Boundary effects in kernel regression

Xingyu Tao, Xuan Li, Sven Jacobs

January 19, 2022

Research Module in Econometrics & Statistics

Introduction

Kernel regression

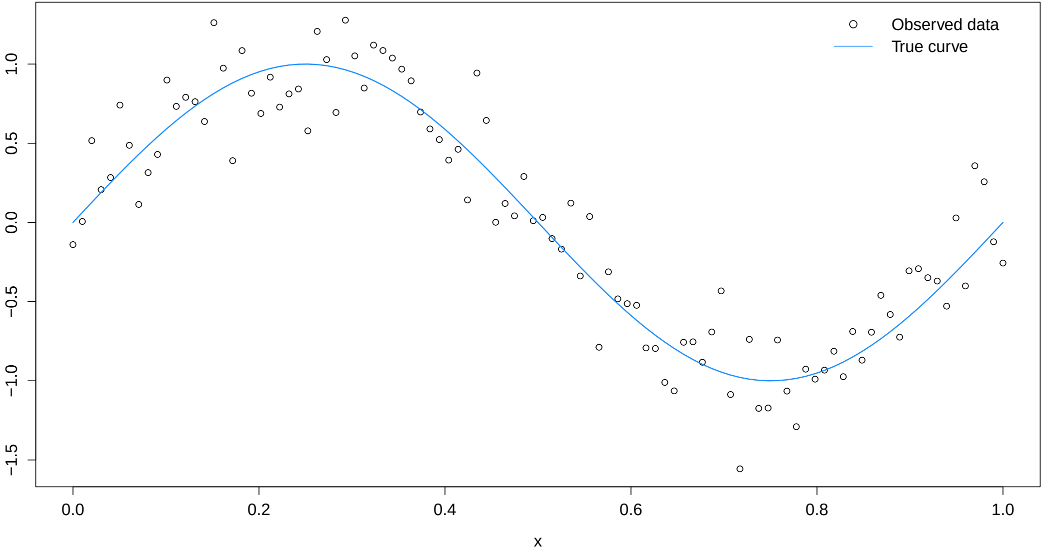
Setting

- iid data: $\{(X_i, Y_i)\}_{i=1,\ldots,n}$
- Random design: $X \sim f_X, X_i \in [a,b] \subset \mathbb{R}$

$$Y_i = \underbrace{m(X_i)}_{\text{target}} + \epsilon_i$$

- $m(X_i) = \mathrm{E}[Y_i | X = X_i]$ (CEF)
- $\mathrm{E}[\epsilon_i | X = X_i] = 0, \mathrm{Var}(\epsilon_i | X = X_i) = \sigma^2(X_i)$

Example



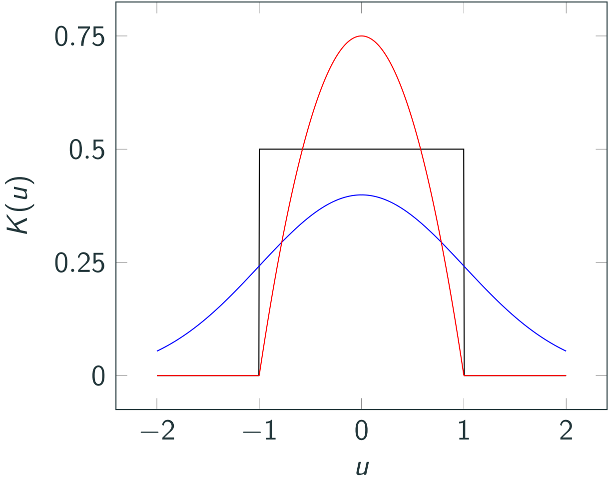
Nadaraya-Watson estimator

- Approximation of $m(x)$ by a weighted local average of the Y_i's:

$$\hat{m}_{\mathrm{NW}}(x) = \sum_{i=1}^{n} w(x, X_i, h) Y_i = \sum_{i=1}^{n} \frac{K\left(\frac{x-X_i}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right)} Y_i$$

- Kernel K: – Weight of Y_i smaller if $|x - X_i|$ larger
  – Most often: symmetric density
- Bandwidth h: – Smoothing parameter
  – Determines how fast weights decrease in $|x - X_i|$

Selection of kernels



Uniform

$$K(u) = \begin{cases} \frac{1}{2} & \text{if } |u| \leq 1 \\ 0 & \text{else} \end{cases}$$

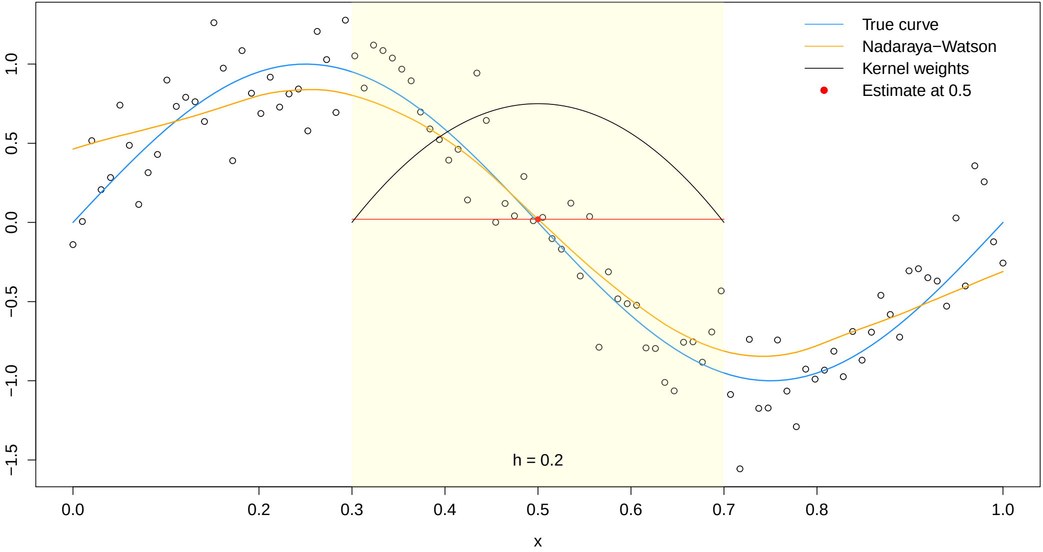
Gaussian

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

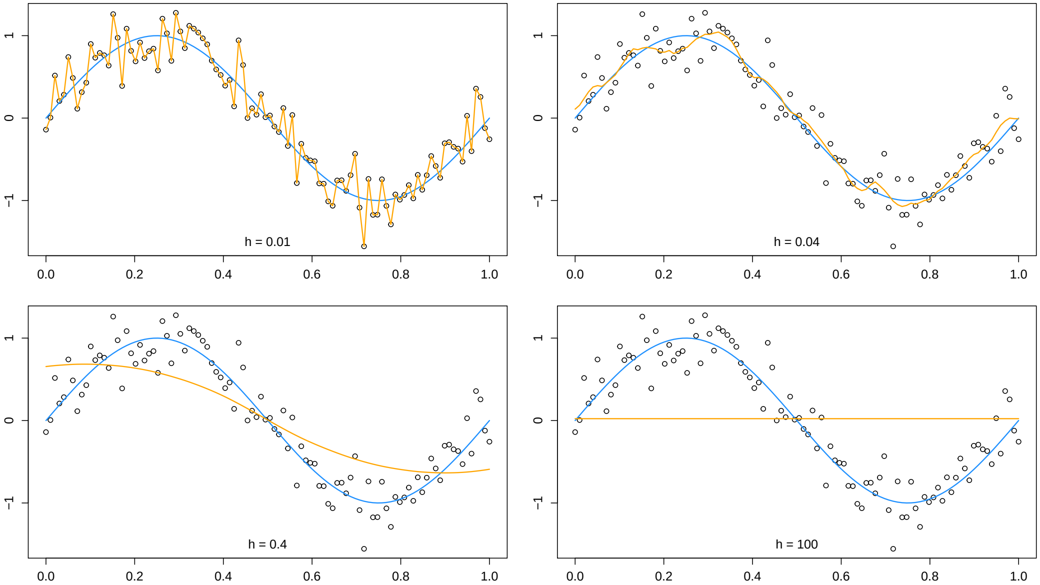
Epanechnikov

$$K(u) = \begin{cases} \frac{3}{4}(1 - u^2) & \text{if } |u| \leq 1 \\ 0 & \text{else} \end{cases}$$

Example continued: Nadaraya-Watson



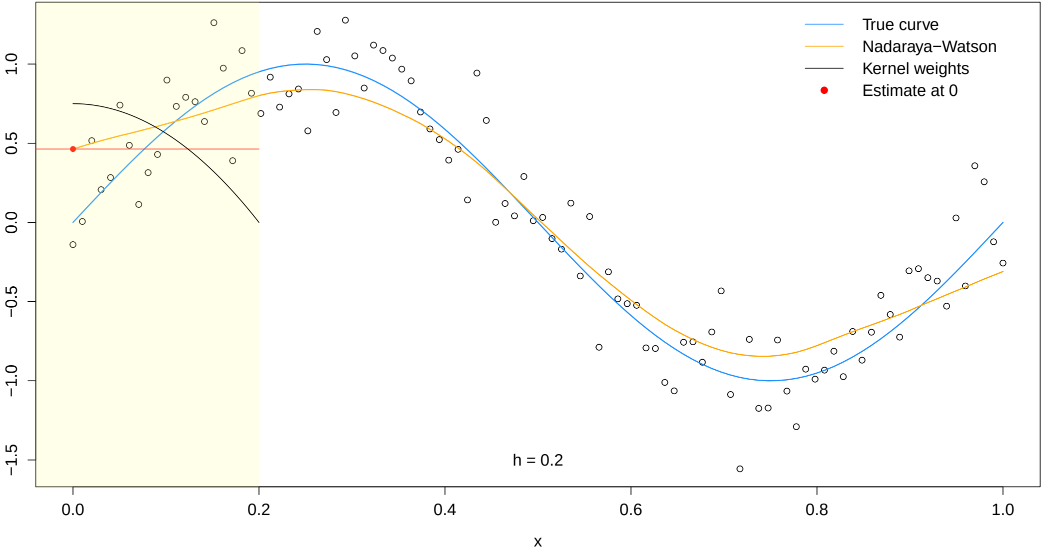
Role of bandwidth



Introduction

Motivation

Example continued: Boundary effects



Solution approaches

1. Local linear regression
   - Automatic boundary adaption

Solution approaches

1. Local linear regression
   - Automatic boundary adaption
2. Special boundary kernels
   - Adjustment of NW estimator near boundaries

Local linear regression

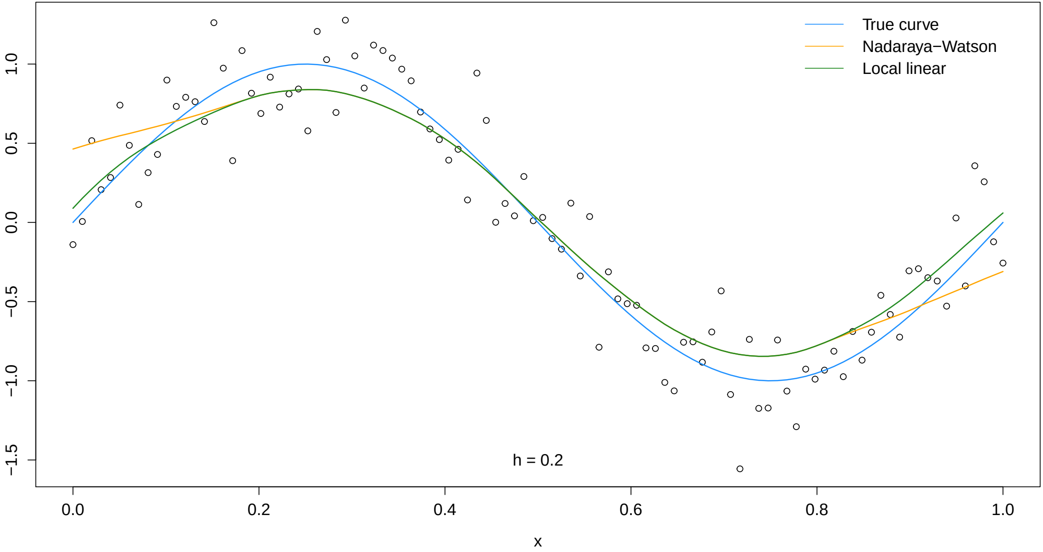
Definition

- Weighted local fitting of a line

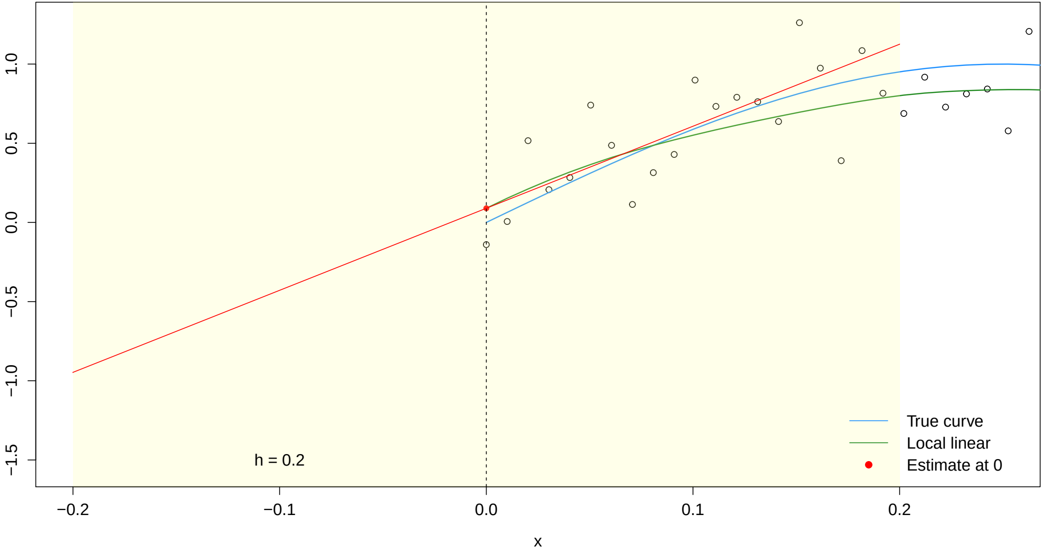
$$
\begin{aligned}
\hat{m}_{\text{NW}}(x) &= \arg\min_{\beta_0} \sum_{i=1}^{n} (Y_i - \beta_0)^2 K\left(\frac{x - X_i}{h}\right) \\
\hat{m}_{\text{LL}}(x) &= \arg\min_{\beta_0} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1(X_i - x))^2 K\left(\frac{x - X_i}{h}\right) \\
&= \sum_{i=1}^{n} w(x, X_i, h) Y_i
\end{aligned}
$$

$\Rightarrow$ Weighted Least Squares (WLS)

Example continued: Local linear



Example continued: Local linear



Local linear regression

Theoretical comparison

Estimation in the interior: Assumptions

(A1) m is twice continuously differentiable

(A2) Regularity conditions on f, σ^2 and K

(A3) $h \to 0, nh \to \infty$ as $n \to \infty$

(A4) $x \in [a + h, b - h]$

Estimation in the interior: Asymptotics

$$\text{ABias}(\hat{m}_{\text{NW}}(x)|\boldsymbol{X}) = \frac{1}{2}\kappa_2(K)m''(x)h^2 + \kappa_2(K)\frac{m'(x)f'(x)}{f(x)}h^2 \qquad = \mathcal{O}\left(h^2\right)$$

$$\text{ABias}(\hat{m}_{\text{LL}}(x)|\boldsymbol{X}) = \frac{1}{2}\kappa_2(K)m''(x)h^2 \qquad = \mathcal{O}\left(h^2\right)$$

$$\text{AVar}(\hat{m}_{\text{NW}}(x)|\boldsymbol{X}) = \text{AVar}(\hat{m}_{\text{LL}}(x)|\boldsymbol{X}) = \frac{R(K)\sigma^2(x)}{f(x)}\frac{1}{nh} \qquad = \mathcal{O}\left(\frac{1}{nh}\right)$$

- Bias of LL typically smaller but of same order
- Asymptotic variance is the same

Estimation at the boundary: Assumptions

(A1) m is twice continuously differentiable

- $\lim_{x \downarrow a} m''(x) = m''(a)$ and $\lim_{x \uparrow b} m''(x) = m''(b)$

(A2) Regularity conditions on f, σ^2 and K

(A3) $h \to 0, nh \to \infty$ as $n \to \infty$

(A4) $x \in \{a, b\}$

Estimation at the boundary: Asymptotics

$$\mathrm{ABias}(\hat{m}_{\mathrm{NW}}(a)|\boldsymbol{X}) = 2\int_0^\infty uK(u)\,\mathrm{d}u\, m'(a)h \qquad = \mathcal{O}(h)$$

$$\mathrm{ABias}(\hat{m}_{\mathrm{NW}}(b)|\boldsymbol{X}) = -2\int_0^\infty uK(u)\,\mathrm{d}u\, m'(b)h \qquad = \mathcal{O}(h)$$

$$\mathrm{ABias}(\hat{m}_{\mathrm{LL}}(a)|\boldsymbol{X}) = \frac{1}{2}\kappa_2(\tilde{K})m''(a)h^2 \qquad = \mathcal{O}\left(h^2\right)$$

$$\mathrm{AVar}(\hat{m}_{\mathrm{LL}}(a)|\boldsymbol{X}) = \frac{R(\tilde{K})\sigma^2(a)}{f(a)}\frac{1}{nh} \qquad = \mathcal{O}\left(\frac{1}{nh}\right)$$

- ABias of NW is $\mathcal{O}(h)$ instead of $\mathcal{O}(h^2)$
- LL preserves the order (but constants differ)

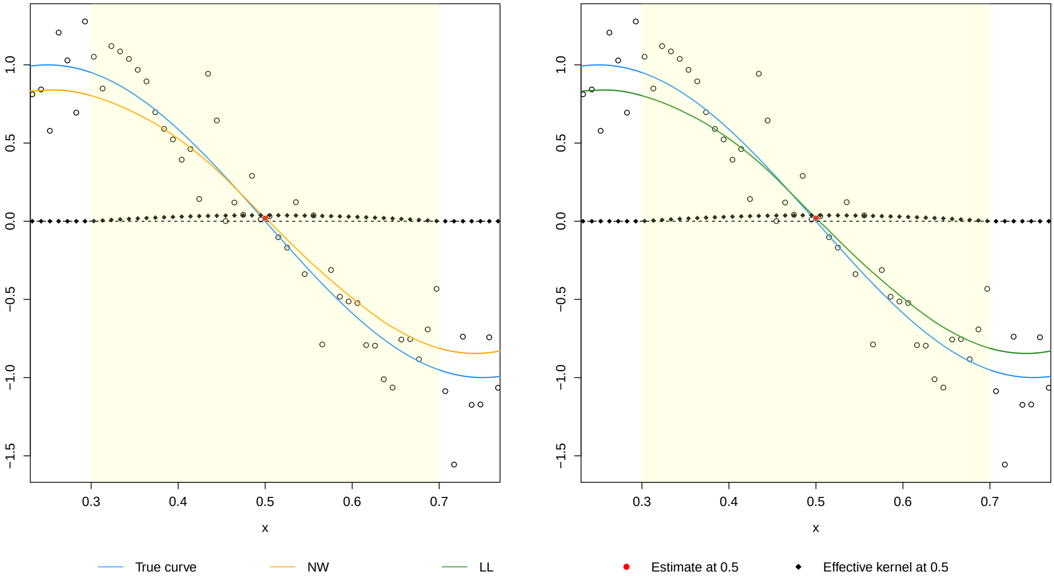
Local linear regression

Graphical comparison

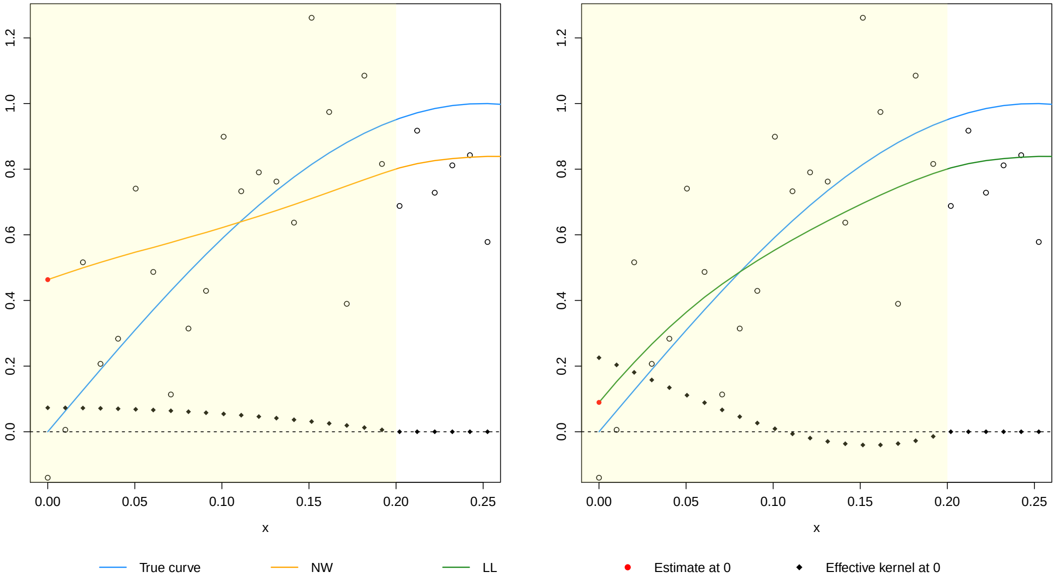
Effective kernels

- NW and LL both linear smoothers: $\hat{m}(x) = \sum_{i=1}^{n} w(x, X_i, h) Y_i$
- $\{w(x, X_i, h)\}_{i=1}^{n}$: Effective kernel at x

Effective kernel in the interior



Effective kernel at the boundary



Boundary kernels

Idea

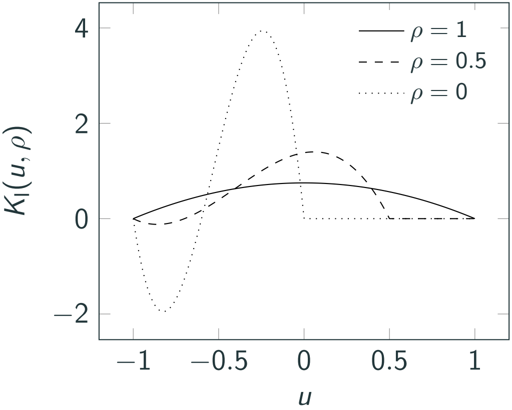
- Assume that $\operatorname{supp}(K) = [-1, 1]$ and w.l.o.g. $X_i \in [0, 1]$
  $\Rightarrow$ Boundary region: $[0, h) \cup (1 - h, 1]$
- Idea: Construct asymmetric kernel that restores the original moment properties

- For $x = \rho h$ a left boundary kernel $K_\mathrm{l}(u, \rho)$ satisfies for $\rho \in [0, 1]$

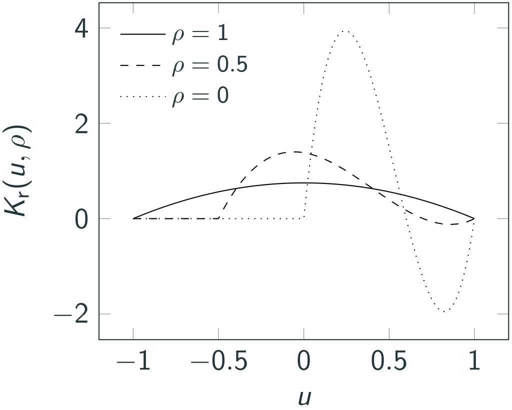
$$\int_{-1}^{\rho} u^j K_\mathrm{l}(u, \rho)\,\mathrm{d}u = \begin{cases} 1 & j = 0 \\ 0 & j = 1 \\ \kappa_{2,\rho}(K_\mathrm{l}) \neq 0 & j = 2 \end{cases}$$

- Note: Kernel shape changes with relative location of x to boundary!

Epanechnikov boundary kernels, Müller (1991)

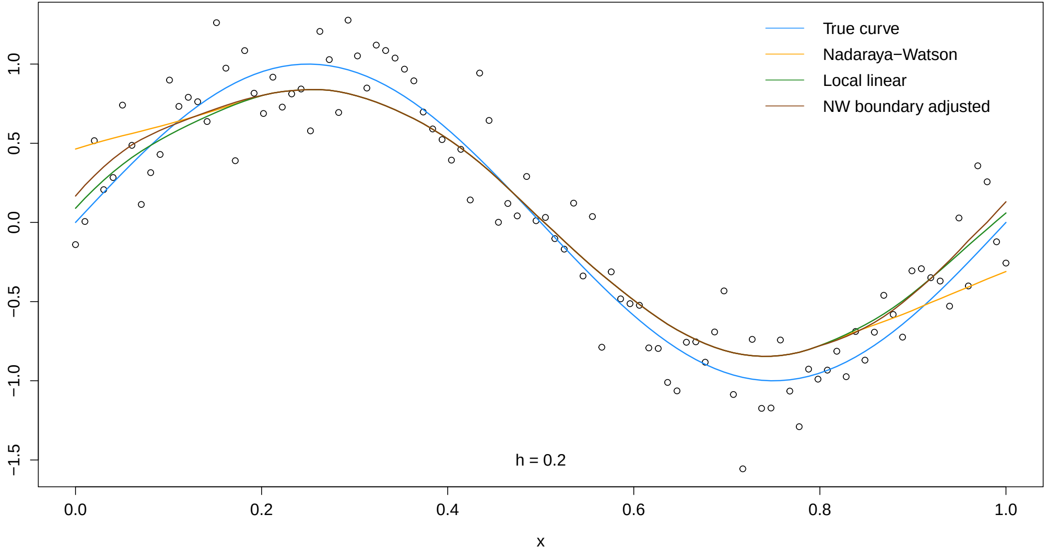


Left boundary kernels



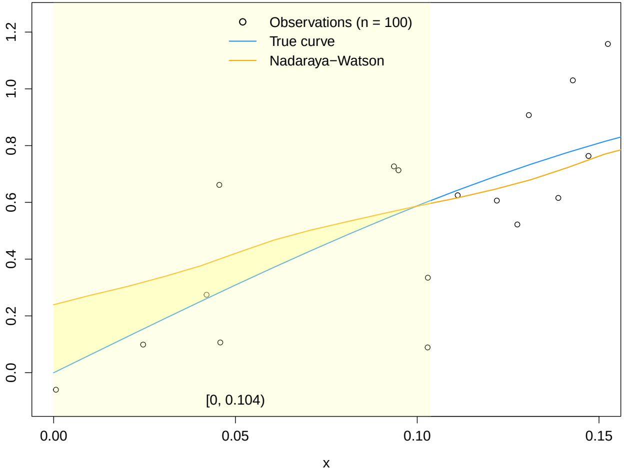
Right boundary kernels

Example continued: Boundary kernels



Simulation

Simulation set-up



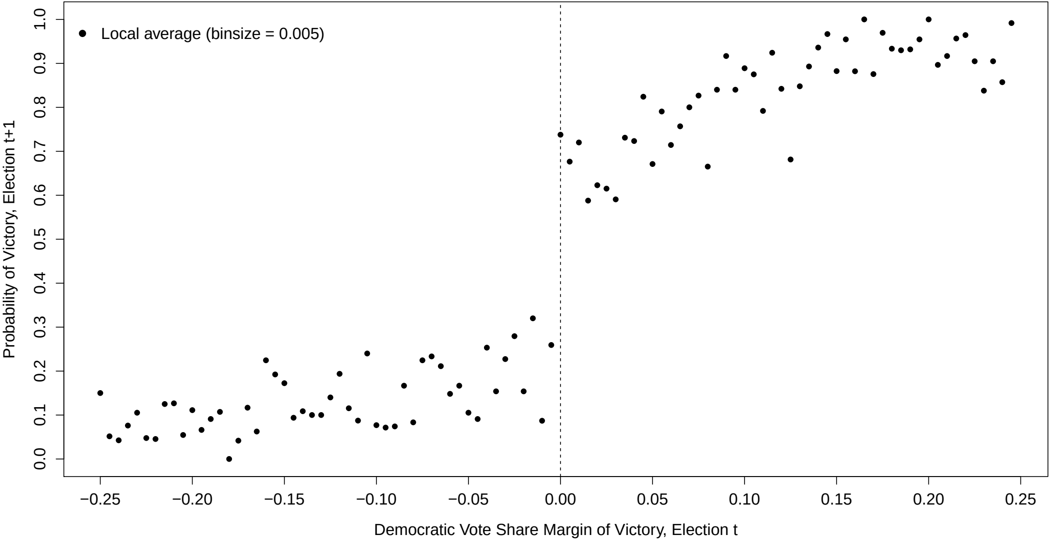
- $m(x) = \sin(2\pi x)$
- $X \sim \mathcal{U}(0, 1)$
- $\epsilon \sim \mathcal{N}(0, 0.25^2)$
- Target: Mean Integrated Squared Error (MISE) over boundary region
- Asymptotic optimal bandwidth: $h_{\text{AMISE}}(n)$
- 10000 Monte-Carlo repetitions
- Sample sizes from 50 to 5000

Simulation results

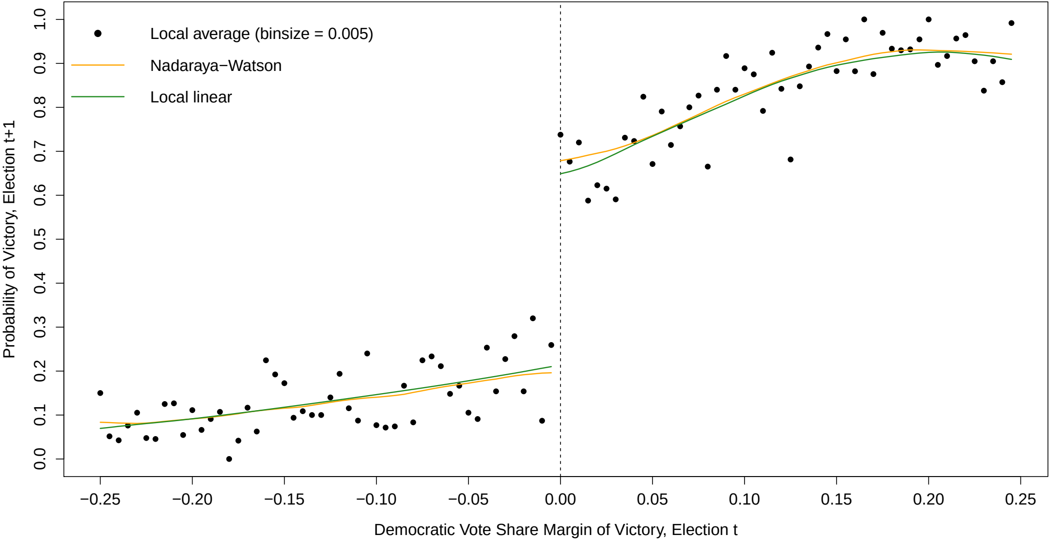
n	Boundary region	Change in MISE to NW [%]	
		Local linear	NW boundary adjusted
50	[0, 0.119)	789.29	1578.07
100	[0, 0.104)	−41.90	1143.28
250	[0, 0.086)	−65.13	1982.67
500	[0, 0.075)	−72.98	14.92
1000	[0, 0.065)	−78.80	−63.95
2500	[0, 0.055)	−85.12	−77.80
5000	[0, 0.047)	−88.50	−83.76

Application

Regression discontinuity design (RDD), Lee (2008)



Regression discontinuity design (RDD), Lee (2008)



Conclusion

Take-away

- Often in econometrics the boundaries are of great interest
- NW with potentially severe boundary bias $\rightarrow$ Adjustment!
- Boundary kernels with asymptotic correction, but complicated and impractical
- LL regression as a simple and intuitive method to automatically reduce boundary effects
    - Plus favorable asymptotic properties in interior

$\Rightarrow$ Recommendation: Local linear regression

Contact

Slides and codes are hosted on GitHub.

For further questions contact us via email.

github.com/svjaco

s6xitaoo(at)uni-bonn.de (Xingyu)

s6xuliii(at)uni-bonn.de (Xuan)

s.jacobs(at)uni-bonn.de (Sven)

Appendix

Local polynomial regression: Definition

$$
\begin{aligned}
\hat{m}_{\text{LP}}(x) &= \arg\min_{\beta_0} \sum_{i=1}^{n} \left(Y_i - \sum_{j=0}^{p} \beta_j (X_i - x)^j \right)^2 K\left(\frac{x - X_i}{h}\right) \\
&= \sum_{i=1}^{n} w(x, X_i, h) Y_i
\end{aligned}
$$

Local polynomial regression: Motivation

- Minimization of RSS

$$\sum_{i=1}^{n}(Y_i - \hat{m}(X_i))^2$$

- p-th order Taylor expansion:

$$m(X_i) \approx \sum_{j=0}^{p} \frac{m^{(j)}(x)}{j!}(X_i - x)^j$$

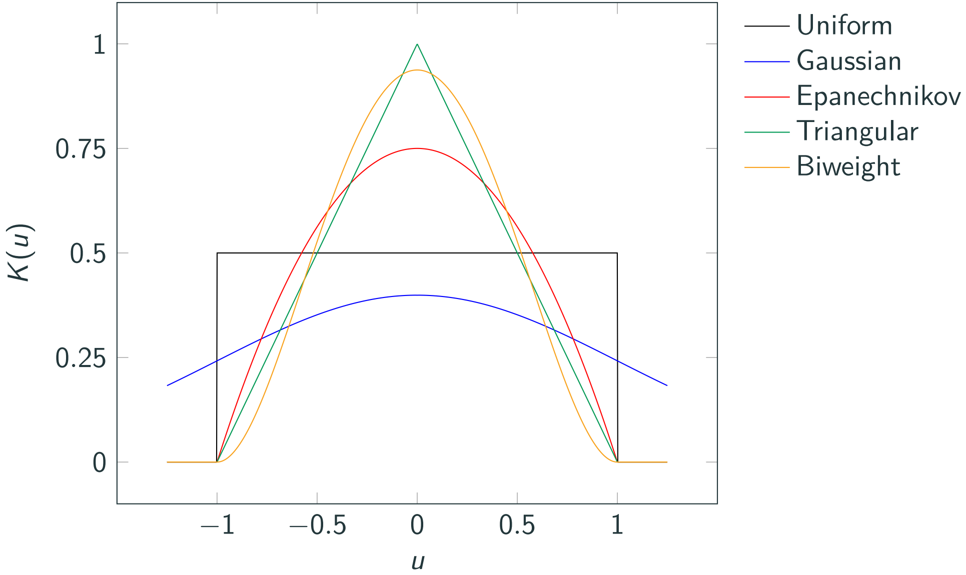
Local polynomial regression: Motivation

$$\sum_{i=1}^{n} \left(Y_i - \sum_{j=0}^{p} \underbrace{\frac{m^{(j)}(x)}{j!}}_{\equiv \beta_j} (X_i - x)^j \right)^2$$

- Weighting of observation (X_i, Y_i) using a kernel:

$$\hat{m}_{\text{LP}}(x) = \underset{\beta_0}{\arg\min} \sum_{i=1}^{n} \left(Y_i - \sum_{j=0}^{p} \beta_j (X_i - x)^j \right)^2 K\left(\frac{x - X_i}{h} \right)$$

Kernels: Prominent examples

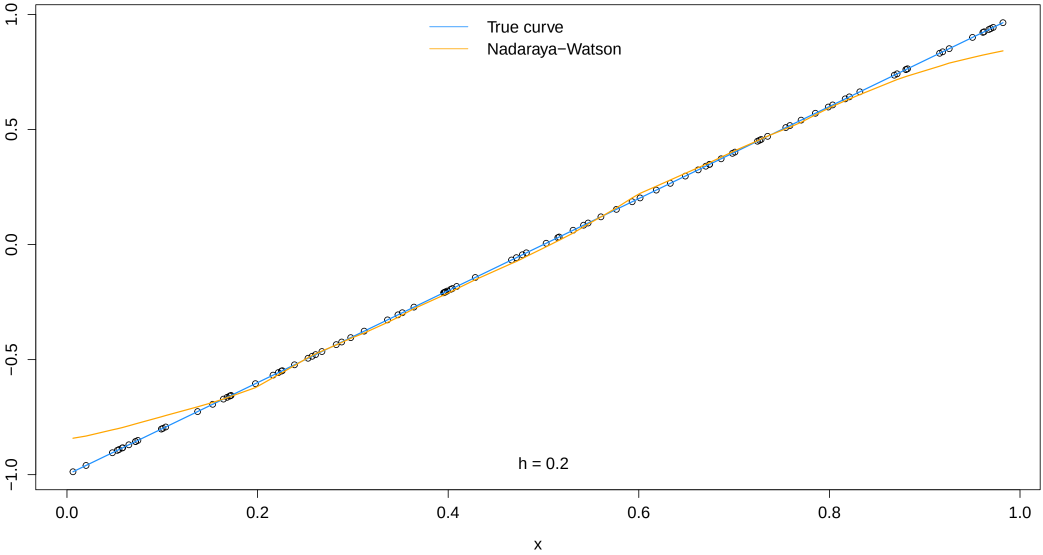


Kernels: Efficiency

Kernel	Function	$R(K)$	$\kappa_2(K)$	Efficiency
Epanechnikov	$K_{\mathrm{E}}(u) = \frac{3}{4}(1 - u^2)$	$\frac{3}{5}$	$\frac{1}{5}$	100.0%
Biweight	$K_{\mathrm{B}}(u) = \frac{15}{16}(1 - u^2)^2$	$\frac{5}{7}$	$\frac{1}{7}$	99.39%
Triangular	$K_{\mathrm{T}}(u) = 1 - \lvert u \rvert$	$\frac{2}{3}$	$\frac{1}{6}$	98.59%
Gaussian	$K_{\mathrm{G}}(u) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{u^2}{2}}$	$\frac{1}{2\sqrt{\pi}}$	1	95.12%
Uniform	$K_{\mathrm{U}}(u) = \frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	92.95%

$$\mathrm{Eff}(K) \equiv \frac{\kappa_2(K_{\mathrm{E}})^{\frac{1}{2}} R(K_{\mathrm{E}})}{\kappa_2(K)^{\frac{1}{2}} R(K)}$$

Example II: Boundary effects



Others: Assumptions Section 2

- (A1) m is twice continuously differentiable
- (A2) f is continuously differentiable and $f(x) > 0$
- (A3) σ^2 is continuous and $\sigma^2(x) > 0$
- (A4) K is a symmetric and bounded pdf with $\kappa_2(K) \equiv \int u^2 K(u)\,\mathrm{d}u < \infty$ (finite variance) and $R(K) \equiv \int K(u)^2\,\mathrm{d}u < \infty$ (square integrability)
- (A5) $h \to 0, nh \to \infty$ as $n \to \infty$
- (A6) $x \in [a+h, b-h]$ / $x \in \{a, b\}$

Effective kernels: Insights

- LL regression automatically modifies the kernel to correct the bias exactly to first order when there is asymmetry in the smoothing window

$$
\begin{aligned}
\mathrm{E}[\hat{m}(x)|\boldsymbol{X}] &= \sum_{i=1}^{n} w_i(x) m(X_i) \\
&= m(x) \underbrace{\sum_{i=1}^{n} w_i(x)}_{=1} + m'(x) \underbrace{\sum_{i=1}^{n} w_i(x)(X_i - x)}_{=0 \text{ for LL}} + R
\end{aligned}
$$

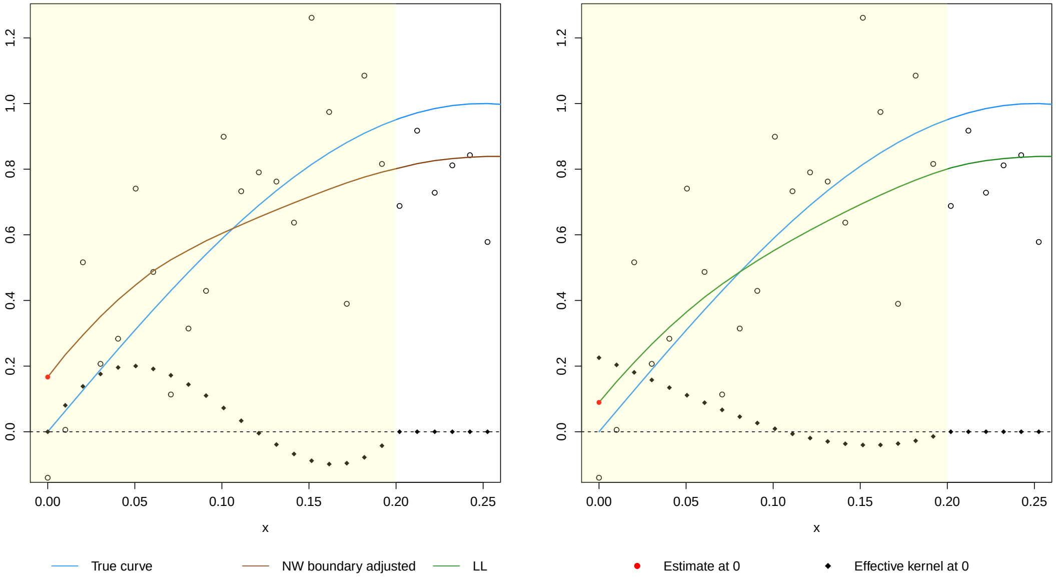
Others: Epanechnikov boundary kernel, Müller (1991, Table 1)

For $0 \le \rho \le 1$:

$$K_{\mathrm{l}}(u,\rho) = 6(1+u)(\rho-u)\frac{1}{(1+\rho)^3}\left\{1+5\left(\frac{1-\rho}{1+\rho}\right)^2+10\frac{1-\rho}{(1+\rho)^2}u\right\} \quad , u \in [-1,\rho]$$

$$K_{\mathrm{r}}(u,\rho) = K_{\mathrm{l}}(-u,\rho) \quad , u \in [-\rho,1]$$

Others: Effective kernel at the boundary



Others: Optimal bandwidths

- Minimization of (weighted) AMISE:

$$h_{\mathrm{AMISE}} \equiv \arg\min_{h>0} \mathrm{AMISE}(\hat{m}|\boldsymbol{X})$$

$$h^{\mathrm{LL}}_{\mathrm{AMISE}}(n) = \left\{ \frac{R(K)\int \sigma^2(x)\,\mathrm{d}x}{\kappa_2(K)^2 \int (m''(x))^2 f(x)\,\mathrm{d}x} \right\}^{1/5} \cdot n^{-1/5}$$

$$h^{\mathrm{NW}}_{\mathrm{AMISE}}(n) = \left\{ \frac{R(K)\int \sigma^2(x)\,\mathrm{d}x}{\kappa_2(K)^2 4 \int \left[\frac{1}{2} m''(x) + \frac{f'(x)m'(x)}{f(x)}\right]^2 f(x)\,\mathrm{d}x} \right\}^{1/5} \cdot n^{-1/5}$$

Simulation: Extended results

n	Boundary region	Change in MISE to NW [%]	
		Local linear	NW boundary adjusted
50	[0, 0.119)	789.29	1578.07
100	[0, 0.104)	−41.90	1143.28
250	[0, 0.086)	−65.13	1982.67
500	[0, 0.075)	−72.98	14.92
1000	[0, 0.065)	−78.80	−63.95
2500	[0, 0.055)	−85.12	−77.80
5000	[0, 0.047)	−88.50	−83.76
10000	[0, 0.041)	−91.14	−88.26

Note: For $n = 10000$ the number of Monte-Carlo repetitions is 1000 instead of 10000.

Simulation: Robustness check h_{AMISE}

CV-optimal bandwidths for NW and LL compared to the asymptotically optimal bandwidths

n	$h_{\text{CV}}^{\text{NW}}$	$h_{\text{CV}}^{\text{LL}}$	h_{AMISE}	Change to h_{AMISE} [%]	
				Nadaraya-Watson	Local linear
50	0.097	0.128	0.119	−18.59	6.99
100	0.080	0.105	0.104	−22.96	1.25
250	0.062	0.085	0.086	−28.44	−1.55

Note: CV-optimal bandwidths are computed as a mean over 1000 repetitions.

Simulation: Robustness check MISE

n	Boundary region	Change in MISE to NW [%]	
		Local linear	NW boundary adjusted
Results for cross-validated bandwidths			
50	[0, 0.097)	703.01	3373.15
100	[0, 0.080)	−20.87	4561.98
250	[0, 0.062)	−41.54	443.39
Results for asymptotically optimal bandwidths			
50	[0, 0.119)	789.29	1578.07
100	[0, 0.104)	−41.90	1143.28
250	[0, 0.086)	−65.13	1982.67

Others: Cross Validation (CV)

- Leave One Out Cross Validation (LOOCV):

$$\text{CV}(h) \equiv \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{m}_{-i}(X_i, h))^2$$

$$\hat{h}_{\text{CV}} \equiv \arg\min_{h>0} \text{CV}(h)$$

- $\hat{h}_{\text{CV}}$ is essentially an unbiased estimator of h_{MISE}

Extensions

- Local polynomial regression (e.g. cubic)
- Multivariate regression
  - Boundary issues more severe
- Estimation of other functionals (e.g. derivatives)

Introductory literature

Monographs

- Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and its Applications*. Monographs on Statistics and Applied Probability 66. Boca Raton: Chapman & Hall/CRC.
- Wand, M. and M. Jones (1995). *Kernel Smoothing*. Monographs on Statistics and Applied Probability 60. Boca Raton: Chapman & Hall/CRC.

Papers

- Fan, J. (1992). “Design-adaptive nonparametric regression”. *Journal of the American Statistical Association* 87 (420), pp. 998–1004.
- Müller, H.-G. (1991). “Smooth optimum kernel estimators near endpoints”. *Biometrika* 78 (3), pp. 521–530.