Hassan Doosti *Editor*

Flexible Nonparametric Curve Estimation

Springer

Flexible Nonparametric Curve Estimation

Hassan Doosti
Editor

Flexible Nonparametric Curve Estimation

Springer

Editor
Hassan Doosti
School of Mathematical and Physical Sciences
Macquarie University
Sydney, NSW, Australia

ISBN 978-3-031-66500-4 ISBN 978-3-031-66501-1 (eBook)
https://doi.org/10.1007/978-3-031-66501-1

Mathematics Subject Classification: 62G08, 62G09, 62E20, 62G20, 62G30, 62E17, 62G05, 62G07, 62N02

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

In the realm of statistical inference, we often find ourselves navigating the delicate balance between assumptions and reality. While classical statistics typically rests upon the premise of known population distributions with unknown parameters, the truth of many situations diverges sharply from these assumptions. Take, for instance, the ubiquitous standard regression models where Gaussian distribution of error terms is presumed. Yet, there are myriad scenarios where such assumptions fail to hold, casting doubts on the validity of our inferences.

It is precisely in these challenging contexts that the realm of nonparametric statistics emerges as a beacon of possibility. Rather than tethering ourselves to predefined distributional forms, nonparametric statistics bravely ventures into the terrain where the true distribution of the population remains enigmatic. It is within this landscape that our edited volume, "Flexible Nonparametric Curve Estimation," finds its purpose.

The heart of this book lies in the endeavor to estimate crucial functions or curves when the distributions of populations evade our grasp. These curves find relevance across a spectrum of disciplines, from medicine to engineering, and economics to environmental sciences. Each chapter within this compendium is dedicated to presenting a novel nonparametric estimation approach for a specific function, coupled with an exploration of its significance in various contexts.

Organized into 12 chapters, this volume embarks on a journey through a spectrum of methodologies and applications. From the revisiting of convolution processes to the exploration of Bayesian nonparametrics, each chapter offers a unique lens through which to view the complexities of real-world data analysis.

Each chapter is meticulously crafted to not only propose innovative estimation techniques but also to delve into the theoretical underpinnings and practical implications of the methodologies. Through numerical studies and inclusion of code, the authors endeavor to empower readers to replicate and extend their findings, fostering a spirit of exploration and advancement.

This volume is intended to serve as a valuable resource for researchers immersed in the multifaceted world of data analysis across diverse domains. Whether delving into medical sciences, econometrics, environmental sciences, or beyond, the methodologies presented herein offer a flexible toolkit for navigating the complexities of real-world data. Furthermore, this book holds promise as a companion text for postgraduate students in statistical fields, offering both theoretical insights and practical guidance in the realm of nonparametric curve estimation.

We extend our gratitude to the contributing authors for their dedication and expertise, which have enriched this volume immeasurably. It is our hope that this book will inspire further exploration, innovation, and collaboration in the ever-evolving landscape of statistical inference.

Sydney, NSW, Australia Hassan Doosti

Contents

Tilted Nonparametric Regression Function Estimation

Farzaneh Boroumand, Mohammad T. Shakeri, Nino Kordzakhia, Mahdi Salehi, and Hassan Doosti

Abstract This paper provides the theory about the convergence rate of the tilted version of linear smoother. We study tilted linear smoother, a class of nonparametric regression function estimators, which is obtained by minimizing the distance to an infinite order flat-top trapezoidal kernel estimator. We prove that the proposed estimator achieves a high level of accuracy. Moreover, it preserves the attractive

The first version of this chapter was initially published on the preprint server arXiv, and it can be accessed at https://arxiv.org/abs/2102.02381

F. Boroumand (✉)
School of Mathematical and Physical Sciences, Macquarie University, Macquarie Park, NSW, Australia

Department of Biostatistics, Health School, Mashhad University of Medical Sciences, Mashhad, Iran

Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Camperdown, NSW, Australia
e-mail: Farzaneh.Boroumand@mq.edu.au

M. T. Shakeri
Department of Biostatistics, Health School, Mashhad University of Medical Sciences, Mashhad, Iran
e-mail: shakerimt@mums.ac.ir

N. Kordzakhia
School of Mathematical and Physical Sciences, Macquarie University, Macquarie Park, NSW, Australia
e-mail: Nino.Kordzakhia@mq.edu.au

M. Salehi
Department of Mathematics and Statistics, University of Neyshabur, Neyshabur, Iran

Department of Statistics, University of Pretoria, Pretoria, South Africa
e-mail: salehi.sms@neyshabur.ac.ir

H. Doosti
School of Mathematical and Physical Sciences, Macquarie University, Sydney, Australia
e-mail: hassan.doosti@mq.edu.au

H. Doosti (ed.), *Flexible Nonparametric Curve Estimation*,
https://doi.org/10.1007/978-3-031-66501-1_1

properties of the infinite order flat-top kernel estimator. We also present an extensive numerical study for analysing the performance of two members of the tilted linear smoother class named tilted Nadaraya-Watson and tilted local linear for finite samples. The simulation study shows that tilted Nadaraya-Watson and tilted local linear perform better than their classical analogs, under some specified conditions, in terms of Median Integrated Squared Error (MISE). Next, the performance of these estimators as well as the conventional estimators are illustrated by curve fitting to COVID-19 data for 12 countries and a dose-response data set. Finally, the R codes for obtaining various regression estimators mentioned above are given as an appendix.

Keywords Tilted estimators · Nonparametric regression function estimation · Rate of convergence · Infinite order flat top kernels · COVID-19 curve fitting

1 Introduction

Let the regression model be

$$Y_i = r(X_i) + \epsilon_i, \quad 1 \leq i \leq n, \tag{1}$$

where $(Y_1, X_1), (Y_2, X_2), \ldots, (Y_n, X_n)$, are the data pairs, the design variable $X \sim f_X$, X and ϵ are independent, ϵ_i's are independent and identically distributed (iid) errors with zero mean $E(\epsilon) = 0$ and variance $E(\epsilon^2) = \sigma^2$. The regression function r and f_X are unknown. In this paper, we will focus on a nonparametric approach to estimate r. The main subject of this study is a class of nonparametric estimators called linear smoother. An estimator $\check{r}$ of r, is said to be a linear smoother if it can be written in a form of linear function of weighted Y sample. Let the weight-vector be

$$l(x) = (l_1(x), \ldots, l_n(x))^T.$$

Then the linear smoother $\check{r}$ can be written as

$$\check{r}_n(x) = l(x)^T Y = \sum\nolimits_{i=1}^{n} l_i(x) Y_i, \tag{2}$$

where $\sum_{i=1}^{n} l_i(x) = 1$, see Buja et al. (1989). Nadaraya-Watson estimator and local linear estimator are two prevailing members of this class of estimators. The weight functions for Nadaraya-Watson smoother, see Nadaraya (1964), Watson (1964), are

$$l_{i,NW}(x) = \frac{K(\frac{X_i - x}{h})}{\sum_{j=1}^{n} K(\frac{X_j - x}{h})}, \quad i = 1, \ldots, n. \tag{3}$$

For the standard local linear smoother the weight functions are defined as follows,

$$l_{i,ll}(x) = \frac{b_i(x)}{\sum_{j=1}^{n} b_j(x)}, \; i = 1, \ldots, n, \tag{4}$$

$$b_i(x) = K(\frac{X_i - x}{h})(S_{n,2}(x) - (X_i - x)S_{n,1}(x)), \; i = 1, \ldots, n,$$

$$S_{n,j}(x) = \sum_{i=1}^{n} K(\frac{X_i - x}{h})(X_i - x)^j, \; j = 1, 2.$$

where K is a kernel function. The kernel function depends on the bandwidth, or smoothing, parameter h and assigns weights to the observations according to the distance to the target point x. The small values of h cause the neighboring points of x to have the larger influence on the estimate leading to curvature changes in the estimated curve. The larger values of h imply that the distanced data points will have the same effect as the neighboring points on the local fit, resulting in a smoother estimate. Thus finding an optimal h is the essential task in the estimation procedure, see Wasserman (2006). One of the ways finding the optimal h is by minimising the leave-one-out cross validation score function, (Wasserman, 2006). The leave-one-out cross validation score is defined by

$$CV = \hat{R}(h) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \check{r}_{(-i)}(X_i))^2, \tag{5}$$

where $\check{r}_{(-i)}(X_i)$ is obtained from (2) by omitting the ith pair (X_i, Y_i). In this work, we will present the tilted versions of linear smoother. A tilting technique applied to an empirical distribution, leads to replacing $1/n$ data weights from uniform distribution by $p_i, \; 1 \le i \le n$, from general multinomial distribution over data. Hall and Yao (2003) studied asymptotic properties of the tilted regression estimator with autoregressive errors using generalized empirical likelihood method, which typically involves solving a non-linear and high dimensional optimization problem. Grenander (1956) introduced a tilted method to impose restrictions on the density estimates. There are two approaches to estimating of the tilting parameters: Empirical likelihood and Distance Measure based approaches. The empirical likelihood-based method is a semi-parametric method which provides a convenience of adding a parametric model through estimating equations. Owen (1988) proposed an empirical likelihood to be used as an alternative to the likelihood ratio tests, and derived its asymptotic distribution. Chen (1997), Zhang (1998), Schick and Wefelmeyer (2009), Müller et al. (2005) further developed the empirical likelihood-based method for estimating the tilting parameters. Chen (1997) applied the empirical likelihood method to estimate the tilting parameters $p_i, \; 1 \le i \le n$, under the constraints on the shape of distribution. In his kernel-based estimator, n^{-1} was replaced by the weights obtained from the empirical likelihood method.

In Chen (1997) it was proved that the proposed estimator has a smaller variance than the conventional kernel estimators. Schick and Wefelmeyer (2009), also used the similar approach obtaining the consistent tilted estimator with higher efficiency than that of conventional estimators in the autoregression framework. In contrast in the Distance Measure approach, the tilted estimators are defined by minimizing distances, conditional to various types of constraints. Hall and Presnell (1999), Hall and Huang (2001), Carroll et al. (2011), Doosti and Hall (2016), Doosti et al. (2018) used the setup-specific Distance Measure approaches for estimating the tilting parameters. Carroll et al. (2011), proposed a new approach for density function estimation, and regression function estimation as well as hypothesis testing under shape constraints in the model with measurement errors. A tilting method used in Carroll et al. (2011) led to curve estimators under some constraints. Doosti and Hall (2016) introduced a new higher order nonparametric density estimator, using tilting method, where they used L_2-metric between the proposed estimator and a consistent 'Sinc' kernel based estimator. Doosti et al. (2018), have introduced a new way of choosing the bandwidth and estimating the tilted parameters based on the cross-validation function. In Doosti et al. (2018), it was shown that the proposed density function estimator had improved efficiency and was more cost-effective than the conventional kernel-based estimators studied in this paper.

In this work, we propose a new tilted version of a linear smoother which is obtained by minimising the distance to a comparator estimator. The comparator estimator is selected to be an infinite order flat-top kernel estimator. This class of estimators is characterized by a Fourier transform, which is flat near the origin and infinitely differentiable elsewhere, see McMurry and Politis (2008). We prove that the tilted estimators achieve a high level of accuracy, yet preserving the attractive properties of an infinite-order flat-top kernel estimator.

The rest of this paper contains the additional four sections and Appendix. In the Sect. 2, we provide the notation, definitions and preliminary results. The Sect. 2 also includes the definition of an infinite-order estimator, as a comparator estimator. Section 3 contains the main results formulated in Theorems 1–3. We present a simulation study in the Sect. 4. The real data applications are provided in the Sect. 5. The proof of the main theorem is accommodated in the Appendix.

2 Notation and Preliminary Results

Definition 1 A general infinite order flat-top kernel K is defined

$$K(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lambda(s) e^{-isx} ds, \tag{6}$$

where $\lambda(s)$ is the Fourier transform of kernel K, and $c > 0$ is a fixed constant.

$$\lambda(s) = \begin{cases} 1, & | s | \leq c \\ g(| s |), & | s | > c \end{cases},$$

and g is not unique and it should be chosen to make $\lambda(s)$, $\lambda^2(s)$, and $s\lambda(s)$ integrable (McMurry & Politis, 2004).

2.1 Infinite Order Flat-Top Kernel Regression Estimator

McMurry and Politis (2004) proposed Infinite order flat-top kernel regression estimator as follows; we used this estimator as the comparator estimator in Sect. 3,

$$\check{r} = \sum_{i=1}^{n} \frac{K(\frac{X_i - x}{h})}{\sum_{j=1}^{n} K(\frac{X_j - x}{h})} Y_i, \tag{7}$$

where K is an infinite order flat-top kernel from (6) such as trapezoidal kernel with the following definition

$$K(x) = \frac{2(cos(x/2) - cos(x))}{\pi x^2},$$

this K satisfying Definition 1 since the Fourier transform of $K(x)$ is

$$\lambda(s) = \begin{cases} 1 & | s | \leq 1/2, \\ 2(1 - | s |) & 1/2 < | s | \leq 1, \\ 0 & | s | > 1. \end{cases}$$

2.2 Tilted Linear Smoother

We define tilted linear smoother as follows

$$\hat{r}_n(x|h, p) = \sum_{i=1}^{n} n p_i l_i(x_i) Y_i, \tag{8}$$

where p_i's are tilting parameters, $p_i \geq 0$ and $\sum_i^n p_i = 1$. The bandwidth parameter h and the vector of tilting parameters $p = (p_1, \cdots, p_n)$, are to be estimated. In Sect. 4, we evaluate the performance of two member of this class of estimators including tilted versions of Nadaraya-Watson (2) and standard local linear estimators (3) in finite samples.

3 Main Results

Let $\hat{r}_n(.|\theta)$ be the tilted linear smoother from (8) for the regression function r, where $\theta = (h, p)$ is a vector of unknown parameters. Further $\check{r}$ from (7) will be used as a comparator estimator of r, $\check{r}$ can be any estimator with an optimal convergence rate (McMurry & Politis, 2008). We will estimate θ by minimising the L_2-distance between $\hat{r}_n(.|\theta)$ and $\check{r}$ preserving the convergence rate of $\check{r}$, provided the following assumptions hold

(a) $\|\check{r} - r\| = O_p(\delta_n)$
(b) There exists $\tilde{\theta}$ such that $\hat{r}_n(.|\tilde{\theta})$ and $\check{r}$ possess the same convergence rates, i.e. $\|\hat{r}_n(.|\tilde{\theta}) - r\| = O_p(\delta_n)$,

where $\delta_n \geq 0$ converges to 0 as n tends to ∞, e.g. $\delta_n = n^{-c}$ for some $c \in (0, 1/2)$. A further discussion on the assumptions (a)–(b) can be found in Doosti and Hall (2016).

We define $\hat{\theta}$ as the solution to the optimisation problem as

$$\hat{\theta} = \arg\min_{\theta} \|\hat{r}_n(.|\theta) - \check{r}\| \tag{9}$$

subject to the constraints for the bandwidth parameter $h > 0$ and vector p introduced in Sect. 2.2.

In Theorem 1, we show that the convergence rate of $\hat{r}_n(.|\hat{\theta})$ and $\check{r}$ is $O_p(\delta_n)$.

Theorem 1 *If the assumptions (a)–(b) hold then for any $\hat{\theta}$ which fulfills (9) we have*

$$\|\hat{r}_n(.|\hat{\theta}) - r\| = O_p(\delta_n).$$

Proof Due to assumption (a), there exists $\tilde{\theta}$ such that

$$\|\hat{r}_n(.|\tilde{\theta}) - \check{r}\| \leq \|\hat{r}_n(.|\tilde{\theta}) - r\| + \|r - \check{r}\| = O_p(\delta_n),$$

in which the first equation is a result of the triangle inequality, and specifically from the fact that

$$\|r - \check{r}\| = O_p(\delta_n); \tag{10}$$

see Assumption (a). If $\tilde{\theta}$ is as in assumption (b) then

$$\|\hat{r}_n(.|\hat{\theta}) - \check{r}\| \leq \|\hat{r}_n(.|\tilde{\theta}) - \check{r}\| = O_p(\delta_n). \tag{11}$$

Together, results (10) and (11) imply Theorem 1.

Theorem 1 implies that the convergence rate of $\hat{r}_n(.|\hat{\theta})$ estimator coincides with that of $\check{r}$ with the bandwidth parameter h replaced by its 'plug-in' type estimate similar to that from McMurry and Politis (2004, 2008).

The regression function $r \in \mathcal{C}$, where $\mathcal{C}$ is a class of regression functions, if

$$\lim_{C\to\infty} \limsup_{n\to\infty} \sup_{r\in\mathcal{C}} [P\{\|\hat{r}_n(.|\tilde{\theta}) - r\| \geq C\delta_n\} + P\{\|\check{r} - r\| \geq C\delta_n\}] = 0, \tag{12}$$

subject to existence of $\tilde{\theta}$.

Theorem 2 *If* (12) *holds for regression functions from* $\mathcal{C}$ *then*

$$\lim_{C\to\infty} \limsup_{n\to\infty} \sup_{r\in\mathcal{C}} P\{\|\hat{r}_n(.|\hat{\theta}) - r\| \geq C\delta_n\} = 0. \tag{13}$$

Theorem 2 states that $\hat{r}_n(.|\hat{\theta})$ and $\check{r}$ converge to r uniformly in $\mathcal{C}$.

Let $X_1, X_2, \ldots, X_n$ be iid random variables with probability density function (pdf) $f(x)$ and $\hat{g}(x)$ be its kernel based density function estimator

$$\hat{g}(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x - X_i}{h})$$

and $g(x) = E_f \hat{g}(x)$.

Suppose that (c)–(d) hold for ϕ_K and ϕ_q, Fourier transforms for K and $q = r \cdot g$, respectively,

(c) $\phi_K(t)^{-1} = 1 + \sum_{j=1}^{k} c_j t^{2j}$, $c_1, \ldots, c_k$ are real numbers;
(d) $\int |\phi_q(t)||t|^{2k} dt < \infty$;
(e) For constants $C_1, \ldots, C_5 > 0$, and $j = 1, \ldots, k$ the derivatives $q^{(2j)}$ exist, $|q^{(2j)}(x)| \leq C_1$ and $\int |q^{(2j)}(x)| \leq C_1$, and either
  (1) $|q^{(2j)}(x)/r \cdot f(x)| \leq C_1$ for all x, or
  (2) $|q^{(2j)}(x)/r \cdot f(x)| \leq C_1(1 + |x|)^{C_2}$ for all x and $P(|X| \geq x) \leq C_3 \exp(-C_4 x^{C_5}),\ x > 0$.
(f) Under assumption (e)–(1), $\delta_n \to 0$ as $n \to \infty$, so that $n^{1/2}\delta_n \to \infty$, and under assumption (e)–(2), $n^{1/2} log(n)^{-C_2/4C_5} \delta_n \to \infty$, where C_2 and C_5 are defined in (e)–(2).

Assumption (c)–(f) are ordinary and reasonable assumptions. These assumptions were considered in tilted density function estimation by Doosti and Hall (2016). We need these assumption to find the rate of convergence of tilted linear smoother in Theorem 3. It is anticipated that $\|\hat{r}_n(.|\hat{\theta}) - r\| = O_p(\delta_n)$, where δ_n converges to 0 slower than $n^{-1/2}$ as shown in Theorem 3. Next we formulate the assumption using the first term of the expression in the left hand side of (12)

$$\lim_{C\to\infty} \limsup_{n\to\infty} \sup_{r\in\mathcal{C}} P_r(\|\check{r} - r\| > C\delta_n) = 0. \tag{14}$$

Theorem 3 *Let (c)–(f) be valid and* $\hat{\theta}$ *be defined in* (9).

I. *If in addition* $\|\check{r} - r\| = O_p(\delta_n)$ *then* $\|\hat{r}_n(.|\hat{\theta}) - r\| = O_p(\delta_n)$.

II. *If the assumptions in (e) hold uniformly for* $q \in \mathcal{C}$ *and obtain* (14) *then* (13) *is valid.*

Both Theorems 1 and 3 provide rate of convergence for tilted linear smoother. Theorem 3 instead of the assumption (b), requires the assumptions (c)–(f). The proof of Theorem 3 is given in the Appendix.

4 Simulation Study

We present the results of simulation study of the performance of tilted estimators in various settings. In the simulation study, estimators NW p4 (p10) and LL p4 (p10) refer to tilted Nadaraya-Watson and tilted local linear estimators with 4 (10) distinct values of $p_1, \cdots, p_n$, respectively. Data were generated using exponential and sin regression functions with normal and uniform design distributions. Four samples of sizes $n = (60, 100, 200, 1000)$ were sampled from the population with regression errors which had standard deviations $\sigma = (0.3, 0.5, 0.7, 1, 1.5, 2)$. In the setting for each set of σ and n, we generated 500 data sets.

We used 'constrOptim' in R for solving the optimization problem in (9) subject to the conditions in Sect. 2.2. This function is suitable for minimising a function subject to linear inequality constraints using an adaptive barrier algorithm (R Core Team, 2020). The Median Integrated Squared Error (MISE) was estimated using the Monte Carlo method.

The leave-one-out cross validation score from (5) was employed to choose the optimal bandwidths for NW and LL estimators, (Wasserman, 2006). For an infinite order flat-top kernel estimator (IO), bandwidth was selected using the rule of thumb introduced by McMurry and Politis (2004) as part of 'iosmooth' (McMurry & Politis, 2017). The bandwidth parameters for tilted estimators were estimated using our suggested procedure. To study the design bias two different design density were selected. For the exponential regression function $r_1(x) = x + 4exp(-2x^2)/\sqrt{2\pi}$, the design densities were taken to be uniform on $[-2, 2]$ and $N(0, 1)$. The Integrated Squared Error (ISE) was calculated over the interval $[-2, 2]$. The sin regression function $r_2(x) = sin(4\pi x)$ was paired with the uniform design density on $[0, 1]$. The ISE has been calculated over $[0, 1]$ and $[0.15, 0.85]$, the latter is chosen for addressing boundary effect.

In this simulation study for carrying out the MISE analysis, we had 500 replications done using Monte Carlo method. For MISE evaluation at each combination of an estimator, a function, a standard deviation and for a fixed sample size we had to solve the optimization problem. For the numerical implementation, we used the parallel computing technique in R facilitated through 'snow' (Tierney et al., 2018), 'doParallel' (Microsoft Corporation & Weston, 2019), and 'foreach' (Microsoft & Weston, 2020) packages. We did our parallel computations is Gadi, the NCI's supercomputer, Australia.

Table 1 MISE for Infinite Order (IO) estimator with the trapezoidal kernel, Nadaraya-Watson (NW) estimator, standard local linear (LL) estimator, tilted NW estimator with 4 (NW p4) and 10 (NW p10) weighting nodes, tilted LL estimator with 4 (LL p4) and 10 (LL p10) weighting nodes, Exponential regression function and normal design density function. In each row the minimum MISE is highlighted in bold

n	σ	IO	NW	LL	NW p4	NW p10	LL p4	LL p10
60	0.3	0.2070	0.0861	**0.0742**	0.1422	0.1676	0.1194	0.1566
	0.5	0.2771	0.1630	**0.1486**	0.2152	0.2294	0.1930	0.2203
	0.7	0.3755	0.2710	**0.2432**	0.3039	0.3275	0.3009	0.3223
	1	0.5888	0.4626	**0.4110**	0.5055	0.5597	0.5042	0.5261
	1.5	1.1451	0.9115	**0.8514**	0.9576	1.0860	0.9590	1.0620
	2	1.9295	1.5667	**1.4383**	1.6168	1.7774	1.6119	1.7292
100	0.3	0.0814	0.0571	**0.0462**	0.0609	0.06478	0.05156	0.0595
	0.5	0.1382	0.1044	**0.0893**	0.1131	0.1168	0.1055	0.1107
	0.7	0.2233	0.1649	**0.1454**	0.1879	0.1903	0.1744	0.1846
	1	0.3955	0.2842	**0.2534**	0.3343	0.3616	0.3166	0.3480
	1.5	0.8122	0.5757	**0.5228**	0.6764	0.7340	0.6802	0.7217
	2	1.4227	0.9750	**0.8954**	1.1569	1.3134	1.1310	1.2404
200	0.3	0.04982	0.02886	**0.0267**	0.03518	0.0435	0.0293	0.0377
	0.5	0.0740	0.0632	0.0553	0.0589	0.0640	**0.0552**	0.05780
	0.7	0.1071	0.1092	**0.0857**	0.0907	0.0936	0.08777	0.0867
	1	0.1743	0.2046	**0.1407**	0.1526	0.1568	0.1528	0.1527
	1.5	0.3395	0.43460	**0.2751**	0.3022	0.3232	0.3007	0.3165
	2	0.5732	0.7590	**0.4676**	0.5068	0.5470	0.4991	0.5310
1000	0.3	0.02481	**0.0078**	0.0121	0.0127	0.0222	0.0101	0.0187
	0.5	0.0293	0.0173	0.0197	0.0194	0.0251	**0.0166**	0.0225
	0.7	0.0353	0.0287	0.0288	0.02748	0.0304	**0.0247**	0.0277
	1	0.0502	0.0494	0.0466	0.0443	0.0435	**0.0420**	0.04245
	1.5	0.0825	0.08977	0.0812	0.0799	0.0776	0.0814	**0.0752**
	2	0.1279	0.1442	0.1254	0.1292	**0.1243**	0.1299	0.1246

In Table 1, we provide the MISEs for the proposed estimators, the comparator estimator, and the conventional estimators. Data were generated using $r_1(x)$ regression function along with normal design density and normal distribution for the error term. It is evident that for moderate sample size (n=200 and σ=0.5) the tilted estimators outperformed other estimators. Moreover, for large sample size (n=1000), as standard deviation of error terms increases, the MISEs of tilted estimators are lower than other competitors. Generally, NW p10 (LL p10) was inferior to NW p4 (LL p4). This leads the tilted linear smoother with fewer distinct values of $p_1, \cdots, p_n$ performs better, in terms of MISE.

In Table 2, we provide the MISEs for simulated data using the exponential regression function $r_1(x)$ With the uniform design density and the random normal error term. Overall, for fixed sample size (n=60, 100, 200), as the standard deviation increases, the tilted estimators outperform others. Although, for large sample sizes,

Table 2 MISE for Infinite Order (IO) estimator with the trapezoidal kernel, Nadaraya-Watson (NW) estimator, standard local linear (LL) estimator, tilted NW estimator with 4 (NW p4) and 10 (NW p10) weighting nodes, tilted LL estimator with 4 (LL p4) and 10 (LL p10) weighting nodes, exponential regression function and uniform design density. In each row the minimum MISE is highlighted in bold

n	σ	IO	NW	LL	NW p4	NW p10	LL p4	LL p10
60	0.3	0.1559	**0.0663**	0.0566	0.1308	0.1529	0.1237	0.1470
	0.5	0.1980	**0.1398**	0.1362	0.1724	0.1953	0.1690	0.1901
	0.7	0.2515	**0.2152**	0.2098	0.2316	0.2492	0.2406	0.2433
	1	0.3588	0.3697	0.3599	**0.3418**	0.3650	0.3691	0.3608
	1.5	0.6530	0.6281	0.6821	**0.6197**	0.6520	0.6676	0.6692
	2	1.0524	0.9892	2.2171	**0.9871**	1.0597	1.0287	1.0287
100	0.3	0.1195	0.0442	**0.0360**	0.1034	0.1191	0.0982	0.1156
	0.5	0.1432	0.0914	**0.0809**	0.1253	0.1426	0.1218	0.1372
	0.7	0.1781	0.1443	**0.1376**	0.1607	0.1766	0.1581	0.1695
	1	0.2490	0.2324	0.2438	0.2305	0.2469	0.2497	0.2460
	1.5	0.4165	0.4366	0.5221	**0.4041**	0.4144	0.4373	0.4198
	2	0.6487	0.6780	0.8402	**0.6107**	0.6371	0.6619	0.6401
200	0.3	0.0991	0.0232	**0.0209**	0.0891	0.0997	0.0833	0.0972
	0.5	0.1089	0.0470	**0.0441**	0.0993	0.1086	0.0944	0.1063
	0.7	0.1256	0.0822	**0.0757**	0.1172	0.1253	0.1107	0.1228
	1	0.1577	**0.1299**	0.1351	0.1533	0.1589	0.1554	0.1590
	1.5	0.2401	0.2542	0.3349	**0.2386**	0.2416	0.2587	0.2426
	2	0.3568	0.3878	0.4218	**0.3464**	0.3534	0.3938	0.3573
1000	0.3	0.0801	0.0058	**0.0056**	0.0776	0.0800	0.0724	0.0797
	0.5	0.0823	**0.0125**	0.01286	0.0790	0.0825	0.0718	0.0819
	0.7	0.0853	0.0207	**0.0206**	0.0801	0.0845	0.07170	0.0843
	1	0.0922	**0.0356**	0.0359	0.0830	0.0917	0.0728	0.0895
	1.5	0.1080	0.0716	**0.0670**	0.0972	0.1074	0.0911	0.1041
	2	0.1294	0.1286	**0.1056**	0.1209	0.1308	0.1235	0.1271

the conventional estimators tend to perform better than tilted estimators. For smaller sample sizes and the moderate standard deviation levels, the tilted NW estimator remains superior to the conventional estimators at some extent.

Table 3 presents the MISEs for the simulated data using sin function with uniform design density and normal random errors. For fixed sample size and moderate standard deviations 0.5 and 0.7, the tilted estimators perform better than conventional estimators. For sample size n=1000 with and increasing standard deviation, the tilted estimators demonstrate better performance over others.

For studying boundary effect the results provided in Tables 3 and 4 were evaluated under the identical experimental specifications except the MISEs in Table 4 were evaluated over [0.15,0.85]. According to the results when the sample size and standard deviation increased, the tilted estimators demonstrated improved performance.

Table 3 MISE for Infinite Order (IO) estimator with the trapezoidal kernel, Nadaraya-Watson (NW) estimator, standard local linear (LL) estimator, tilted NW estimator with 4 (NW p4) and 10 (NW p10) weighting nodes, tilted LL estimator with 4 (LL p4) and 10 (LL p10) weighting nodes, sin regression function, uniform design density, edges included. In each row the minimum MISE is highlighted in bold

n	σ	IO	NW	LL	NW p4	NW p10	LL p4	LL p10
60	0.3	0.04034	0.0286	0.0234	0.0315	0.0331	**0.0228**	0.0264
	0.5	0.0663	0.0638	0.0570	0.0585	0.0597	**0.0507**	0.0534
	0.7	0.1085	**0.0840**	0.1329	0.0971	0.0969	0.0897	0.090
	1	0.1958	0.1703	**0.1424**	0.1731	0.1749	0.1714	0.1692
	1.5	0.4036	**0.2410**	0.2652	0.3616	0.3718	0.3600	0.3604
	2	0.7003	**0.3739**	0.3958	0.6198	0.6463	0.6203	0.6307
100	0.3	0.0222	0.0166	**0.0129**	0.0193	0.0197	0.0130	0.0152
	0.5	0.0371	0.0326	0.0286	0.0348	0.0349	**0.0282**	0.0308
	0.7	0.0595	0.0533	**0.0498**	0.0560	0.0554	0.0506	0.0517
	1	0.1054	0.0936	**0.0868**	0.1008	0.1004	0.0989	0.09720
	1.5	0.2183	**0.1541**	0.1728	0.2078	0.2072	0.2067	0.2032
	2	0.3784	**0.2869**	0.2985	0.3520	0.3580	0.3573	0.3549
200	0.3	0.0123	0.0093	0.0070	0.0112	0.0114	**0.0067**	0.0080
	0.5	0.0191	0.0190	0.0161	0.0195	0.0194	**0.0143**	0.0150
	0.7	0.0299	0.0306	0.0273	0.0308	0.0296	**0.0251**	0.0260
	1	0.0522	0.0492	0.0494	0.0533	0.0517	0.0484	**0.0471**
	1.5	0.1073	**0.0885**	0.1043	0.1046	0.1048	0.1044	0.1011
	2	0.1830	**0.1334**	0.1432	0.1770	0.1789	0.1812	0.1752
1000	0.3	0.0056	0.0023	0.0019	0.0046	0.0051	**0.0018**	0.0027
	0.5	0.0070	0.0049	0.0042	0.0065	0.0066	**0.0034**	0.0043
	0.7	0.0089	0.0082	0.0071	0.0091	0.0091	**0.0056**	0.0064
	1	0.0132	0.0145	0.0127	0.01402	0.01346	**0.0101**	0.0108
	1.5	0.0234	0.0254	0.0241	0.0250	0.0242	**0.0214**	0.0212
	2	0.0376	0.0424	0.0386	0.0404	0.0388	0.0369	**0.0358**

It is a known fact that the performance of NW estimator deteriorates near edges, (Martinez & Martinez, 2007). This effect is often referred to as a boundary problem. The results presented in Tables 3 and 4, illustrate that the for scenarios ($n = 60$, $\sigma = 0.5$), ($n = 200$, $\sigma = 0.7$) and ($n = 1000$, $\sigma = \{1, 1.5, 2\}$) the tilted NW estimator outperformed its classical counterpart. From the boxplot in Fig. 1 it is evident that the tilted estimators have smaller median ISEs. The extreme values of the ISEs for the tilted estimators are smaller than these of the conventional estimators. Similarity between the ISE distributions and their spreads of the IO and tilted estimators can also be seen in Fig. 1.

Simulation results can be summarized as follows

- Generally NW p10 (LL p10) was inferior to NW p4 (LL p4). This leads the tilted linear smoother with fewer distinct values of $p_1, \cdots, p_n$ performs better, in terms of MISE.

Table 4 MISE for Infinite Order (IO) estimator with the trapezoidal kernel, Nadaraya-Watson (NW) estimator, standard local linear (LL) estimator, tilted NW estimator with 4 (NW p4) and 10 (NW p10) weighting nodes, tilted LL estimator with 4 (LL p4) and 10 (LL p10) weighting nodes, sin regression function, uniform design density, edges excluded. In each row the minimum MISE is highlighted in bold

n	σ	IO	NW	LL	NW p4	NW p10	LL p4	LL p10
60	0.3	0.0261	0.0182	0.0143	0.0201	0.0213	**0.0142**	0.0184
	0.5	0.0461	0.0421	0.0327	0.03747	0.0394	**0.0319**	0.0365
	0.7	0.0740	**0.0498**	0.0791	0.0628	0.0638	0.05783	0.0618
	1	0.1315	0.1130	**0.0796**	0.1189	0.1209	0.1095	0.1184
	1.5	0.2736	**0.1461**	0.1468	0.2519	0.2610	0.2420	0.2513
	2	0.4747	0.2646	**0.2208**	0.4271	0.4508	0.4215	0.4406
100	0.3	0.0132	0.0101	**0.0078**	0.0108	0.0109	0.0082	0.0096
	0.5	0.0233	0.0194	**0.0168**	0.0201	0.0207	0.0173	0.0197
	0.7	0.0380	0.0307	**0.0285**	0.0340	0.0346	0.0308	0.0331
	1	0.0688	0.0562	**0.0493**	0.0626	0.0641	0.0587	0.0626
	1.5	0.1459	**0.0912**	0.0943	0.1319	0.1365	0.1270	0.1349
	2	0.2536	0.1750	**0.1584**	0.2304	0.2391	0.2228	0.2362
200	0.3	0.0063	0.0054	0.0044	0.0054	0.0054	**0.0040**	0.0049
	0.5	0.0112	0.0121	0.0095	0.0100	0.0102	**0.0084**	0.0097
	0.7	0.0183	0.01733	0.0164	0.0162	0.0166	**0.0149**	0.0164
	1	0.0330	0.0296	0.0300	0.0298	0.0307	**0.0288**	0.0302
	1.5	0.0679	**0.0531**	0.0627	0.0636	0.0653	0.0610	0.0645
	2	0.1160	**0.0816**	0.0834	0.1116	0.1125	0.1094	0.1132
1000	0.3	0.0013	0.0014	0.0012	0.0011	0.0011	**0.0009**	0.0010
	0.5	0.0022	0.0028	0.0026	0.0019	0.0018	**0.0017**	0.0018
	0.7	0.0035	0.0045	0.0044	0.0032	0.0033	**0.0029**	0.0031
	1	0.0064	0.0089	0.0077	0.0060	0.0062	**0.0057**	0.0060
	1.5	0.0136	0.0152	0.0144	0.0126	0.0131	**0.0123**	0.0130
	2	0.0236	0.0247	0.0229	0.0219	0.0229	**0.0218**	0.0225

- In all the scenarios, tilted estimators are superior to the comparator estimator, IO, which means with the same convergence rate in theoretical sense, the proposed estimators perform better than the comparator in terms of MISE.
- For $r_1(x)$ and $X \sim N(0, 1)$ for large sample sizes (n=200, and 1000) and for small standard deviation of error term (σ= 0.5, and 0.7), tilted LL p4 outperformed other estimators.
- For $r_1(x)$ and $X \sim U(-2, 2)$ for fixed sample size, as the standard deviation increases, the tilted estimators outperform their analogs.
- For $r_2(x)$ and $X \sim U(0, 1)$, edge included, for fixed sample size and for small standard deviation of error term (σ=0.5, and 0.7), LL p4 demonstrate better performance over others.

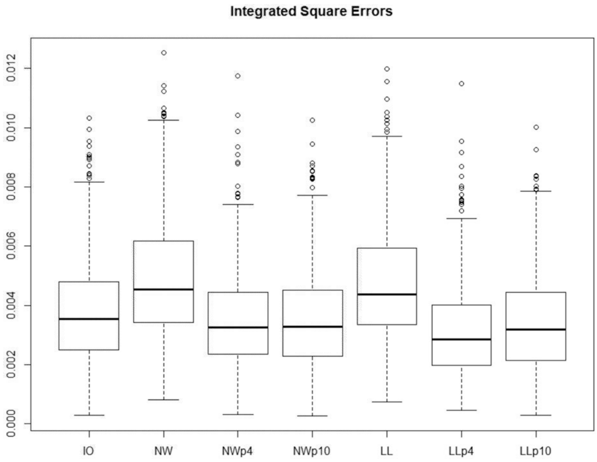


Fig. 1 Boxplots of Integrated Square Errors (ISE) for Infinite Order (IO) estimator with the trapezoidal kernel, Nadaraya-Watson (NW) estimator, standard local linear (LL) estimator, tilted NW estimator with 4 (NW p4) and 10 (NW p10) weighting nodes, tilted LL estimator with 4 (LL p4) and 10 (LL p10) weighting nodes, sin regression function, edges excluded, $n = 1000$ and $\sigma = 0.7$

- For $r_2(x)$ and $X \sim U(0, 1)$, for scenarios $(n = 60,\ \sigma = 0.5)$, $(n = 200,\ \sigma = 0.7)$ and $(n = 1000,\ \sigma = \{1, 1.5, 2\})$ the tilted NW p4 estimator outperformed NW estimator in boundaries. In the other words, the tilted NW p4 estimator addresses the boundaries problem in some conditions.

To sum up, the simulation results illustrate that under some conditions the tilted estimators (tilted NW and tilted LL) are superior to conventional estimators (NW and LL) while in all the scenarios the tilted estimators perform better than the comparator (IO).

5 Real Data

In this section, we study the performance of tilted estimators in the real data environment.

5.1 COVID-19 Data

The tilted NW estimator along with two other kernel-based estimators are being used for a curve fitting to the COVID-19 data. We shall apply the tilted NW estimator approach to daily confirmed new cases and number of daily death for 12 countries including Iran, Australia, Italy, Belgium, Germany, Spain, Brazil, United Kingdom, Canada, Chile, South Africa and United States of America, from 23 February 2020 to 28 October 2020, downloaded from https://www.ecdc.europa.eu. The logarithmic transformation has been applied, and when the number of deaths or new confirmed cases were zero we altered these observations by a positive value eliminating the associated singularity issue. The optimal bandwidth for each Nadaraya-Watson estimator was found through minimization of relevant cross-validation function, (Wasserman, 2006), at the same time we kept the bandwidth fixed for an infinite order flat-top kernel estimator which was found using Mcmurry and Politis rule of thumb, (McMurry & Politis, 2004). Along with the tilted Nadaraya-Watson estimator, we applied the Nadaraya-Watson, and Infinite Order flat-top kernel estimators. The tilted Nadaraya-Watson estimator performed the best in terms of the Mean Square Errors (MSE). Tables 5 and 6 provide the MSE for each estimator for the confirmed new cases and number of deaths. In terms of minimising the MSEs, the tilted Nadaraya-Watson estimator ranked first, followed by Infinite

Table 5 COVID-19 daily confirmed Cases: MSE for Nadaraya-Watson (NW), Infinite Order (IO), and tilted (NW p4) estimators. In each row the minimum MSE is highlighted in bold

Country	IO	NW	NW p4
Iran	324	326	**305**
Australia	5.48	5.41	**5.01**
Italy	445	1590	**372**
Belgium	2357	3071	**2232**
Germany	641	1029	**594**
Spain	41,621	41,754	**41,241**
Brazil	108,590	108,246	**107,781**
United Kingdom	1789	2083	**1737**
Canada	175	183	**173**
Chile	6040	6685	**5991**
South Africa	1158	1403	**1133**
United States of America	43,910	46,512	**41,694**

Table 6 COVID-19 death: MSE for Nadaraya-Watson (NW), Infinite Order (IO), and tilted (NW p4) estimators. In each row the minimum MSE is highlighted in bold

Country	IO	NW	NW p4
Iran	1.36	1.39	**1.32**
Australia	0.0225	0.0223	**0.0222**
Italy	1.97	3.66	**1.89**
Belgium	0.0699	0.3077	**0.0690**
Germany	0.71	0.82	**0.68**
Spain	37.74	37.92	**36.77**
Brazil	63.99	64.21	**63.86**
United Kingdom	13.57	14.50	**13.01**
Canada	0.40	0.47	**0.39**
Chile	9.10	9.73	**8.95**
South Africa	3.32	3.50	**3.22**
United States of America	201.39	205.98	**197.76**

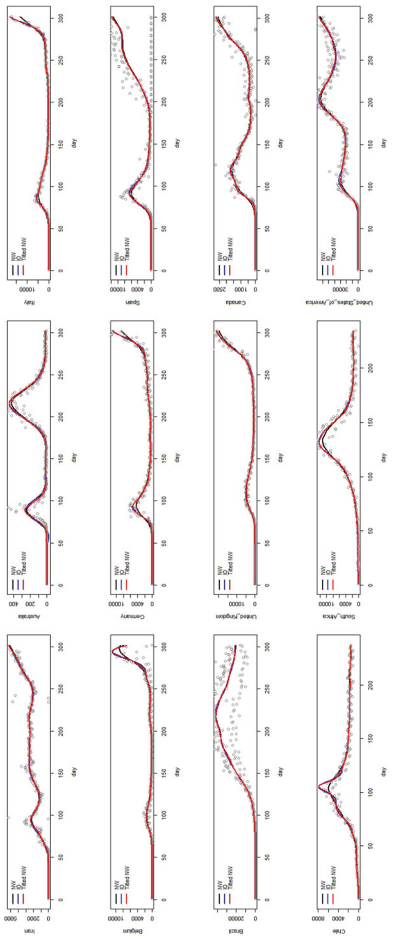


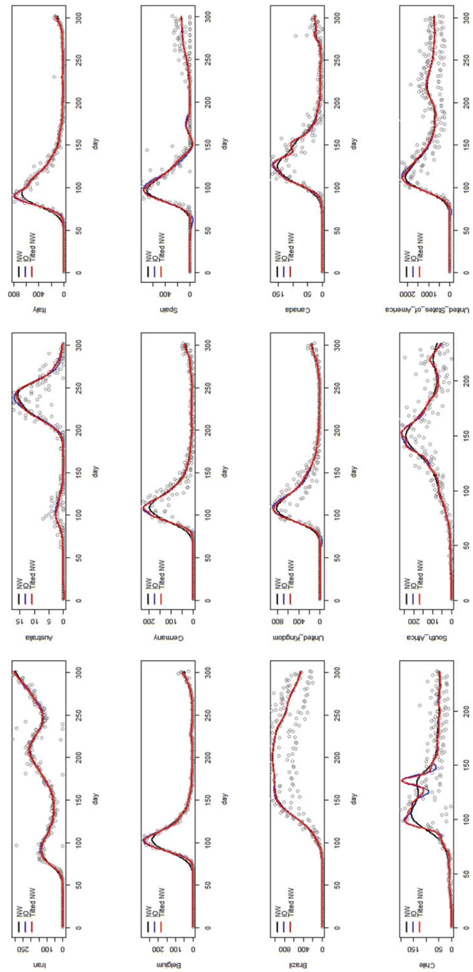
Fig. 2 Daily confirmed cases for 12 countries including Iran, Australia, Italy, Belgium, Germany, Spain, Brazil, United Kingdom, Canada, Chile, South Africa and United States of America

Order flat-top kernel and Nadaraya-Watson estimators. Figures 2 and 3 show curve fit to the daily confirmed cases and daily deaths respectively. Tables 5 and 6 provide MSE of all estimators.

5.2 Dose-Response Data

The dose-response data refers to a study of phenylephrine effects on rat corpus cavernosum strips. This data first appeared in Boroumand et al. (2016) where the dose-response curves to phenylephrine (0.1–300 μM) were obtained by applying the robust four-parameter logistic (4PL) regression. Here we have used a tilted smoother approach to dose-response curve fitting. In terms of Mean Square Errors (MSEs) the tilted local linear estimators performed better than local linear, infinite order flat-top

Fig. 3 Daily deaths for 12 countries including Iran, Australia, Italy, Belgium, Germany, Spain, Brazil, United Kingdom, Canada, Chile, South Africa and United States of America



kernel estimators including the robust 4PL model. The fitted dose-response curves using the tilted local linear, local linear, infinite order flat-top kernel estimator, and 4PL model are plotted in Fig. 4. The corresponding MSEs are listed in the caption of Fig. 4.

The original dose-response data contained the outliers and the standard 4PL model had a poor fit. Due to this we compared the performance of the tilted estimator with the robust 4PL model. The tilted local linear estimator outperformed the robust 4PL model in terms of MSE.

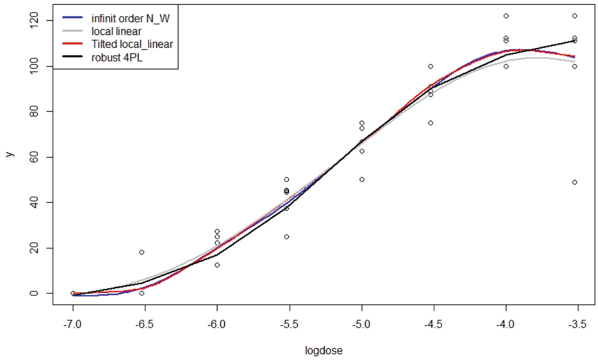


Fig. 4 Dose-response curves: MSE for tilted local linear, local linear, infinite order flat-top kernel estimator and 4PL model are 95.1023, 95.3267, 95.53077 and 110.7539, respectively

Acknowledgments The numerical results in this paper were obtained using the computational facilities of the National Computational Infrastructure (at Australian National University), of which Macquarie University is a member. This research forms part of the first author's PhD thesis approved by the ethics committee of Mashhad University of Medical Sciences with the project code 971017, Mashhad, Iran and Macquarie University, Sydney, Australia.

Appendix

Proof of Theorem 3

In this section we provide proof of Theorem 3 for tilted Nadaraya-Watson estimator as a member of the class of tilted linear smoother from (8). The proof for any tilted linear smoother could be obtained analogously.

Proof Let $p_i = \frac{1}{n}\pi(X_i)$ where $\pi \geq 0$ and is a smooth function, $\sum_{i=1}^{n} p_i = 1$ which is equivalent $\int \pi(X) f_X(x)dx = 1$ for continuous X. For simplicity, we replace $f_X(x)$ by f then

$$\begin{aligned}
\sum_{i=1}^{n} p_i &= \frac{1}{n}\sum_{i=1}^{n} \pi(X_i) \\
&= E\pi(X) + O_p(n^{-1/2}) \\
&= \int \pi(X) f dx + O_p(n^{-1/2}) \\
&= 1 + O_p(n^{-1/2}),
\end{aligned}$$

thus for normalising $p_i s$ so that $\sum_i p_i = 1$ we need to multiply the estimator (8) by $1 + O_p(n^{-1/2})$. The factor $O_p(n^{-1/2})$ is negligibly small. We choose π such that $\hat{r}_n(x|h, p)$ in (8) is unbiased estimator for r, i.e

$$E\hat{r}(x|h, p) = r. \tag{15}$$

From (15) we have

$$\begin{aligned} E\hat{r}(x)\hat{g}(x) &= \frac{1}{h}\sum_{i=1}^{n} E p_i Y_i K(\frac{x - X_i}{h}) \\ &= \frac{1}{nh}\sum_{i=1}^{n} EE\{Y_i \pi(X_i) K(\frac{x - X_i}{h})|X_i\} \\ &= \frac{1}{h}\int_{-\infty}^{\infty} r(t)\pi(t) K(\frac{x - t}{h}) f_X(t)dt, \end{aligned} \tag{16}$$

where

$$\hat{g}(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - X_i}{h})$$

and $g(x) = E\hat{g}(x)$. It can be shown that the left-hand side of (16) is converging to $r(x)g(x)$.

We have

$$r(x)g(x) = \frac{1}{h}\int_{-\infty}^{\infty} r(t')\pi(t') K(\frac{x - t'}{h}) f_X(t')dt',$$

multiplying both sides by e^{-itx} and integrating over x, we deduce

$$\Phi_{rg}(t) = \frac{1}{h}\int_{-\infty}^{\infty}\int e^{-itx} r\pi f(t') K(\frac{x - t'}{h})dx'dt',$$

by changing variable $\frac{x-t'}{h} = u$, we have

$$\begin{aligned} \Phi_{rg}(t) &= \int_{-\infty}^{\infty} e^{-itx} r\pi f(t')\Phi_k(t)dt' \\ &= \Phi_k(t)\Phi_{r\pi f}(t). \end{aligned}$$

$$\Phi_{r\pi f} = \frac{\Phi_{rg}(t)}{\Phi_k(t)},$$
$$\pi r f = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{\Phi_{rg}(t)}{\Phi_k(t)},$$
$$\pi(X) = \frac{1}{2\pi r f}\int_{-\infty}^{\infty} e^{-itx}\frac{\Phi_{rg}(t)}{\Phi_k(t)}dt, \tag{17}$$

if kernel K holds the assumption (c) and $q = r \cdot g$ meet the assumption (d), then

$$\pi = 1 + \sum_{j=1}^{n} C_j(-h^2)^j \frac{rg^{(2j)}}{rf}, \tag{18}$$

with π from (17), then $\hat{r}_n$ in unbiased. Next we show that π satisfies $0 < \pi(X) < 1$.

If the assumption (e) relaxed then there exist C_6 and $h_0 \geq 0$, for all h, $0 \leq h \leq h_0$, $\pi > 0$ and $\sup \pi \leq C_6 < \infty$ then for unbiased $\hat{r}_n$

$$\begin{aligned}\int var\{\hat{r}_n(x|h,p)\}dx &\leq \frac{1}{nh^2}\int E\{\pi^2(X)K^2(\frac{x-X}{h})\}dx,\\ &\leq \frac{1}{nh}(\sup\pi)^2\int K^2 dx,\\ &= O\{nh^{-1}\}.\end{aligned} \tag{19}$$

So MSE can be written as

$$\begin{aligned}MSE\{\hat{r}_n(x|h,p)\} &= \int E\{\hat{r}_n(x|h,p) - r\}^2 dx,\\ &= O\{nh^{-1}\}.\end{aligned}$$

We recall that

(f) Under the assumption (e)–(1), $\delta_n \to 0$ as $n \to \infty$, so that $n^{1/2}\delta_n \to \infty$, and under assumption (e)–(2), $n^{1/2}log(n)^{-C_2/4C_5}\delta_n \to \infty$, where C_2 and C_5 are defined in (e)–(2).

Then under the assumption (e)–(1) and (f), we have that $n^{1/2}\delta_n \to \infty$ thus $n^{-1} = o(\delta_n^2)$. Consequently, there exists $h(n) \downarrow 0$ as $n \to \infty$ such that $(nh)^{-1} = O(\delta_n^2)$, for some large n, $h < h_0$ since $0 \leq h \leq h_0$. Next, by replacing $O(\delta_n^2)$ in the right-hand side of (19), we have a new form of (19) which is true for specific choice of π defined at (17), and considering $\tilde{\theta} = (h, p)$ in the case of (8):

$$\lim_{C\to\infty}\limsup_{n\to\infty} P\{\|\hat{r}_n(.|\tilde{\theta}) - r\| \geq C\delta_n\} = 0. \tag{20}$$

For this version of $p_i = n^{-1}\pi(X_i)$, $\sum_i p_i = 1$ does not satisfy. However, this issue can be fixed by normalisation similar to that done in the first paragraph of the proof.

Property (20) implies part **I** of Theorem 3 and part **II** can be concluded under uniformity of (20) over $\mathcal{C}$.

Under assumption (e)–(2) and (A.4), $|rg^{(2j)}(x)/rf| \leq C_1(1+|x|)^{C_2}$ for $1 \leq j \leq k$, defining $C_6 = \max(|C_1|, \ldots, |C_k|)$, we have for $0 \leq h \leq 1$

$$|\pi(X) - 1| = |\sum_{j=1}^{n} C_j(-h^2)^j \frac{rg^{(2j)}}{rf}|,$$
$$\leq C_1 C_6 K(1+|x|)^{C_2} h_2,$$

so, if $\lambda_{1n} \to \infty$ and $\lambda_{1n}^{C_2} h^2 \to 0$, then

$$\sup |\pi(X) - 1| \to 0 \tag{21}$$

it means that whenever $|X| \leq \lambda_{1n}$, $0 < \pi(X) < 1$.

In the first paragraph of the proof, we showed that $\pi(X) \geq 0$. Then we found an upper bound for $\pi(X)$ in (21) when $X \in [-\lambda_{1n}, \lambda_{1n}]$. Now, we want to show that the probability of X being out of this interval is almost zero which means for all X, $0 < \pi(X) < 1$.

Assumption (e)–(2) implies that

$$P(|X| \geq \lambda_{1n}) \leq C_3 exp(-C_4 \lambda_{1n}^{C_5}). \tag{22}$$

Using (f), $n^{-1/2}(\log n)^{C_2/2C_5}\delta_n \to \infty$, or equivalently, $\delta_n^2 = \lambda_{2n} n^{-1}(\log n)^{C_2/2C_5} \to \infty$, where λ_{2n} exists. We choose h so that $(nh)^{-1} = O(\delta_n^2)$; or for simplicity, $(nh^{-1}) = \delta_n^2$, then $h = \lambda_{2n}^{-1}(\log n)^{-C_2/2C_5}$; let $\lambda_{1n} = \{\lambda_{2n}^{2-\eta}(\log n)^{C_5/C_2}\}^{1/C_2}$, where $\eta \in (0, 2)$, so

$$\exp(-C_4\lambda_{1n}^{C_5}) = \exp(-C_4\lambda_{2n}^{(2-\eta)C_5/C_2} \log n),$$
$$= O(n^{-C}),$$

for all $C > 0$. Therefore by (22),

$$P(|X| \geq \lambda_{1n}) = O(n^{-C}),$$

for all $C > 0$.

R Codes

This section provides the main R codes written for obtaining Tables 1, 2, and 3.

```r
# The true curve
Trufunc <- function(x) x + 4 * exp(-2 * x^2) / sqrt(2 * pi)

# Infinite order estimator
kernelTrap <- function(x) ifelse(x == 0, 1.5 / pi, (cos(x) - cos(2 * x)) / (pi * x^2))
l <- function(x, h, X) kernelTrap((X - x) / h) / sum(kernelTrap((X - x) / h))
rn_Iorder <- function(x, h, X, Y) sum(l(x, h, X) * Y)
rn_Iorder <- Vectorize(rn_Iorder, "x")

# Nadaraya Watson estimator
l_nw <- function(x, X, h) dnorm((X - x) / h) / length(X) / mean(dnorm((X - x) / h))
l_nw <- Vectorize(l_nw, "x")
rn_nw <- function(x, Y, X, h) sum(l_nw(x, X, h) * Y)
rn_nw <- Vectorize(rn_nw, "x")

# Tilted Nadaraya Watson estimator
l_Tnwp <- function(x, X, h, p) dnorm((X - x) / h) * p / mean(dnorm((X - x) / h))
l_Tnwp <- Vectorize(l_Tnwp, "x")
rn_Tnwp <- function(x, Y, X, h, p) {
  Y1 <- sort(Y)
  X1 <- X[order(Y)]
  sum((l_Tnwp(x, X1, h, p)) * Y1)
}
rn_Tnwp <- Vectorize(rn_Tnwp, "x")

MS_nw <- function(x, h, p, X, Y) (rn_Iorder(x, h_kt, X, Y) - (rn_Tnwp(x, Y, X, h, p)))^2
MS_nw <- Vectorize(MS_nw, "x")

targetfunc_nw <- function(theta, X, Y) {
  Y1 <- sort(Y)
  X1 <- X[order(Y)]
  n <- length(Y)
  k <- length(theta)
  h <- theta[k]
  pk <- k / n - sum(theta[-k])
  p <- rep(c(theta[-k], pk), each = n / k)
  integrate(MS_nw, a, b, h = h, p = p, X = X1, Y = Y1)$value
}

# Local Linear estimator
s1 <- function(x, h, X) sum(dnorm((X - x) / h) * (X - x))
s2 <- function(x, h, X) sum(dnorm((X - x) / h) * (X - x)^2)
bl <- function(x, h, X) dnorm((X - x) / h) * (s2(x, h, X) - (X - x) * s1(x, h, X))
l_ll <- function(x, X, h) bl(x, h, X) / sum(bl(x, h, X))
rn_ll <- function(x, Y, X, h) sum(l_ll(x, X, h) * Y)
rn_ll <- Vectorize(rn_ll, "x")

# Tilted Local Linear estimator
sp1 <- function(x, h, X, p) sum(p * dnorm((X - x) / h) * (X - x))
sp2 <- function(x, h, X, p) sum(p * dnorm((X - x) / h) * (X - x)^2)
bp <- function(x, h, X, p) {
  p * dnorm((X - x) / h) * (sp2(x, h, X, p) - (X - x) *
    sp1(x, h, X, p))
}
l_Tllp <- function(x, X, h, p) bp(x, h, X, p) / sum(bp(x, h, X, p))
rn_Tllp <- function(x, Y, X, h, p) {
  Y1 <- sort(Y)
  X1 <- X[order(Y)]
  sum(l_Tllp(x, X1, h, p) * Y1)
}
rn_Tllp <- Vectorize(rn_Tllp, "x")

MS_ll <- function(x, h, p, X, Y) (rn_Iorder(x, h_kt, X, Y) - rn_Tllp(x, Y, X, h, p))^2
MS_ll <- Vectorize(MS_ll, "x")
```

```
targetfunc_ll <- function(theta, X, Y) {
  n <- length(Y)
  k <- length(theta)
  h <- theta[k]
  pk <- k / n - sum(theta[-k])
  p <- rep(c(theta[-k], pk), each = n / k)
  integrate(MS_ll, a, b, h = h, p = p, X = X[order(Y)], Y = sort(Y))$value
}

# Obtaining optimum values of pi's and bandwidths for tilted estimators
tilting <- function(k, X, Y, h0, type = "nw") {
  n <- length(Y)
  ui <- rbind(diag(k), n / k * c(rep(-1, k - 1), 0))
  ci <- c(rep(0, k), -1)
  tf <- if (type == "nw") targetfunc_nw else if (type == "ll") targetfunc_ll
  theta_opt <- as.double(
    try(
      constrOptim(
        c(rep(1 / n, k - 1), h0),
        tf,
        gr = NULL, ui = ui, ci = ci, Y = Y, X = X
      )$par,
      silent = T
    )
  )
  pk <- k / n - sum(theta_opt[-k])
  p <- rep(c(theta_opt[-k], pk), each = n / k)
  return(p_opt = p, h_opt = theta_opt[k])
}

# Computing integrated square errors (ISEs) for various estimators

mainfunction <- function(X, Y, h_kt, h_nw, h_ll) {
  n <- length(Y)
  ISE_Iorder <- try(
    integrate(function(x) (Trufunc(x) - (rn_Iorder(x, h_kt, X, Y)))^2, a, b)$value,
    silent = T
  )
  ISE_nw <- try(
    integrate(function(x) (Trufunc(x) - (rn_nw(x, Y, X, h_nw)))^2, a, b)$value,
    silent = T
  )
  ISE_ll <- try(
    integrate(function(x) (Trufunc(x) - (rn_ll(x, Y, X, h_ll)))^2, a, b)$value,
    silent = T
  )
  Tnwp4_tilted <- tilting(k = 4, X = X, Y = Y, h_nw, type = "nw")
  ISE_Tnwp4 <- try(
    integrate(
      function(x) (Trufunc(x) - rn_Tnwp(x, Y, X, Tnwp4_tilted$h_opt, Tnwp4_tilted$p_opt))^2,
      a, b
    )$value,
    silent = T
  )
  Tnwp10_tilted <- tilting(k = 10, X = X, Y = Y, h_nw, type = "nw")
  ISE_Tnwp10 <- try(
    integrate(
      function(x) (Trufunc(x) - rn_Tnwp(x, Y, X, Tnwp10_tilted$h_opt, Tnwp10_tilted$p_opt))^2,
      a, b
    )$value,
    silent = T
  )
  Tllp4_tilted <- tilting(k = 4, X = X, Y = Y, h_ll, type = "ll")
  ISE_Tllp4 <- try(
    integrate(
      function(x) (Trufunc(x) - rn_Tllp(x, Y, X, Tllp4_tilted$h_opt, Tllp4_tilted$p_opt))^2,
      a, b
    )$value,
    silent = T
```