameter increases) dominates the behaviour of the estimator. As expected, increasing the sample size leads to a reduction of the average value of $\log(MSE)$. All four estimators attain almost identical $\log(MSE)$ for very large values of the bandwidth parameters. However, for moderate values of the smoothing parameters, it is possible to establish a clear ranking amount all four methods. The LLR with beta kernel clearly dominates the other estimators. The Gamma and Gaussian kernels lead to comparable performance in terms of $\log(MSE)$, but the LLR is comparatively much worse if the Epanechnikov kernel is used.

When the logistic mode (DGP 2) is considered, the estimators perform worse when the magnitude of the bandwidth is small. As the amount of smoothing increases, the precision of the estimators increases, but only up to a point, beyond which the mean square error begins to increase. This is clearer for $N = 2000$. In this case, all four methods attain a minimum error at $\log(MSE) \approx 3.2$. However, this minimum is attained at different values of the smoothing parameter. The optimal value of b is about 0.7 when using the beta kernel and approximately 2 if using the gamma kernel. In contrast with this, the optimal value of h was attained at 1.5 and 2.7 if using the Gaussian and Epanechnikov kernels, respectively. Once again, when a small bandwidth parameter was used in all four methods, then the LLR with Beta kernel would be the most reliable estimator, followed by the Gamma, Gaussian and Epanechnikov kernels. Large bandwidths, on the other hand, will make the performance of the four methods comparable (although, for large N, the Gaussian kernel seem to be the least accurate among the estimators).

When data were generated with the cubic model (DGP 3), the value of the smoothing parameter leading to the optimal performance was significantly smaller than in the previous two cases. Again, all four methods attained a comparable minimum $\log(MSE)$: This minimum value was located at about -0.3 for $N = 500$, -0.6 for $N = 1000$ and -0.85 for $N = 2000$. However all estimators required different values of the smoothing parameter to obtain their best performance. Thus, for $N = 2000$, the optimal value of b was 0.12 for the LLR-Gamma estimator and 0.25 for the LLR-Beta estimator. Similarly, the optimal h was 0.15 and 0.37 for the LLR-Gaussian and LLR-Epanechnikov estimators respectively. Outside these values, is is difficult to establish a ranking among the estimators. If researchers would be oversmoothing the data, then the LLR-Beta estimator dominates all of the other methods, followed by the LLR-Epanechnikov, LLR-Gamma and LLR-Gaussian estimators. For small to moderate amounts of smoothing, then the LLR-Beta estimator tends to dominate the others, while the Gaussian and Gamma kernels generate estimates of comparable quality. The LLR-Epanechnikov estimator, on the other hand, can behave quite poorly. In general, the Gamma and Beta kernels lead to more stable estimation, and their $\log(MSE)$ curves change less dramatically with b than those of the Gaussian and Epanechnikov kernels.

The results suggest that all four methods described here can provide comparable accuracy. However, for this to happen, researchers must be able to estimate the optimal value of the smoothing parameter with great accuracy. Otherwise, all four estimators behave quite differently. In practice, this is not a minor point, since the problem of optimal bandwidth selection is largely unresolved. To illustrate the problem, we repeated the experiment using model 3 and a Gaussian kernel with $N = 2000$, but in this occasion, the amount of smoothing was selected using simple Cross-Validation, which is one of the better know data-driven selectors. The mean bandwidth was 1.41, with a median value of 0.1345. The distribution of bandwidths was wide and heavily skewed. The median value of 0.13 was close to the optimal value of 0.15 mentioned above, however the Gaussian kernel is only marginally preferred to the Gamma kernels is a neighborhood about 0.13.

Given the difficulty of choosing the amount of smoothing in practice, the question is which of the above methods is most reliable. Our experiment seems to favour the LLR-Beta. This estimator dominates the others in most settings, and it seems to be less sensitive to the amount of smoothing in the sample. The LLR-Gamma estimator exhibits, for small bandwidths, a performance comparable to that achieved by the LLR-Gaussian estimator. However, if oversmoothing, Gamma kernel is, again, less sensitive to the choice of bandwidth parameter. The LLR-Epanechnikov, on the other hand, can be quite inaccurate, and its performance varies widely with the magnitude of h, making it the least preferred of the methods.

4 Conclusion

This article has discussed the use of Local Linear Regression (LLR) with asymmetric kernels in order to estimate the effect of a policy, intervention or treatment in a Regression Discontinuity design. Following Chen (2001), gamma and beta kernels have been proposed to replace symmetric kernels in the LLR. These kernels circumvent the problem of boundary bias even in finite samples (by distributing their weight only within the support of the regression function) and they increase the effective amount of data used by the estimator (by matching the support of the regression function), thus making the LLR more stable.

The article provides the expressions of the approximate large sample bias and variance of the estimator and, in doing so, results in Chen (2001) have been extended by establishing the asymptotic normality (at the boundary) of the LLR estimator with a gamma and beta kernels. Although the results are circumscribed to the boundary of the distribution of the running variable, a similar approach could be used to prove the asymptotic normality of these estimator at any point of the domain of the regression function. The asymptotic distribution of the estimators allowed us to obtain the theoretically optimal value of the smoothing parameter. It turns out that the optimal rate of convergence to 0 is $O(n^{-1/5})$, which is the rate commonly exhibited when the LLR estimator is combined with symmetric kernels.

A Monte Carlo experiment was used to evaluate the performance of the new method and study the impact of the amount of smoothing. Two are the main findings of our simulation. Firstly, the new estimators can dominate existing methods based on Gaussian or Epanechnikov kernels in a variety of scenarios and, secondly, we found that the new methods are less sensitive to the magnitude of the bandwidth parameter than LLR based on Gaussian or Epanechnikov kernels, making them attractive methods for empirical applications.

Although our results focus on beta and gamma kernels, other choices of kernel are available. Scaillet (2004) has used Inverse Gaussian and Reciprocal Inverse Gaussian kernels for density estimation. Other choices might be also feasible. However an study of alternative asymmetric kernels and their peformance is beyond the scope of the present article, and it is left for future consideration.

References

Almond, D., Doyle, J., Kowalski, A., and Williams, H. (2010). Estimating marginal returns to medical care: Evidence from at-risk newborns. *Quarterly Journal of Economics*, 125:591–634.

Angrist, J. and Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*, 114:533–575.

Brown, B. and Chen, S. X. (1999). Beta-Bernstein smoothing for regression curves with compact supports. *Scandinavian Journal of Statistics.*, 26:47–59.

Chen, S. X. (2000). Beta kernel smoothers for regression curves. *Statistica Sinica*, 10:73–91.

Chen, S. X. (2001). Local linear smoothers using asymmetric kernels. *Annals of the Institute of Statistical Mathematics.*, 54:312–323.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Annals of Statistics*, 21:196–216.

Hahn, J., Todd, P., and van der Klaauw, W. (1999). Evaluating the effect of an antidiscrimination law using a regression-discontinuity design. *NBER Working Paper*, 7131.

Hahn, J., Todd, P., and van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression discontinuity design. *Econometrica*, 69:201–209.

Hart, J. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests.* Springer-Verlag New York.

Imbens, G. and Lemieux, T. (2008). Regression Discontinuity Designs: A guide to practice. *Journal of Econometrics*, 142:615–635.

Ludwig, J. and Miller, D. (2007). Does a head start improve childrens' life chances? evidence from a Regression Discontinuity Design. *Quarterly Journal of Economics*, 122:159–208.

Porter, J. (2003). Estimation in the Regression Discontinuity model. Technical report, Harvard University.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Annals of Statistics*, 22:1346–1370.

Scaillet, O. (2004). Density estimation using inverse and Reciprocal Inverse Gaussian kernels. *Journal of Nonparametric Statistics*, 16:217–226.

Seifert, B. and Gasser, T. (1996). Finite sample variance of local polynomials: Analysis and solutions. *Journal of the American Statistical Association*, 91:267–275.

Thistlethwaite, D. and Campbell, D. (1960). Regression-Discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51:309–317.

van der Klaauw, W. (1996). A Regression-Discontinuity evaluation of the effect of financial aid offers on college enrollment. Technical report, New York University, Economics Department.

Wand, M. and Jones, M. (1994). *Kernel smoothing.* Chapman and Hall/CRC.

A Proofs

This appendix contains the proofs of the main results. We first obtain the asymptotic distribution of $\hat{m}^+$ at the boundary. The asymptotic distribution of the estimators of m^-, p^+ and p^- is obtained identically, and so the proof concerning these cases is omitted.

Define

$$y_i^* = y_i - m^+(z) + m'^+(z)(z_i - z) = y_i - Z_i'\theta,$$

where $Z_i' = (1, (z_i - z))$, $\theta = (m^+(z), m'^+(z))'$ and $y_i = m(z_i) + u_i$. Here the point $z = z_o$ is the boundary on which estimation of $m^+(z)$ is taking place and equals 0 or 1, depending on the kernel employed. Define Z as the $n \times 2$ matrix with rows Z_i', K is the diagonal matrix with elements $K_{z,b}(z_i)\mathbb{I}(z_i > z)$ and let M and u be $n \times 1$ matrices with elements $m(z_i)$ and u_i respectively. The design points, z_i have been normalized to fall within $[0, \infty)$ (when using Gamma kernels) or $[0, 1]$ (if using a beta kernel); this can be done without loss of generality. Finally let $e_1' = (1, 0)$.

The estimator of m^+ that solves equation (2.4) can be written as:

$$\begin{aligned} \hat{m}^+(z) - m^+(z) &= e_1'(Z'KZ)^{-1}Z'K(M + u - Z\theta) \\ &= e_1'(Z'KZ)^{-1}Z'K(M - Z\theta) \\ &\quad + e_1'(Z'KZ)^{-1}Z'Ku \qquad \text{(A-1)} \end{aligned}$$

Lemma A.1 *Under assumptions 2.1, 2.2 and property 2.1,*

$$e_1'(Z'KZ)^{-1}Z'K(M - Z\theta) = b^2 m''^+(z) + o_p(1) \qquad \text{(A-2)}$$

at z, regardless of whether a beta or gamma kernel is used in the estimation.

Proof. Consider first the LLR with the gamma kernel. The matrix $Z'KZ/n$ has typical element

$$S_l(z) = n^{-1} \sum_{i=1}^{n} K_{z,b}(z_i)\mathbb{I}(z_i > z)(z_i - z)^l$$

for $l = 0, 1, 2$. Chen (2001) shows that $S_l(z) = E(S_l(z)) + o_p(1)$, where

$$\begin{aligned} E\left(S_l(z)\right) &= \int_0^\infty K_{z,b}(z_i)(z_i - z)^l f(z_i) dz_i \\ &= E(f(\xi_i)(\xi_i - z)^l) \\ &= E\left\{\left(f(z) + f'(z)(z_i - z) + \frac{f''(z)}{2}(\xi_i - z)\right)(\xi_i - z)^l\right\} + O(b^2) \\ &= f(z)p_l(z) + f'(z)p_{l+1}(z) + \frac{f''(z)}{2}p_{l+2}(z) + O(b^2) \qquad \text{(A-3)} \end{aligned}$$

where ξ_i is a random variable with density $K_{z,b}(.)$. The third equality follows since $p_l = O(b^2)$ for $l \geq 2$. From this expression Chen provides the asymptotic approximation to the elements in $(Z'KZ/n)^{-1}$ (see Chen (2001), pg. 322). Secondly, $n^{-1}Z'K(M - Z\theta)$ has typical element

$$n^{-1}\sum_i K_{z,b}(z_i)\mathbb{I}(z_i > z)(z_i - z)^l(m(z_i) - m^+(z) - m'^+(z)(z_i - z))$$

for $l = 0, 1$. From assumption 2.2.3 it follows that

$$\begin{aligned} & E\left\{\frac{1}{n}\sum_i K_{z,b}(z_i)\mathbb{I}(z_i > z)(z_i - z)^l(m(z_i) - m^+(z) - m'^+(z)(z_i - z))\right\} \\ = \; & E\left\{\frac{1}{n}\sum_i K_{z,b}(z_i)\mathbb{I}(z_i > z)(z_i - z)^l\left(\frac{1}{2}m''^+(z)(z_i - z)^2 + \zeta(z_i)\right)\right\} \\ = \; & \frac{1}{2}m''^+(z)E(f(\xi_i)(\xi_i - z)^{2+l}) + o(b^2)E(f(\xi_i)(\xi_i - z)^l) \\ = \; & \frac{1}{2}m''^+(z)\left(f(z)p_{2+l}(z) + O(b^2)\right) + o(b^2)\left(f(z)p_l + f'(z)p_{l+1} + \frac{f''(z)}{2}p_{l+2} + O(b^2)\right) \\ = \; & \frac{1}{2}m''^+(z)f(z)p_{2+l}(z) + o(1) \text{ for } l = 0, 1. \qquad \text{(A-4)} \end{aligned}$$

The variance of each of these terms is such that,

$$\begin{aligned}
& var\left(\frac{1}{n}\sum_i K_{z,b}(z_i)\mathbb{I}(z_i > z)(z_i - z)^l \left\{\frac{1}{2}m''^{+}(z)(z_i - z)^2 + \zeta(z_i)\right\}\right) \\
\leq\ & \frac{1}{n}E\left\{K^2_{z,b}(z_i)\mathbb{I}(z_i > z)(z_i - z)^{4+2l}\frac{1}{4}\left(m''^{+}(z)\right)^2\right\} \\
+\ & \left\{o(b^2)\right\}^2 \frac{1}{n}E\left\{K^2_{z,b}(z_i)(z_i - z)^{2l}\right\} \\
\propto\ & \frac{A_b(z)}{n}E\left\{(\xi_i - z)^{4+2l}f(\xi_i)\right\} + \left\{o(b^2)\right\}^2 \frac{A_b(z)}{n}E\left\{(\xi_i - z)^{2l}f(\xi_i)\right\} \\
=\ & \frac{A_b(z)}{n}O(b^2) + \left\{o(b^2)\right\}^2 \frac{A_b(z)}{n}E\left\{f(z)p_{2l}(z) + O(b^2)\right\} \qquad \text{(A-5)}
\end{aligned}$$

At the boundary, $A_b(z) = O(1/b)$ and so the first term above is $o(1)$, while for $l = 0, 1$, the second term is $o(1/nb)$. Therefore the above variance is bounded by a quantity of order $o(1)$. The result then follows from (A-4), since $p_2(z) = b^2(2 + (z/b)) \to b^2(2 + \kappa)$ at the boundary (see property 2.1). However, since $z = 0$, the above simplifies to $p_2(0) = 2b^2$, and the result then follows.

The above results depend on the kernel density only through $p_j(z)$. However, in the boundary, property 2.2 establishes that the Beta and Gamma kernels have identical moments p_1 and p_2 (up to a negligible term) and therefore the expression of the bias above coincides in both cases. ■

Proof of Lemma 2.1.

We only need to establish a Central Limit Theorem for the second term in equation (A-1). Begin by noting that the term $n^{-1}Z'Ku$ has elements $T_l(z) = \sum_{i=1}^n K_{z,b}(z_i)\mathbb{I}(z_i > z)(z_i - z)^l u_i$, for $l = 0, 1$ with zero mean and conditional variance,

$$\begin{aligned}
var(T_l(z)|z_1, \ldots, z_n) &= \frac{1}{n}E\left(K^2_{z,b}(z_i)\mathbb{I}(z_i > z)(z_i - z)^{2l}\sigma^2(z_i)\right) \\
&= \frac{A_b(z)}{n}E\left(f(\xi_i)\sigma^2(\xi_i)(z_i - z)^{2l}\right) \\
&= \frac{A_b(z)}{n}\left[f(z)\sigma^{2+}(z)p_{2l}(z) + O(b^2)\right] \\
&= \frac{\Gamma(2\kappa + 1)}{nb\Gamma^2(\kappa + 1)2^{2\kappa+1}}f(z)\sigma^{2+}(z)p_{2l}(z) + o(1) \qquad \text{(A-6)}
\end{aligned}$$

for $l = 0, 1$. Let $\lambda = (\lambda_1, \lambda_2)'$ be such that $\lambda'\lambda = 1$, and consider,

$$\sqrt{nb}\lambda' \frac{Z'Ku}{n} = \sum_{i=1}^{n} \frac{t_{ni}}{\sqrt{n}}$$

where

$$t_{ni} = \sqrt{b}\lambda_1 K_{z,b}(z_i)\mathbb{I}(z_i > z)u_i + \sqrt{b}\lambda_2 K_{z,b}(z_i)\mathbb{I}(z_i > z)(z_i - z)u_i$$

A sufficient Lyapounov condition is

$$\lim_{n\to\infty} \sum_{i=1}^{n} E\left|\frac{t_{ni}}{\sqrt{n}}\right|^{2+\delta} = \lim_{n\to\infty} \frac{1}{n^{2+\delta/2}} \sum_{i=1}^{n} E\left|t_{ni}\right|^{2+\delta} = \lim_{n\to\infty} n^{-1-\delta/2} E|t_{ni}|^{2+\delta} = 0,$$

and from this it follows that

$$\begin{aligned} |t_{ni}|^{2+\delta} &= E\left|\sqrt{b}\lambda_1 K_{z,b}(z_i)\mathbb{I}(z_i > z)u_i + \sqrt{b}\lambda_2 K_{z,b}(z_i)\mathbb{I}(z_i > z)(z_i - z)u_i\right|^{2+\delta} \\ &\leq 2^{1+\delta} E\left|\sqrt{b}\lambda_1 K_{z,b}(z_i)\mathbb{I}(z_i > z)u_i\right|^{2+\delta} + 2^{1+\delta} E\left|\sqrt{b}\lambda_2 K_{z,b}(z_i)\mathbb{I}(z_i > z)(z_i - z)u_i\right|^{2+\delta} \end{aligned}$$

Now, the first term is such that,

$$E\left|K_{z,b}(z_i)u_i\right|^{2+\delta} = |u|_i^{2+\delta} \int_0^\infty |K_{z,b}(z_i)|^{2+\delta} f(z_i)dz_i \qquad \text{(A-7)}$$

Note that

$$\begin{aligned} K_{z,b}(z_i)^{2+\delta} &= \frac{z_i^{(2+\delta)z/b} e^{-(2+\delta)z_i/b}}{b^{z(2+\delta)/b+(2+\delta)}\Gamma^{(2+\delta)}(z/b+1)} \frac{\Gamma((2+\delta)z/b+1)(2+\delta)^{-(2+\delta)z/b-1}}{\Gamma((2+\delta)z/b+1)(2+\delta)^{-(2+\delta)z/b-1}} \\ &= A_b^*(z) K_{x,b}^*(z_i), \end{aligned} \qquad \text{(A-8)}$$

where now,

$$\begin{aligned} A_b^*(z) &= \frac{(2+\delta)^{\frac{-(2+\delta)z}{b}} \Gamma\left(\frac{(2+\delta)z}{b} + 1\right)}{b^{1+\delta}\Gamma^{(2+\delta)}\left(\frac{x}{b} + 1\right)} \\ K_{x,b}^*(z_i) &= \frac{z_i^{k-1} e^{-z_i/\theta}}{\theta^k \Gamma(k)} \text{ for } \theta = \frac{b}{(2+\delta)} \text{ and } k = \frac{(2+\delta)z}{b} + 1 \end{aligned}$$

so that $p_1(z) = 1$ and $p_2(z) \propto zb(2+\delta) + b^2$. Given the boundary condition $z/b \to \kappa$,

$$A_b^*(z) \sim \frac{(2+\delta)^{-(2+\delta)\kappa}\Gamma((2+\delta)\kappa + 1)}{b^{1+\delta}\Gamma^{(2+\delta)}(\kappa+1)} = O(1/b^{1+\delta}). \qquad \text{(A-9)}$$

Thus, at the boundary,

$$\begin{aligned} E\left|K_{z,b}(z_i)u_i\right|^{(2+\delta)} &= |u|_i^{(2+\delta)} A_b^*(z) E\left\{f(\xi_i)\right\} \\ &= |u|_i^{(2+\delta)} A_b^*(z) f(z) + |u|_i^{(2+\delta)} A_b^*(z) f'(z) + O(1/b^{\delta}) \\ &= O(1/b^{1+\delta}) \qquad \text{(A-10)} \end{aligned}$$

and similarly,

$$\begin{aligned} E\left|K_{z,b}(z_i)(z_i - z)u_i\right|^{(2+\delta)} &= |u_i|^{(2+\delta)} A_b^*(z) E\left\{f(\xi_i)(\xi_i - z)^{(2+\delta)}\right\} \\ &= |u_i|^{(2+\delta)} A_b^*(z) f(z) p_{(2+\delta)}(z) + O(b^2) \\ &= O(1/b^{1+\delta}) O(b^2) = o(1) \qquad \text{(A-11)} \end{aligned}$$

Therefore,

$$\lim_{n\to\infty} \frac{1}{n^{1+\delta/2}} E|t_{ni}|^{2+\delta} \leq \lim_{n\to\infty} \frac{b^{(2+\delta)/2}}{n^{1+\delta/2}} \left(O\left(\frac{1}{b^{1+\delta}}\right) + o(1) \right) = o(1)$$

provided that $nb \to \infty$ from which it follows that

$$\sqrt{nb}\left\{\hat{m}(x)^+ - m(x)^+\right\} - \frac{\sqrt{nb}}{2}(2+\kappa) m''^+(x) b^2 \sim N\left(0, \frac{\sigma^{2+}(x)\Gamma(2\kappa+1)}{2^{2\kappa+1}\Gamma(\kappa+1) f(x)}\right) \qquad \text{(A-12)}$$

Noting that, by definition, $z = 0$ (so that $\kappa = 0$), the lemma follows. Once again, the above result depends on the type of kernel only through $p_l(z)$ for $l = 0, 1, 2$, and by property 2.2 it folows that the one can extrapolate the conclusion to the case of a LLR with a beta kernel.

Proof of Theorem 2.1.

To prove the theorem, we first establish the covariance between $\hat{m}^+(z)$ and $\hat{p}^+(z)$. Consider firstly the terms $n^{-1}Z'Ku$ and $n^{-1}Z'Kv$ and note that,

$$\begin{aligned} & E\left(\frac{1}{n}\sum_{i=1}^{n} K_{z,b}(z_i)(z_i - z)^l u_i \frac{1}{n}\sum_{i=1}^{n} K_{z,b}(z_i)(z_i - z)^l v_i | z_1, \ldots, z_n\right) \\ = \; & \frac{1}{n} E\left\{K_{z,b}^2(z_i)(z_i - z)^{2l}\eta(z_i)\right\} \\ = \; & \frac{A_b(z)}{n}\left(f(z)\eta^+(z) p_{2l}(z) + O(b^{l+1})\right) \qquad \text{(A-13)} \end{aligned}$$

The final expression for the covariance between $\hat{m}^+$ and $\hat{p}^+$ follows from A-1, A-3 and results in page 322 of Chen (2001).

With the expression for the bias and the central limit theorem one deduces that

$$\sqrt{nb}\begin{pmatrix} (\hat{m}^+ - \hat{m}^-) - (m^+ - m^-) \\ (\hat{p}^+ - \hat{p}^-) - (p^+ - p^-) \end{pmatrix} \rightarrow \mathcal{N}\left(\begin{pmatrix} \tau_m \\ \tau_p \end{pmatrix}, \begin{pmatrix} \lambda_m & \lambda_{mp} \\ \lambda_{mp} & \lambda_p \end{pmatrix}\right) \tag{A-14}$$

where $\tau_m = \frac{b^2\sqrt{nb}}{2}(2+\kappa)(m''_+(z) - m''_-(z))$, $\lambda_m = f^{-1}(z)C(\sigma^2_+(z) - \sigma^2_-(z))$, $\lambda_{mp} = Cf^-1(z)(\eta^+(z) - \eta^-(z))$, and similarly for the remaining terms. Finally, a Taylor expansion (see Hahn et al. (1999) or proposition 1 in Porter (2003)), yields,

$$\sqrt{nb}\left(\frac{\hat{m}^+ - \hat{m}^-}{\hat{p}^+ - \hat{p}^-} - \frac{m^+ - m^-}{p^+ - p^-}\right) \rightarrow \mathcal{N}(\lambda, \tau) \tag{A-15}$$

where,

$$\lambda = \frac{b^2\sqrt{nb}}{2}(2+\kappa)\left(\frac{1}{p^+ - p^-}(m''_+ - m''_-) - \frac{m^+ - m^-}{(p^+ - p^-)^2}(p''_+ - p''_-)\right) \tag{A-16}$$

and

$$\begin{aligned} \tau &= \frac{C}{f(z)}\left(\frac{1}{(p^+ - p^-)^2}(\sigma^2_+ - \sigma^2_-)\right. \\ &- 2\frac{m^+ - m^-}{(p^+ - p^-)^3}(\eta^+ - \eta^-) \\ &+ \left.\frac{(m^+ - m^-)^2}{(p^+ - p^-)^4}(p^+(1-p^+) - p^-(1-p^-))\right) \end{aligned} \tag{A-17}$$

for $C = \frac{\Gamma(2\kappa+1)}{\Gamma^2(\kappa+1)2^{2\kappa+1}}$. The result then follows by noting that κ is the limit of z/b (if using gamma kernels or if estimation is done using beta a beta kernel in the boundary about 0) or $(1-z)/b$ (if estimation is done with the beta kernel within a neighborhood of 1). In either case, $\kappa = 0$.

Figure 1: log(MSE). Model 1, N=500, R=10.000

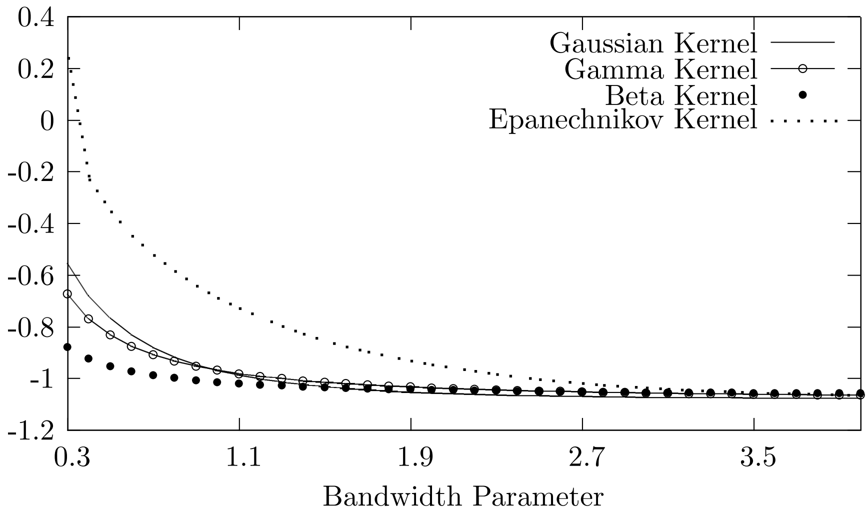

Figure 2: log(MSE). Model 1, N=1000, R=10.000

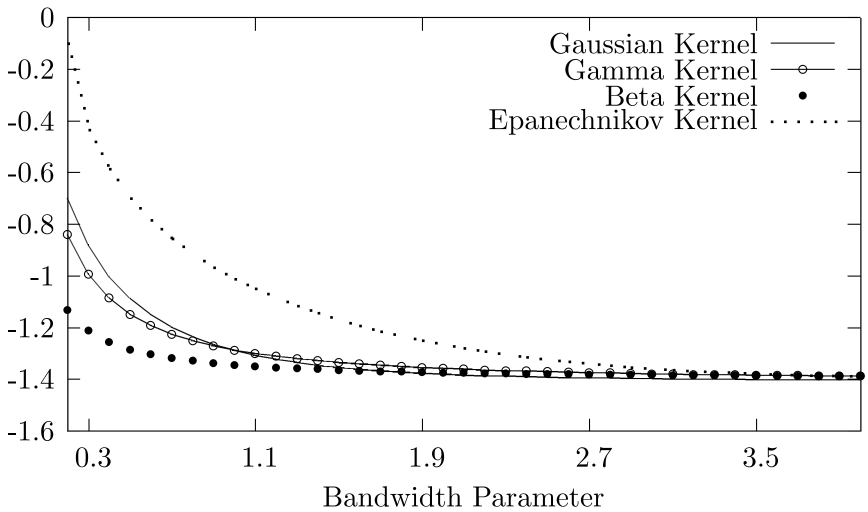

Figure 3: log(MSE). Model 1, N=2000, R=10.000

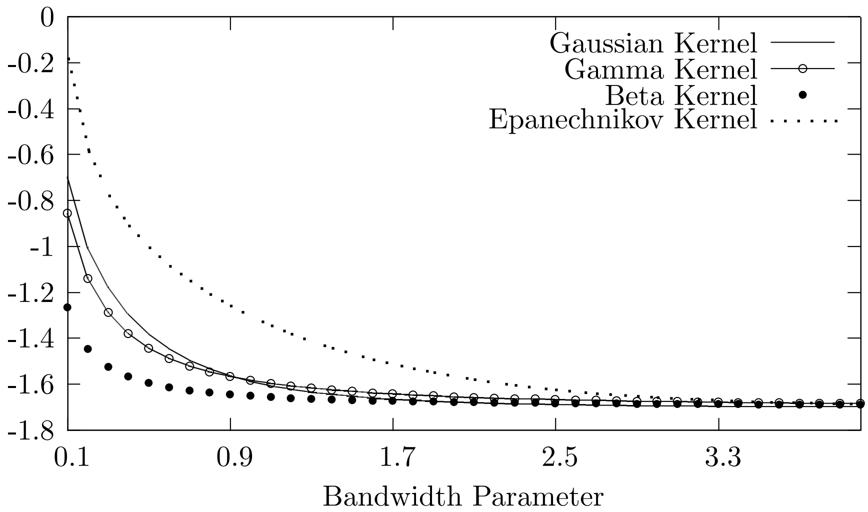

Figure 4: log(MSE). Model 2, N=500, R=10.000

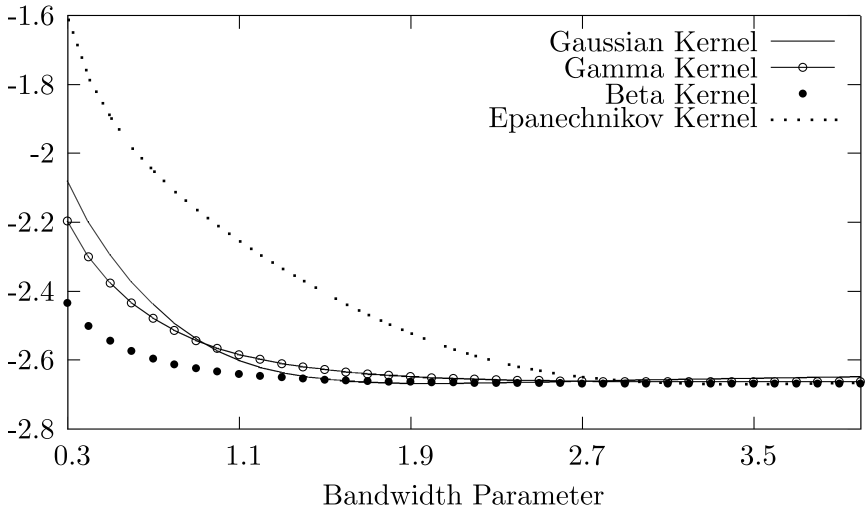

Figure 5: log(MSE). Model 2, N=1000, R=10.000

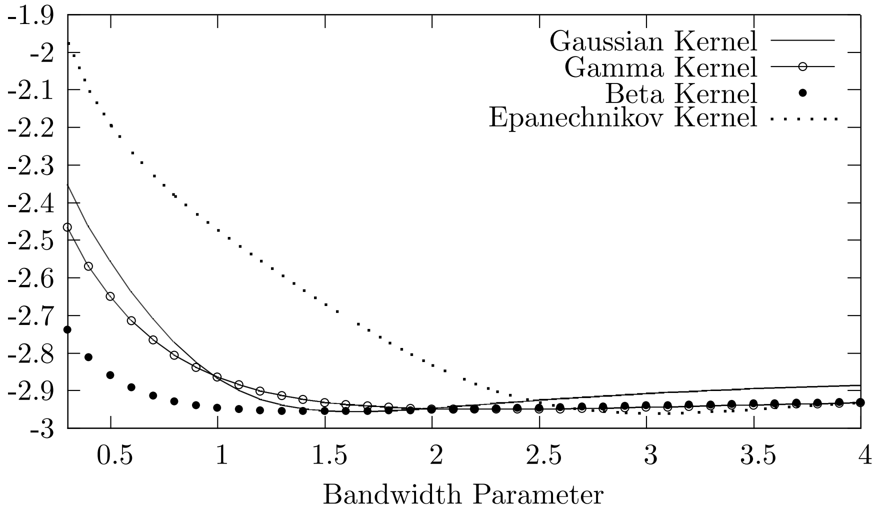

Figure 6: log(MSE). Model 2, N=2000, R=10.000

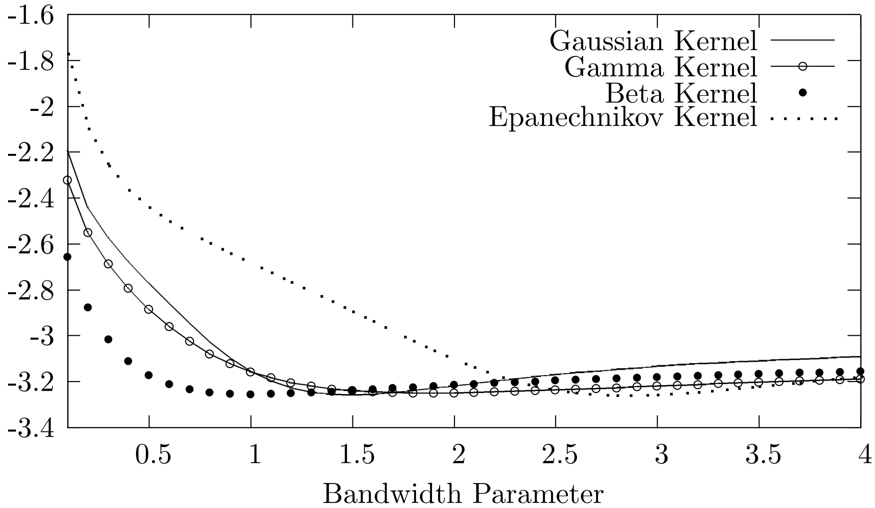

Figure 7: log(MSE). Model 3, N=500, R=10.000

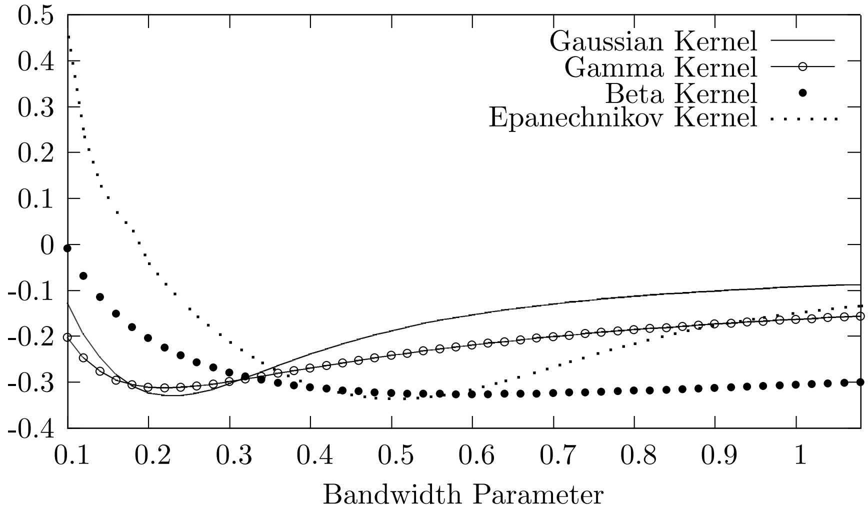

Figure 8: log(MSE). Model 3, N=1000, R=10.000

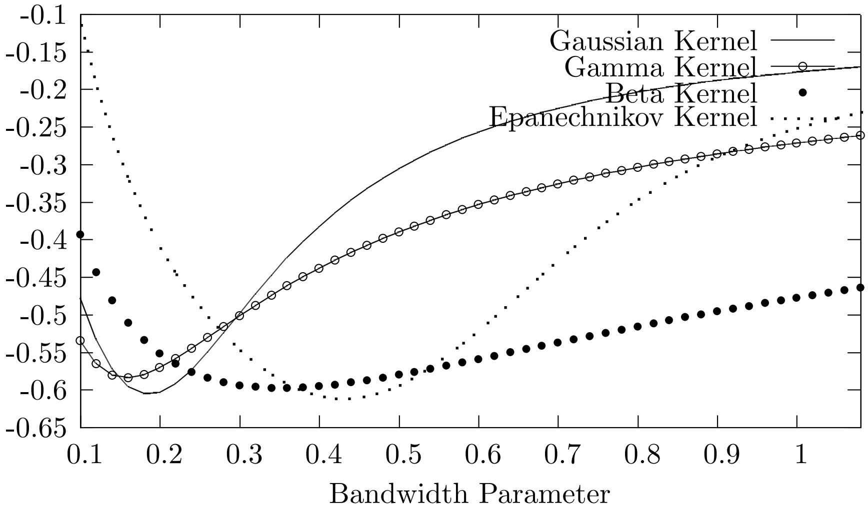

Figure 9: log(MSE). Model 3, N=2000, R=10.000