Smoothed Nonparametric Derivative Estimation Using Weighted Difference Quotients

Paper Author: *Yu Liu* and *Kris De Brabanter*

Presented By **Yikun Zhang**

Department of Statistics,
University of Washington

May 5, 2023

Department of STATISTICS

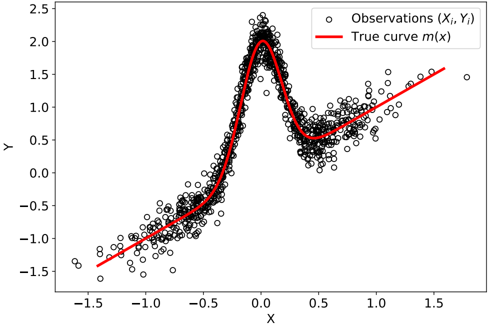
UNIVERSITY of WASHINGTON

Introduction

Data setting:

$$Y_i = m(X_i) + e_i, \quad \text{with} \quad X_i \in [a,b] \subset \mathbb{R} \quad \text{for} \quad i = 1, ..., n,$$

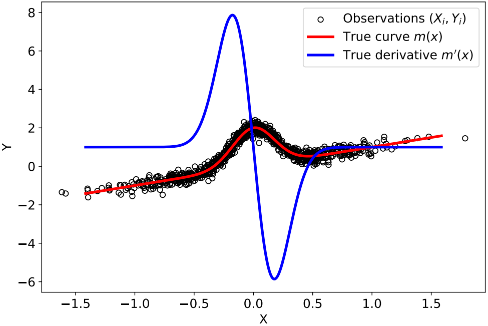
where e_i is independent of X_i and $\mathbb{E}(e_i) = 0$, $\mathrm{Var}(e_i) = \sigma_e^2 < \infty$.



Data setting:

$$Y_i = m(X_i) + e_i, \quad \text{with} \quad X_i \in [a,b] \subset \mathbb{R} \quad \text{for} \quad i = 1, \ldots, n,$$

where e_i is independent of X_i and $\mathbb{E}(e_i) = 0$, $\mathrm{Var}(e_i) = \sigma_e^2 < \infty$.



Question: How do we estimate $m^{(1)}(x) = \lim_{h \to 0} \frac{m(x+h) - m(x)}{h}$ from the data $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$?

Estimating $m^{(1)}(x)$ has significant impacts within and beyond **Statistics**:

- Explore the structures in curves (Chaudhuri and Marron, 1999) or the changing trend in time series (Rondonotti et al., 2007).

- Correct the bias term for a regression estimator in order to conduct valid inference (Eubank and Speckman, 1993; Calonico et al., 2018; Cheng and Chen, 2019).

Motivations of Derivative Estimation

Estimating $m^{(1)}(x)$ has significant impacts within and beyond **Statistics**:

- Explore the structures in curves (Chaudhuri and Marron, 1999) or the changing trend in time series (Rondonotti et al., 2007).
- Correct the bias term for a regression estimator in order to conduct valid inference (Eubank and Speckman, 1993; Calonico et al., 2018; Cheng and Chen, 2019).
- **Economics:** Quantify the relations between Marginal Propensity to Consume and other labor factors (Haavelmo, 1947).
- **Biomechanics:** Facilitate the kinematic analysis of human movements (Woltring, 1985).
- ...

Challenges of Derivative Estimation

Good news: The data $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$ from the model

$$Y = m(X) + e$$

are generally available in practice.

Challenges of Derivative Estimation

Good news: The data $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$ from the model

$$Y = m(X) + e$$

are generally available in practice.

Bad news: We don't have any data directly from the derivative (De Brabanter et al., 2013), *e.g.*, from the model

$$Y^{(1)} = m^{(1)}(X) + e'.$$

Challenge: We need to extract the derivative information from the original data $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$.

Existing Methods for Estimating the Derivatives

Parametric Methods: Assume $m(x)$ lying in some parametric family $\{f(x;\theta) : \theta \in \Theta\}$ and fit

$$\widehat{\theta} \in \arg\min_{\theta\in\Theta} \sum_{i=1}^{n} \left[Y_i - f(X_i;\theta)\right]^2 \quad \Longrightarrow \quad \widehat{m}^{(1)}(x) = f^{(1)}(x;\widehat{\theta}).$$

- *Drawback:* It is difficult to prosit a correct family $\{f(x;\theta) : \theta \in \Theta\}$.

Nonparametric Methods: Make no model assumptions on $m(x)$ and estimate $m^{(1)}(x)$ from the data $\mathcal{D}$.

- **Smoothing Splines (Zhou and Wolfe, 2000):** Estimate $m^{(q)}(x)$ for $q \geq 1$ through the derivative of smoothing splines (*i.e.*, piecewise polynomial curves).

- **Local Polynomial Regression (Fan and Gijbels, 1996):** It solves the weighted least-square problem at each query point x as: [1]

$$\begin{aligned}\widehat{\boldsymbol{\beta}}(x) &\equiv \left(\widehat{\beta}_0(x), ..., \widehat{\beta}_p(x)\right)^T \\ &= \underset{\boldsymbol{\beta}(x)\in\mathbb{R}^{p+1}}{\arg\min} \sum_{i=1}^{n}\left[Y_i - \sum_{j=0}^{p}\beta_j(x)\cdot(X_i - x)^j\right]^2 \bar{K}\left(\frac{X_i - x}{h}\right)\end{aligned}$$

and estimate the q-th order derivative $m^{(q)}(x)$ as $\widehat{m}^{(q)}(x) = q!\,\widehat{\beta}_q(x)$ for $q \leq p$.

[1] $\bar{K}$ is a symmetric kernel and $h > 0$ is the bandwidth parameter.

Nonparametric Methods: Difference Quotients

We order the data $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$ according to the increasing order of $X_i, i = 1, ..., n$:

$$Y_i = m(X_{(i)}) + e_i, \quad i = 1, ..., n.$$

The first-order difference quotients are defined as (Müller et al., 1987; Härdle, 1990):

$$\widehat{q}^{(1)}(X_{(i)}) = \frac{Y_i - Y_{i-1}}{X_{(i)} - X_{(i-1)}}, \quad i = 2, ..., n.$$

We order the data $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$ according to the increasing order of $X_i, i = 1, ..., n$:

$$Y_i = m(X_{(i)}) + e_i, \quad i = 1, ..., n.$$

The first-order difference quotients are defined as (Müller et al., 1987; Härdle, 1990):

$$\widehat{q}^{(1)}(X_{(i)}) = \frac{Y_i - Y_{i-1}}{X_{(i)} - X_{(i-1)}}, \quad i = 2, ..., n.$$

Drawback: The difference quotient $\widehat{q}^{(1)}(X_{(i)})$ estimates $m^{(1)}(X_{(i)})$ in the asymptotic rate:

$$\widehat{q}^{(1)}(X_{(i)}) - m^{(1)}(X_{(i)}) = O_P\left(n^2\right).$$

Nonparametric Methods: Weighted Difference Quotients

To reduce the variance, Iserles (2009); Charnigo et al. (2011) considered

$$\widehat{Y}_i^{(1)} \equiv \widehat{Y}_i^{(1)}(X_{(i)}) = \sum_{j=1}^{k} w_{i,j} \left(\frac{Y_{i+j} - Y_{i-j}}{X_{(i+j)} - X_{(i-j)}} \right)$$

for $k + 1 \leq i \leq n - k$ and $k \leq \frac{(n-1)}{2}$.

- The weights with $\sum_{j=1}^{k} w_{i,j} = 1$ are chosen to minimize the conditional variance $\mathrm{Var}\left(\widehat{Y}_i^{(1)} | X_{(1)}, ..., X_{(n)}\right)$.
- The asymptotic rate of convergence given $\left\{X_{(i)}\right\}_{i=1}^{n}$ becomes

$$\widehat{Y}_i^{(1)} - m^{(1)}(X_{(i)}) = \underbrace{O_P\left(\frac{k}{n}\right)}_{\text{Bias}} + \underbrace{O_P\left(\frac{n}{k^{\frac{3}{2}}}\right)}_{\sqrt{\text{Variance}}}.$$

To reduce the variance, Iserles (2009); Charnigo et al. (2011) considered

$$\widehat{Y}_i^{(1)} \equiv \widehat{Y}_i^{(1)}(X_{(i)}) = \sum_{j=1}^{k} w_{i,j} \left(\frac{Y_{i+j} - Y_{i-j}}{X_{(i+j)} - X_{(i-j)}} \right)$$

for $k+1 \leq i \leq n-k$ and $k \leq \frac{(n-1)}{2}$.

- The weights with $\sum_{j=1}^{k} w_{i,j} = 1$ are chosen to minimize the conditional variance $\mathrm{Var}\left(\widehat{Y}_i^{(1)} | X_{(1)}, ..., X_{(n)}\right)$.
- The asymptotic rate of convergence given $\{X_{(i)}\}_{i=1}^{n}$ becomes

$$\widehat{Y}_i^{(1)} - m^{(1)}(X_{(i)}) = \underbrace{O_P\left(\frac{k}{n}\right)}_{\text{Bias}} + \underbrace{O_P\left(\frac{n}{k^{\frac{3}{2}}}\right)}_{\sqrt{\text{Variance}}}.$$

Drawback: It only estimates $m^{(1)}(x)$ at $x = X_{(i)}$ for $k+1 \leq i \leq n-k$.

Contributions of This Paper

De Brabanter et al. (2013) proposed using local polynomial regression to smoothing out the noisy derivative estimates $\widehat{Y}_i^{(1)}, i = k+1, ..., n-k$.

De Brabanter et al. (2013) proposed using local polynomial regression to smoothing out the noisy derivative estimates $\widehat{Y}_i^{(1)}, i = k+1, ..., n-k$.

Drawback: Their method only works for the equispaced design, *i.e.*,

$$X_{(i)} = a + \frac{i \cdot (b-a)}{n-1}, \quad i = 1, ..., n.$$

De Brabanter et al. (2013) proposed using local polynomial regression to smoothing out the noisy derivative estimates $\widehat{Y}_i^{(1)}, i = k+1, ..., n-k$.

Drawback: Their method only works for the equispaced design, *i.e.*,

$$X_{(i)} = a + \frac{i \cdot (b-a)}{n-1}, \quad i = 1, ..., n.$$

Main contribution: In this paper (Liu and De Brabanter, 2020), the authors will extend the above framework to the random design.

Methodology

Probability Integral Transform to Uniform$[0,1]$

Recall that our i.i.d. data $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$ are generated from the model

$$Y = m(X) + e,$$

where X has unknown density f and CDF F.

Fact: $F(X_i) \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}[0,1]$ for $i = 1, ..., n$ (Casella and Berger, 2002).

Insights:

Probability Integral Transform to Uniform$[0, 1]$

Recall that our i.i.d. data $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$ are generated from the model

$$Y = m(X) + e,$$

where X has unknown density f and CDF F.

Fact: $F(X_i) \overset{\text{i.i.d.}}{\sim} \text{Uniform}[0, 1]$ for $i = 1, ..., n$ (Casella and Berger, 2002).

Insights:

- Estimate derivatives of the transformed function $r(U) = m(F^{-1}(U))$.

Probability Integral Transform to Uniform$[0, 1]$

Recall that our i.i.d. data $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$ are generated from the model

$$Y = m(X) + e,$$

where X has unknown density f and CDF F.

Fact: $F(X_i) \overset{\text{i.i.d.}}{\sim} \text{Uniform}[0, 1]$ for $i = 1, ..., n$ (Casella and Berger, 2002).

Insights:

- Estimate derivatives of the transformed function $r(U) = m(F^{-1}(U))$.
- Refer back to the derivatives of $m(X)$ by the chain rule:

$$\begin{aligned} m^{(1)}(X) &= f(X) \cdot r^{(1)}(U), \\ m^{(2)}(X) &= f^{(1)}(X) \cdot r^{(1)}(U) + [f(X)]^2 \, r^{(2)}(U), ... \end{aligned}$$

Probability Integral Transform to Uniform$[0, 1]$

Recall that our i.i.d. data $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$ are generated from the model

$$Y = m(X) + e,$$

where X has unknown density f and CDF F.

Fact: $F(X_i) \overset{\text{i.i.d.}}{\sim} \text{Uniform}[0, 1]$ for $i = 1, ..., n$ (Casella and Berger, 2002).

Insights:

- Estimate derivatives of the transformed function $r(U) = m(F^{-1}(U))$.
- Refer back to the derivatives of $m(X)$ by the chain rule:

$$\begin{aligned} m^{(1)}(X) &= f(X) \cdot r^{(1)}(U), \\ m^{(2)}(X) &= f^{(1)}(X) \cdot r^{(1)}(U) + [f(X)]^2 r^{(2)}(U), ... \end{aligned}$$

- Practically, f and F can be estimated by the kernel density estimator (KDE; Chen 2017).

Data Setting: Consider the ordered data $\{(U_{(i)}, Y_i)\}_{i=1}^n$ from the model:

$$Y_i = r(U_{(i)}) + e_i, \quad i = 1, ..., n,$$

where $U_{(1)} \leq \cdots \leq U_{(n)}$ are order statistics from Uniform$[0, 1]$.

First-Order Noisy Derivative Estimator

Data Setting: Consider the ordered data $\{(U_{(i)}, Y_i)\}_{i=1}^n$ from the model:

$$Y_i = r(U_{(i)}) + e_i, \quad i = 1, ..., n,$$

where $U_{(1)} \leq \cdots \leq U_{(n)}$ are order statistics from Uniform$[0, 1]$.

First-order noisy derivative estimator at $u = U_{(i)}$:

$$\widehat{Y}_i^{(1)} = \sum_{j=1}^{k} w_{i,j} \left(\frac{Y_{i+j} - Y_{i-j}}{U_{(i+j)} - U_{(i-j)}} \right) \quad \text{for} \quad k+1 \leq i \leq n-k,$$

where k is a tuning parameter.

- The weights are chosen to minimize $\mathrm{Var}\left(\widehat{Y}_i^{(1)} | U_{(1)}, ..., U_{(n)}\right)$.
- The asymptotic rate of convergence of $\widehat{Y}_i^{(1)}$ given $\{U_{(i)}\}_{i=1}^n$ is

$$\widehat{Y}_i^{(1)} - r^{(1)}(U_{(i)}) = O_P\left(\frac{k}{n}\right) + O_P\left(\frac{n}{k^{\frac{3}{2}}}\right).$$

Second-Order Noisy Derivative Estimator

Second-order noisy derivative estimator at $u = U_{(i)}$:

$$\widehat{Y}_i^{(2)} = 2\sum_{j=1}^{k_2} w_{ij,2} \cdot \frac{\left(\frac{Y_{i+j+k_1}-Y_{i+j}}{U_{(i+j+k_1)}-U_{(i+j)}} - \frac{Y_{i-j-k_1}-Y_{i-j}}{U_{(i-j-k_1)}-U_{(i-j)}}\right)}{U_{(i+j+k_1)} + U_{(i+j)} - U_{(i-j-k_1)} - U_{(i-j)}},$$

for $k_1 + k_2 + 1 \leq i \leq n - k_1 - k_2$, where k_1, k_2 are tuning parameters.

- The weights $w_{ij,2}$ are chosen to minimize the asymptotic leading order of $\text{Var}\left(\widehat{Y}_i^{(2)} | U_{(1)}, ..., U_{(n)}\right)$.
- The asymptotic rate of convergence of $\widehat{Y}_i^{(2)}$ given $\{U_{(i)}\}_{i=1}^n$ is

$$\widehat{Y}_i^{(2)} - r^{(2)}(U_{(i)}) = O_P\left(\frac{k}{n}\right) + O_P\left(\frac{n^2}{k^{\frac{5}{2}}}\right)$$

  when $k_1, k_2 \asymp k$.

Drawbacks of the proposed noisy derivative estimators $\widehat{Y}_i^{(1)}$ and $\widehat{Y}_i^{(2)}$:

1. They are only defined at the (interior) design points $U_{(i)}$ for $k+1 \leq i \leq n-k$.
2. They contain noises from the unknown error $e_i, i = 1, ..., n$.

Drawbacks of the proposed noisy derivative estimators $\widehat{Y}_i^{(1)}$ and $\widehat{Y}_i^{(2)}$:

1. They are only defined at the (interior) design points $U_{(i)}$ for $k+1 \leq i \leq n-k$.

2. They contain noises from the unknown error $e_i, i = 1, ..., n$.

Solution: Apply the local polynomial regression to smoothing out these noisy derivative estimators.

Smoothing the Noisy Derivative Estimators

Take the first-order derivative data $\{(U_{(i)}, \widehat{Y}_i^{(1)})\}_{i=k+1}^{n-k}$ as an example.

At any point $u_0 \in [0,1]$, the solution of the local polynomial regression is

$$\widehat{r}^{(1)}(u_0) = \boldsymbol{\epsilon}_1^T \widehat{\boldsymbol{\beta}}(u_0) = \boldsymbol{\epsilon}_1^T \left(\boldsymbol{U}_u^T \boldsymbol{W}_u \boldsymbol{U}_u\right)^{-1} \boldsymbol{U}_u^T \boldsymbol{W}_u \widehat{\boldsymbol{Y}}^{(1)},$$

where $\boldsymbol{\epsilon}_1 = (1, 0, ..., 0)^T \in \mathbb{R}^{p+1}$, $\widehat{\boldsymbol{Y}}^{(1)} = \left(\widehat{Y}_{k+1}^{(1)}, ..., \widehat{Y}_{n-k}^{(1)}\right)^T \in \mathbb{R}^{n-2k}$, and

$$\boldsymbol{U}_u = \begin{pmatrix} 1 & \left(U_{(k+1)} - u_0\right) & \cdots & \left(U_{(k+1)} - u_0\right)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \left(U_{(n-k)} - u_0\right) & \cdots & \left(U_{(n-k)} - u_0\right)^p \end{pmatrix},$$

$$\boldsymbol{W}_u = \begin{pmatrix} \bar{K}\left(\frac{U_{(k+1)} - u_0}{h}\right) & & \\ & \ddots & \\ & & \bar{K}\left(\frac{U_{(n-k)} - u_0}{h}\right) \end{pmatrix}.$$

Smoothing the Noisy Derivative Estimators

Caveat: $\left\{\widehat{Y}_i^{(1)}\right\}_{i=k+1}^{n-k}$ are no longer independent even when we condition on $\{U_{(i)}\}_{i=1}^n$. Equivalently,

$$\widehat{Y}_i^{(1)} = r^{(1)}(U_{(i)}) + \widetilde{e}_i, \quad i = 1, ..., n.$$

Smoothing the Noisy Derivative Estimators

Caveat: $\left\{\widehat{Y}_i^{(1)}\right\}_{i=k+1}^{n-k}$ are no longer independent even when we condition on $\{U_{(i)}\}_{i=1}^{n}$. Equivalently,

$$\widehat{Y}_i^{(1)} = r^{(1)}(U_{(i)}) + \widetilde{e}_i, \quad i = 1, ..., n.$$

Solution: Use a bimodal kernel $\bar{K}$ with $\bar{K}(0) = 0$ in the local polynomial regression to tackle the correlated errors (De Brabanter et al., 2013).

- Gaussian bimodal kernel: $\bar{K}(u) = \frac{2u^2}{\sqrt{\pi}} \exp(-u^2)$.

Caveat: $\left\{\widehat{Y}_i^{(1)}\right\}_{i=k+1}^{n-k}$ are no longer independent even when we condition on $\{U_{(i)}\}_{i=1}^{n}$. Equivalently,

$$\widehat{Y}_i^{(1)} = r^{(1)}(U_{(i)}) + \widetilde{e}_i, \quad i = 1, ..., n.$$

Solution: Use a bimodal kernel $\bar{K}$ with $\bar{K}(0) = 0$ in the local polynomial regression to tackle the correlated errors (De Brabanter et al., 2013).

- Gaussian bimodal kernel: $\bar{K}(u) = \frac{2u^2}{\sqrt{\pi}} \exp(-u^2)$.
- Bandwidth h is selected by minimizing $\frac{1}{n-2k} \sum_{i=k+1}^{n-k} \left(\widehat{r}^{(1)}(U_{(i)}) - \widehat{Y}_i^{(1)}\right)^2$ with a correction for the bimodal kernel.

Smoothing the Noisy Derivative Estimators

Caveat: $\left\{\widehat{Y}_i^{(1)}\right\}_{i=k+1}^{n-k}$ are no longer independent even when we condition on $\{U_{(i)}\}_{i=1}^n$. Equivalently,

$$\widehat{Y}_i^{(1)} = r^{(1)}(U_{(i)}) + \widetilde{e}_i, \quad i = 1, ..., n.$$

Solution: Use a bimodal kernel $\bar{K}$ with $\bar{K}(0) = 0$ in the local polynomial regression to tackle the correlated errors (De Brabanter et al., 2013).

- Gaussian bimodal kernel: $\bar{K}(u) = \frac{2u^2}{\sqrt{\pi}} \exp(-u^2)$.
- Bandwidth h is selected by minimizing $\frac{1}{n-2k} \sum_{i=k+1}^{n-k} \left(\widehat{r}^{(1)}(U_{(i)}) - \widehat{Y}_i^{(1)}\right)^2$ with a correction for the bimodal kernel.
- The asymptotic rate of convergence of $\widehat{r}^{(1)}(u_0)$ given $\{U_{(i)}\}_{i=1}^n$ is

$$\widehat{r}^{(1)}(u_0) - r^{(1)}(u_0) = O_P\left(h^{p+1}\right) + O_P\left(\frac{k}{n}\right) + O_P\left(\frac{n}{\sqrt{k^3(n-2k)h}}\right).$$

Observed data $\{(X_i, Y_i)\}_{i=1}^n$ from $Y_i = m(X_i) + e_i$

Estimate f and F via KDE as $\widehat{f}_v$ and $\widehat{F}_v$

Weighted difference quotients $Y_i^{(1)} = \sum_{j=1}^{k} w_{ij} \left(\frac{Y_{i+j} - Y_{i-j}}{U_{(i+j)} - U_{(i-j)}} \right)$ with $U_{(i)} = \widehat{F}_v(X_{(i)})$

Smooth out $\left(U_{(i)}, Y_i^{(1)}\right)$ via the local polynomial regression as $\widehat{r}^{(1)}(U)$

Transform back to X as $\widehat{m}^{(1)}(X) = \widehat{f}_v(X) \cdot \widehat{r}^{(1)}(U)$

Figure 1: Summary of the proposed derivative estimation framework in the paper (Liu and De Brabanter, 2020).

Thank you!

Reference I

S. Calonico, M. D. Cattaneo, and M. H. Farrell. On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, 113(522):767–779, 2018.

G. Casella and R. Berger. *Statistical Inference*. Duxbury advanced series. Thomson Learning, 2nd ed. edition, 2002.

R. Charnigo, B. Hall, and C. Srinivasan. A generalized c_p criterion for derivative estimation. *Technometrics*, 53(3):238–253, 2011.

P. Chaudhuri and J. S. Marron. Sizer for exploration of structures in curves. *Journal of the American Statistical Association*, 94(447):807–823, 1999.

Y.-C. Chen. A tutorial on kernel density estimation and recent advances. *Biostatistics & Epidemiology*, 1 (1):161–187, 2017.

G. Cheng and Y.-C. Chen. Nonparametric inference via bootstrapping the debiased estimator. *Electronic Journal of Statistics*, 13(1):2194 – 2256, 2019.

K. De Brabanter, J. De Brabanter, I. Gijbels, and B. De Moor. Derivative estimation with local polynomial fitting. *Journal of Machine Learning Research*, 14(1):281–301, 2013.

R. L. Eubank and P. L. Speckman. Confidence bands in nonparametric regression. *Journal of the American Statistical Association*, 88(424):1287–1301, 1993.

J. Fan and I. Gijbels. *Local polynomial modelling and its applications*, volume 66. Chapman & Hall/CRC, 1996.

T. Haavelmo. Methods of measuring the marginal propensity to consume. *Journal of the American Statistical Association*, 42(237):105–122, 1947.

W. Härdle. *Applied nonparametric regression*. Number 19. Cambridge university press, 1990.

A. Iserles. *A first course in the numerical analysis of differential equations*. Number 44. Cambridge university press, 2009.

Reference II

Y. Liu and K. De Brabanter. Smoothed nonparametric derivative estimation using weighted difference quotients. *Journal of Machine Learning Research*, 21(1):2438–2482, 2020.

H.-G. Müller, U. Stadtmüller, and T. Schmitt. Bandwidth choice and confidence intervals for derivatives of noisy data. *Biometrika*, 74(4):743–749, 1987.

V. Rondonotti, J. S. Marron, and C. Park. SiZer for time series: A new approach to the analysis of trends. *Electronic Journal of Statistics*, 1(none):268 – 289, 2007.

H. J. Woltring. On optimal smoothing and derivative estimation from noisy displacement data in biomechanics. *Human Movement Science*, 4(3):229–245, 1985.

S. Zhou and D. A. Wolfe. On derivative estimation in spline regression. *Statistica Sinica*, pages 93–108, 2000.

Recall that the proposed first-order noisy derivative estimator

$$\widehat{Y}_i^{(1)} = \sum_{j=1}^{k} w_{i,j} \left(\frac{Y_{i+j} - Y_{i-j}}{U_{(i+j)} - U_{(i-j)}} \right)$$

is only defined at $U_{(i)}$ for $k+1 \leq i \leq n-k$.

Issue: There are not enough pairs of observations within the left and right boundary regions $2 \leq i \leq k$ and $n-k+1 \leq i \leq n-1$.

Naive Solution:

$$\widehat{Y}_i^{(1)} = \sum_{j=1}^{k(i)} w_{i,j} \left(\frac{Y_{i+j} - Y_{i-j}}{U_{(i+j)} - U_{(i-j)}} \right),$$

where $k(i) = i-1$ for the left boundary and $k(i) = n-i$ for the right boundary.

Recall that the proposed first-order noisy derivative estimator

$$\widehat{Y}_i^{(1)} = \sum_{j=1}^{k} w_{i,j} \left(\frac{Y_{i+j} - Y_{i-j}}{U_{(i+j)} - U_{(i-j)}} \right)$$

is only defined at $U_{(i)}$ for $k+1 \leq i \leq n-k$.

Issue: There are not enough pairs of observations within the left and right boundary regions $2 \leq i \leq k$ and $n-k+1 \leq i \leq n-1$.

Proposed boundary correction:

$$\widehat{Y}_i^{(1)} = \sum_{j=1}^{k(i)} w_{i,j} \left(\frac{Y_{i+j} - Y_{i-j}}{U_{(i+j)} - U_{(i-j)}} \right) + \sum_{j=k(i)+1}^{k} w_{i,j} \left[\left(\frac{Y_{i+j} - Y_{i}}{U_{(i+j)} - U_{(i)}} \right) \mathbb{1}_{\{2 \leq i \leq k\}} + \left(\frac{Y_{i} - Y_{i-j}}{U_{(i)} - U_{(i-j)}} \right) \mathbb{1}_{\{n-k < i < n\}} \right],$$

where $k(i) = i-1$ for the left boundary and $k(i) = n-i$ for the right boundary.