Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 9(11): 32409–32441.
DOI: 10.3934/math.20241554
Received: 12 August 2024
Revised: 31 October 2024
Accepted: 04 November 2024
Published: 15 November 2024

Research article

A Bernstein polynomial approach of the robust regression

Sihem Semmar[1,2,*], Omar Fetitah[2,3], Mohammed Kadi Attouch[2], Salah Khardani[4] and Ibrahim M. Almanjahie[5]

[1] Ecole Normale Supérieure d'Oran AMMOUR Ahmed, Oran, Algeria
[2] Laboratory of Statistics and Stochastic Processes, University of Djillali Liabes, BP 89, Sidi Bel Abbes 22000, Algeria
[3] Ecole Superieure en Informatique, Sidi Bel Abbes, Algeria
[4] Faculté des Sciences Tunis, Université El-Manar, Laboratoire de Modélisation Mathématique, Statistique et Analyse Stochastique, Tunis, Tunisia
[5] Department of Mathematics, College of Science, King Khalid University, Abha 62223, Saudi Arabia

* **Correspondence:** Email: semmar.sihem@ens-oran.dz; Tel: +213661871702.

Abstract: This paper proposes a new family of robust non-parametric estimators for regression functions by applying polynomials to construct a robust regression estimator. Theoretical results and tests on simulated and real data sets validate the efficiency and practicality of the approach. Moreover, some of its asymptotic properties are discussed and demonstrated. Experimental studies are conducted to compare this new approach with the Bernstein-Nadaraya-Watson estimator and the Nadaraya-Watson estimators. Some simulations are performed to illustrate that our robust estimator has the lowest average integrated squared error ($\overline{AISE}$). In the end, real data is utilized to assess the performance of conventional and newly presented robust regression algorithms regarding their ability to handle sensitivity to outliers.

Keywords: nonparametric regression; Bernstein polynomial; robust estimation; asymptotic normality; Nadaraya-Watson estimators
Mathematics Subject Classification: 62G05, 62G35

1. Introduction

Regression is the most frequently employed technique in nonparametric statistics to examine the association between two variables X and Y. In this context, Y represents the response variable, while

X is a random vector of predictors (covariates) that can assume values in the real number space $\mathbb{R}$. The regression function at a point $x \in \mathbb{R}$ is the conditional expectation of Y given $X = x$, denoted as

$$r(x) := \mathbb{E}(Y|X = x).$$

Various techniques can be employed to estimate a regression function, including kernel estimators, regression spline methods, and others. Nevertheless, these methods lack robustness as they are highly susceptible to outliers. Given that outliers are commonly observed in various fields, such as finance, it is essential to handle outliers properly to emphasize a dataset's unique features. Robust regression is a statistical technique used to address the issue of lack of robustness in regression models. It ensures that the model remains stable and resistant to the influence of outliers.

Robust regression holds significant importance within the realm of statistics. It is employed to overcome certain constraints of non-robust regression, specifically when the data exhibit heteroscedasticity or include outliers. The earliest significant outcome in this field can be traced back to Huber's work in [1]. The regression estimation method mentioned has been extensively researched. For empirical data, notable studies include Robinson [2], Collomb and Härdle [3], Boente and Fraiman [4, 5], and Fan et al. [6] for earlier findings. Recent advancements and references can be found in Laib and Ould-Saïd [7] and Boente and Rodriguez [8]. Traditional kernel estimators often exhibit significant bias near boundaries because the kernel's support can extend beyond them, resulting in inaccurate estimates. Being supported on the entire interval, Bernstein estimators do not suffer from this boundary bias, leading to more accurate estimations near the edges.

The Bernstein polynomial is widely acknowledged as a valuable tool for interpolating functions on a closed interval, rendering it suitable for approximating density functions within that interval.

The use of Bernstein polynomials as density estimators for variables with finite support has been proposed in several articles. Vitale [9] first introduced this concept, followed by Petrone [10, 11]. Further studies on this topic were conducted by Babu, et al. [12], Petrone and Wassermann [13] , and Kakizawa [14].

Recently, Ouimet [15] studied some asymptotic properties of Bernstein cumulative distribution function and density estimators on the d-dimensional simplex and studied their asymptotic normality and uniform strong consistency. Belalia et al. [16] introduced a two-stage Bernstein estimator for conditional distribution functions. Various other statistical topics related to the Bernstein estimator have been treated ; for more references, see Ouimet [15]. Khardani [17] investigated various asymptotic properties (bias, variance, mean squared error, asymptotic normality, uniform strong consistency) for Bernstein estimators of quantiles and cumulative distribution functions when the variable of interest is subject to random right-censoring.

It is essential to mention that several authors have devised Bernstein-based methodologies for addressing non-parametric function estimation problems. Priestley and Chao [18] first proposed the potential application of Bernstein polynomials for regression problems. Tenbusch [19] , Brown and Chen [20], Choudhuri, Ghosal, and Roy [21], Chang, Hsiung, Wu, and Yang [22], Kakizawa [23], and Slaoui and Jmaei [24] have all conducted research on various non-parametric function estimation problems.

In this paper, our contribution is to find asymptotic expressions for the bias, variance, and mean squared error (MSE) for the Bernstein robust regression function estimator defined in (2.4) and (2.3) and also prove their asymptotic normality and convergence. We deduce the asymptotically optimal bandwidth parameter m using the expression for the MSE as well. The results provided by our Bernstein approach for the robust regression function are better than those of the traditional kernel estimators. In future work, using some kernels, such as Dirichlet, Wishart , and inverse Gaussian kernels, and the robust function will be investigated in other spaces, such as the simplex, the space of positive definite matrices, and half-spaces, etc.

The subsequent sections of the paper are structured in the following manner. In the next section, we will introduce our model. Section 3 presents notations, assumptions, and investigates various asymptotic properties of the proposed estimator. Section 4 presents a simulation study that evaluates the proposed approach's performance compared to the Bernstein-Nadaraya-Watson estimator and the Nadaraya-Watson estimator. Section 5 discusses a real data application, while the proofs of the results are provided in the Appendix.

2. Robust estimation with Bernstein polynomial

Let $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ be independent, identically distributed pairs of random variables with joint density function $g(x, y)$, and let f denote the probability density of X, which is supported on $[0, 1]$. Let x be a fixed element of $\mathbb{R}$, and let ρ a real-valued Borel function that satisfies specific regularity conditions outlined below. The robust method used to study the links between X and Y belongs to the class of M-estimates introduced by Huber [1]. The robust nonparametric parameter studied in this work, denoted by θ_x, is implicitly defined as the unique minimizer w.r.t. t of

$$r(x, t) := \mathbb{E}(\rho(Y - t)|X = x), \tag{2.1}$$

that is

$$\theta_x = \arg\min_{t\in\mathbb{R}} r(x, t). \tag{2.2}$$

This definition covers and includes many important nonparametric models, for example, $\rho(t) = t^2$ yields the non-robust regression, $\rho(t) = |t|$ leads to the conditional median function $m(x) = \text{med}(Y \mid X = x)$, and the $\alpha-$th conditional quantile is obtained by setting $\rho(t) = |t| + (2\alpha - 1)(t)$. We return to Stone [25] for other examples of the function ρ.

We utilize the techniques outlined in Vitale [9] and Leblanc [26, 27] for distribution and density estimation. Additionally, we refer to the work of Slaoui [28] and Tenbusch [19, 29] for non-robust regression. Our objective is to establish a Bernstein estimator for robust regression, defined as

$$\widehat{\theta}_x = \arg\min_{t\in\mathbb{R}} \widehat{r}_n(x, t), \tag{2.3}$$

with at a given point $x \in [0, 1]$ such that $f(x) \neq 0$ and

$$\widehat{r}_n(x, t) = \frac{\sum_{i=1}^n \rho(Y_i - t) \sum_{k=0}^{m_n-1} \mathbb{I}_{\left\{\frac{k}{m_n} < X_i \leq \frac{k+1}{m_n}\right\}} B_k(m_n - 1, x)}{\sum_{i=1}^n \sum_{k=0}^{m_n-1} \mathbb{I}_{\left\{\frac{k}{m_n} < X_i \leq \frac{k+1}{m_n}\right\}} B_k(m_n - 1, x)} = \frac{N_n(x, t)}{f_n(x)}, \tag{2.4}$$

where $B_k(m, x) = \binom{m}{k} x^k (1 - x)^{m-k}$ is the Bernstein polynomial of order m. This estimator can be viewed as a generalization of the estimator proposed in Slaoui and Jmaei [28], with

$$N_n(x, t) = \frac{m_n}{n} \sum_{i=1}^n \rho(Y_i - t) \sum_{k=0}^{m_n-1} \mathbb{I}_{\left\{\frac{k}{m_n} < X_i \leq \frac{k+1}{m_n}\right\}} B_k(m_n - 1, x),$$

where f_n is Vitale's estimator of the density f defined, for all $x \in [0,1]$, by

$$\begin{aligned} f_n(x) &= \frac{m_n}{n} \sum_{i=1}^{n} \sum_{k=0}^{m_n-1} \mathbb{I}_{\left\{\frac{k}{m_n} < X \leq \frac{k+1}{m_n}\right\}} B_k(m_n-1, x) \\ &= m_n \sum_{k=0}^{m_n-1} \left\{ F_n\left(\frac{k+1}{m_n}\right) - F_n\left(\frac{k}{m_n}\right) \right\} B_k(m_n-1, x), \end{aligned} \tag{2.5}$$

with F_n, the empirical distribution function of the variable X.

This paper will use the following notations:

$$\begin{aligned} \psi(x) &= (4\pi x(1-x))^{-1/2}, \\ \Delta_1(x) &= \frac{1}{2}\left[(1-2x)f'(x) + x(1-x)f''(x)\right], \\ \Delta_2(x) &= \frac{1}{2}\left\{(1-2x)\left(\frac{\partial r}{\partial x}(x,t)f(x) + f'(x)r(x,t)\right)\right. \\ &\quad \left. + x(1-x)\left(2f'(x)\frac{\partial r}{\partial x}(x,t) + f(x)\frac{\partial^2 r}{\partial x^2}(x,t) + f''(x)r(x,t)\right)\right\}, \\ \Delta(x) &= \frac{1}{2}\left\{x(1-x)\frac{\partial^2 r}{\partial x^2}(x,t) + \left[(1-2x) + 2x(1-x)\frac{f'(x)}{f(x)}\right]\frac{\partial r}{\partial x}(x,t)\right\}, \\ \delta_1 &= \int_0^1 \Delta^2(x)dx, \quad \delta_2 = \int_0^1 \frac{\mathrm{Var}[\rho(Y-t) \mid X=x]}{f(x)}\psi(x)dx. \end{aligned}$$

Moreover, we denote by o the pointwise bound in x (i.e., the error is not uniform in $x \in [0,1]$).

Remark 2.1. *Robust regression is advantageous in real data settings where outliers, non-normal errors, or heteroscedasticity are present, making it a more flexible and resilient choice.*

3. Assumptions and main results

To state our results, we will need to gather some assumptions to make reading our results easier. In what follows, we will assume that the following assumptions hold:

Throughout the paper, C_1, C_2, C_3 represent positive constants, while C denotes a generic constant independent of n. Let $I_0 := \{x \in [0,1] \ : \ f(x) > 0\}$ and S be a compact subset of I_0.

H1: $m_n \geq 2$, $m_n \underset{n\to+\infty}{\longrightarrow} \infty$ and $m_n/n \underset{n\to+\infty}{\longrightarrow} 0$.

H2: $g(s,t)$ is twice continuously differentiable with respect to s.

H3: For $q \in \{0,1,2\}$, $s \mapsto \int_{\mathbb{R}} t^q g(s,t)dt$ is a bounded function continuous at $s = x$.

H4: For $q > 2$, $s \mapsto \int_{\mathbb{R}} |t|^{-q} g(s,t)dt$ is a bounded function.

H5: The function $\rho(.)$ is a bounded, monotone, differentiable function. Its derivative is bounded.

H6: The functions r and f are continuous and admit twice continuous and bounded derivatives such that $|\frac{\partial r}{\partial x}(x,t)| \geq C > 0$, $\forall x \in \mathbb{R}$.

H7: $r(x,.)$ is of class C^1 on $[\theta_x - \tau, \theta_x + \tau]$ and satisfies $\inf_{[\theta_x-\tau,\theta_x+\tau]} \left|\frac{\partial r}{\partial t}(x,.)\right| > C_3$ and uniformly continuous.

The assumptions we make are typical for this type of framework. Assumption (**H1**) is a technical requirement imposed to make proofs more concise. Assumptions (**H2**)–(**H4**) are necessary conditions for the estimation of the regression function in the couple (X, Y), as outlined in the works of Nadaraya [30], Watson [31], and Slaoui and Jmaei [28]. These assumptions pertain to the regularity of the density function. The condition (**H5**) controls the robustness properties of our model. It maintains the same conditions on the function ρ' as those provided by Collomb and Härdle [3] and Boente and Rodriguez [8] in the multivariate case. Assumptions (**H6**) and (**H7**) deal with some regularity of the function $r(.,.)$. Note that condition (**H6**) is used to get the asymptotic normality of our estimator, and condition (**H7**) is somewhat less restrictive compared to that presented in the literature (see Boente and Fraiman [32], L. Aït Hennani, M.Lemdani, and E. Ould Saïd [33], Attouch et al. [34, 35]), needed for the consistency result.

Proposition 3.1. *Under Assumptions (**H1**)–(**H5**), and for $x \in [0,1]$ such that $f(x) > 0$, we have*

$$\mathbb{E}\left[\widehat{r}_n(x,t)\right] - r(x,t) = \Delta(x) m_n^{-1} + o\left(m_n^{-1}\right), \tag{3.1}$$

$$\operatorname{Var}\left[\widehat{r}_n(x,t)\right] = \begin{cases} \frac{m_n^{1/2}}{n}\mathbb{E}\left[(\rho(Y-t))^2 \mid X = x\right] f(x)\psi(x) + o_x\left(\frac{m_n^{3/2}}{n}\right) & \text{for } x \in (0,1), \\ \frac{m_n}{n}\mathbb{E}\left[(\rho(Y-t))^2 \mid X = x\right] f(x) + o_x\left(\frac{m_n}{n}\right) & \text{for} x = 0, 1, \end{cases} \tag{3.2}$$

$$MSE\left[\widehat{r}_n(x,t)\right] = \begin{cases} \Delta^2(x) m_n^{-2} + \frac{m_n^{1/2}}{n}\frac{\operatorname{Var}(\rho(Y-t)|X=x)}{f(x)}\psi(x) + o\left(m_n^{-2}\right) + o_x\left(\frac{m_n^{1/2}}{n}\right) & \text{if } x \in (0,1), \\ \Delta^2(x) m_n^{-2} + \frac{m}{n}\frac{\operatorname{Var}(\rho(Y-t)|X=x)}{f(x)} + o\left(m_n^{-2}\right) + o_x\left(\frac{m_n}{n}\right) & \text{if } x = 0, 1. \end{cases} \tag{3.3}$$

To minimize the MSE of $\widehat{r}_n$, for $x \in [0,1]$ such that $f(x) > 0$, the order m_n must be equal to

$$m_{opt} = \begin{cases} \left[\frac{4\Delta^2(x) f(x)}{\operatorname{Var}(\rho(Y-t)|X=x)\psi(x)}\right]^{2/5} n^{2/5} & \text{if } x \in (0,1), \\ \left[\frac{2\Delta^2(x) f(x)}{\operatorname{Var}(\rho(Y-t)|X=x)}\right]^{1/3} n^{1/3} & \text{if } x = 0, 1. \end{cases}$$

Then,

$$MSE\left[\widehat{r}_{n,m_{opt}}(x,t)\right] = \begin{cases} \frac{5(\Delta(x))^{2/5}(\operatorname{Var}(\rho(Y-t)|X=x)\psi(x))^{4/5}}{(4f(x))^{4/5}} n^{-4/5} + o\left(n^{-4/5}\right) & \text{if } x \in (0,1), \\ \frac{3(\Delta(x)\operatorname{Var}(\rho(Y-t)|X=x))^{2/3}}{(2f(x))^{2/3}} n^{-2/3} + o\left(n^{-2/3}\right) & \text{if } x = 0, 1. \end{cases}$$

Theorem 3.1. *Under conditions of Proposition 3.1, we have*

$$\widehat{\theta}_x \xrightarrow[n\to+\infty]{\mathcal{P}} \theta_x.$$

Proposition 3.2. *Let Assumptions (**H1**)–(**H7**) hold.*

1) For $x \in (0,1)$, we have:

i) If $nm_n^{-5/2} \underset{n\to+\infty}{\longrightarrow} c$ for some constant $c \geq 0$, then

$$n^{1/2} m_n^{-1/4}\left(\widehat{r}_n(x,t) - r(x,t)\right) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(\sqrt{c}\Delta(x), \frac{\operatorname{Var}(\rho(Y-t) \mid X = x)}{f(x)}\psi(x)\right). \tag{3.4}$$

ii) If $nm_n^{-5/2} \underset{n\to+\infty}{\longrightarrow} \infty$, *then*

$$m_n\left(\widehat{r}_n(x,t) - r(x,t)\right) \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} \Delta(x). \tag{3.5}$$

2) For $x = \{0,1\}$, *we have:*

i) If $nm_n^{-3} \underset{n\to+\infty}{\longrightarrow} c$ *for some constant* $c \geq 0$, *then*

$$\sqrt{\frac{n}{m}}\left(\widehat{r}_n(x,t) - r(x,t)\right) \underset{n\to+\infty}{\overset{\mathcal{D}}{\longrightarrow}} \mathcal{N}\left(\sqrt{c}\Delta(x), \frac{\mathrm{Var}(\rho(Y-t) \mid X = x)}{f(x)}\right). \tag{3.6}$$

ii) If $nm_n^{-3} \underset{n\to+\infty}{\longrightarrow} \infty$, *then*

$$m_n\left(\widehat{r}_n(x,t) - r(x,t)\right) \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} \Delta(x), \tag{3.7}$$

where $\underset{n\to+\infty}{\overset{\mathcal{D}}{\to}}$ *denotes the convergence in distribution,* $\mathcal{N}$ *the Gaussian distribution, and* $\underset{n\to+\infty}{\overset{\mathbb{P}}{\to}}$ *the convergence in probability.*

Theorem 3.2. *(The Mean Integrated Squared Error (MISE) of* $\widehat{r}_n$*).*

*Let Assumptions (**H1**)–(**H7**) hold. Then, we have*

$$\mathrm{MISE}\left(\widehat{r}_n\right) = \frac{\Lambda_1}{m_n^2} + \Lambda_2 \frac{m_n^{1/2}}{n} + o\left(\frac{m_n^{1/2}}{n}\right) + o\left(m_n^{-2}\right). \tag{3.8}$$

Hence, the asymptotically optimal choice of m is

$$m_{opt} = \left[\frac{4\Lambda_1}{\Lambda_2}\right]^{2/5} n^{2/5},$$

for which we get

$$\mathrm{MISE}\left(\widehat{r}_{n,m_{opt}}\right) = \frac{5\Lambda_1^{1/5}\Lambda_2^{4/5}}{4^{4/5}} n^{-4/5} + o\left(n^{-4/5}\right).$$

Theorem 3.3. *Assume that (**H1**)–(**H7**) hold. If* $\Gamma(x,\theta_x) = \mathbb{E}\left[\rho'(Y-\theta_x) | X = x\right] \neq 0$, *then* $\widehat{\theta}_x$ *exists and is unique with great probability, and we have:*

i) when $x \in (0,1)$ *and* m_n *is chosen such that* $nm_n^{-5/2} \to 0$, *then*

$$n^{1/2} m_n^{-1/4}(\widehat{\theta}_x - \theta_x) \overset{\mathcal{D}}{\to} \mathcal{N}\left(\frac{\sqrt{c}\Delta(x)}{\Gamma(x,\theta_x)}, \sigma_1^2(x,\theta_x)\right),$$

ii) when $x \in [0,1]$ *and* m_n *is chosen such that* $nm_n^{-3} \to 0$, *then*

$$\sqrt{\frac{n}{m_n}}(\widehat{\theta}_x - \theta_x) \overset{\mathcal{D}}{\to} \mathcal{N}\left(\frac{\sqrt{c}\Delta(x)}{\Gamma(x,\theta_x)}, \sigma_2^2(x,\theta_x)\right),$$

where

$$\sigma_1^2(x,\theta_x) = \frac{\mathrm{Var}\left[\rho(Y-\theta_x) | X = x\right]}{f(x)\Gamma^2(x,\theta_x)} \psi(x), \ \ \sigma_2^2(x,\theta_x) = \frac{\mathrm{Var}\left[\rho(Y-\theta_x) | X = x\right]}{f(x)\Gamma^2(x,\theta_x)},$$

$\underset{n\to+\infty}{\overset{\mathcal{D}}{\to}}$ *denotes the convergence in distribution, and* $\mathcal{N}$ *the Gaussian distribution.*

The following corollary directly follows from the previous theorem and provides the weak convergence rate of the estimator $\widehat{\theta}_x$ *for* $x \in [0,1]$, *where* $f(x) > 0$. *This is specifically for the case when* m_n *is chosen such that* $nm_n^{-5/2} \to 0$ *for* $x \in (0,1)$ *and* $nm_n^{-3} \to 0$ *for* $x \in [0,1]$.

Corollary 3.1. *When $x \in (0,1)$ and m_n is chosen such that $nm_n^{-5/2} \to 0$, then*

$$n^{1/2} m_n^{-1/4} (\widehat{\theta_x} - \theta_x) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_1^2(x, \theta_x)\right).$$

When $x \in [0,1]$ and m_n is chosen such that $nm_n^{-3} \to 0$, then

$$\sqrt{\frac{n}{m_n}} (\widehat{\theta_x} - \theta_x) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_2^2(x, \theta_x)\right),$$

where

$$\sigma_1^2(x, \theta) = \frac{\mathrm{Var}\left[\rho(Y - \theta_x) |\, X = x\right]}{f(x)\Gamma^2(x, \theta_x)} \psi(x), \;\; \sigma_2^2(x, \theta_x) = \frac{\mathrm{Var}\left[\rho(Y - \theta_x) |\, X = x\right]}{f(x)\Gamma^2(x, \theta_x)}.$$

4. Simulation and real data application

This section is divided into two parts: the first shows our estimate's behavior for some particular conditional regression functions, and the second deals with asymptotic normality.

4.1. Consistency

Consider the regression model

$$Y = r(X) + \varepsilon,$$

where $\varepsilon \sim \mathcal{N}(0, 1)$.

A simulation was conducted to compare the proposed estimators $\widehat{\theta_x}$ (robust Bernstein polynomial estimator) with $\widehat{r}_n^{BNW}(x)$ (Bernstein-Nadaraya-Watson estimator) introduced by Slaoui and Jmaei [28] and defined by

$$\widehat{r}_n^{BNW}(x) = \frac{\sum_{i=1}^n Y_i \sum_{k=0}^{m_n - 1} \mathbb{I}_{\left\{\frac{k}{m_n} < X_i \le \frac{k+1}{m_n}\right\}} B_k(m_n - 1, x)}{\sum_{i=1}^n \sum_{k=0}^{m_n - 1} \mathbb{I}_{\left\{\frac{k}{m_n} < X_i \le \frac{k+1}{m_n}\right\}} B_k(m_n - 1, x)}, \tag{4.1}$$

where $B_k(m, x) = \begin{pmatrix} m \\ k \end{pmatrix} x^k (1 - x)^{m-k}$ is the Bernstein polynomial of order m, and $\widehat{r}_n^{NW}(x)$ (Nadaraya-Watson estimator) is defined, for $x \in \mathbb{R}$ such that $f(x) \neq 0$, by

$$\widehat{r}_n^{NW}(x) = \frac{\sum_{i=1}^n Y_i K(\frac{x - X_i}{h})}{\sum_{i=1}^n K(\frac{x - X_i}{h})}, \tag{4.2}$$

where $K : \mathbb{R} \to \mathbb{R}$ is a nonnegative, continuous, bounded function satisfying $\int_{\mathbb{R}} K(z)dz = 1, \int_{\mathbb{R}} zK(z)dz = 1$ and $\int_{\mathbb{R}} z^2 K(z)dz < \infty$ and $h = (h_n)$ is a sequence of positive real numbers that goes to zero.

When using the estimator $\widehat{r}_n^{NW}(x)$, we choose the Gaussian kernel $K(x) = (2\pi)^{-1/2} \exp\left(-x^2/2\right)$ and the bandwidth equal to $(h_n) = m_n^{-1}$.

We consider three sample sizes $n = 20, n = 100$, and $n = 500$, four regression functions

$$Y_i = -2X_i + 5 + \varepsilon_i \quad \text{linear case,}$$
$$Y_i = 2X_i^2 - 1 + \varepsilon_i \quad \text{parabolic case,}$$
$$Y_i = \sin\left(\frac{3}{2} X_i\right) + \varepsilon_i \quad \text{sine case,}$$
$$Y_i = \exp(2X_i - 3) + \varepsilon_i \quad \text{exponential case,}$$

and three densities of X: the truncated standard normal density $\mathcal{N}_{[0,1]}(0,1)$ ($X \in [0,1]$), the exponential density $Exp(2)$ ($X \in [0,\infty)$), and the standard normal density $\mathcal{N}(0,1)$ ($X \in (-\infty,\infty)$). It is also possible to use the transformations $\widetilde{X} = \frac{X}{1+X}$ or $\widetilde{X} = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(X)$ to cover the cases of random variables X with support $\mathbb{R}_+$ and $\mathbb{R}$, respectively. These transformations allow for the application of Bernstein polynomials to smooth the empirical distribution function.

The simulation consists of four parts. In the first three parts, the estimators are compared by their average integrated squared error $\overline{AISE}$. Every $\overline{AISE}$ is calculated by a Monte-Carlo simulation with $N = 1000$ repetitions of sample size n,

$$\overline{AISE} = \frac{1}{N}\sum_{k=1}^{N} \text{ISE}\,[\bar{r}_k]\,,$$

where $\bar{r}_k$ is the estimator ($\widehat{\theta}_x$ or $\widehat{r}_n^{BNW}(x)$ or $\widehat{r}_n^{NW}(x)$) computed from the k^{th} sample, and

$$\text{ISE}\,[\bar{r}_k] = \int_0^1 \{\bar{r}(x) - r(x)\}^2 dx.$$

According to Figures 1–4, it is evident that the robust Bernstein polynomial estimation converges when n is large. This is observed in all cases.

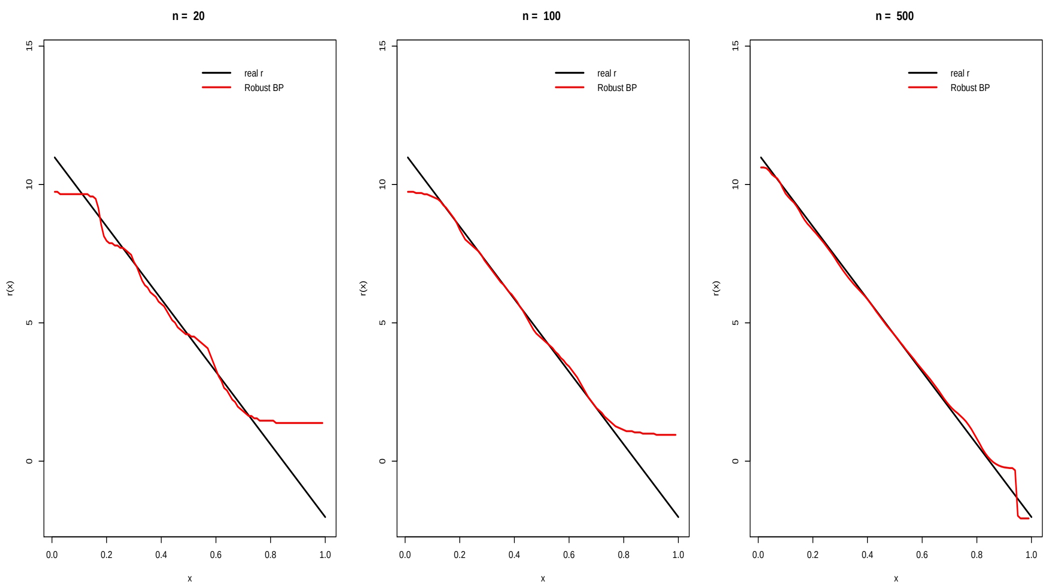


Figure 1. Prediction: linear case.

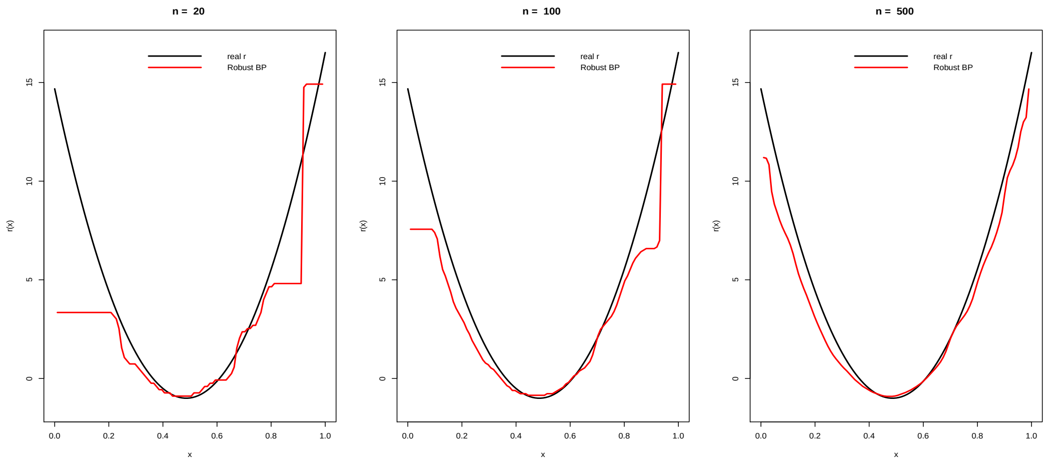


Figure 2. Prediction: parabolic case.

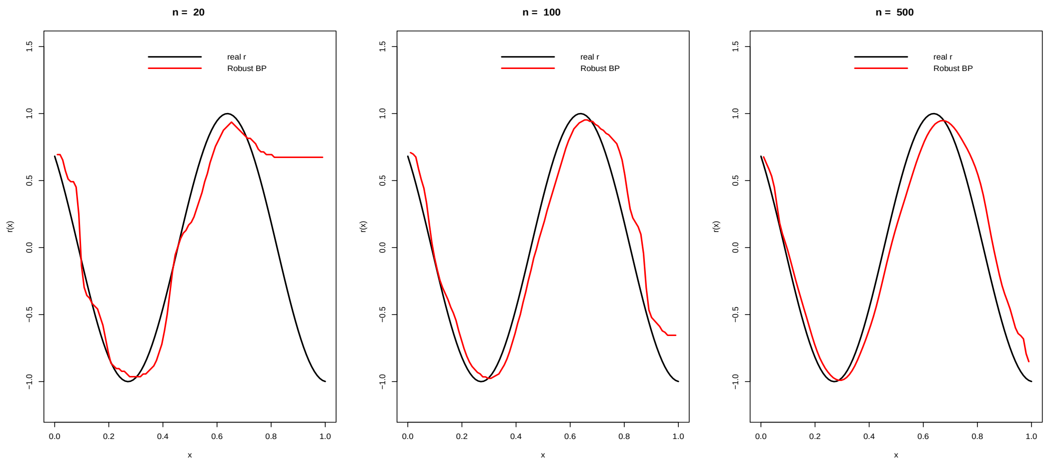


Figure 3. Prediction: sine case.

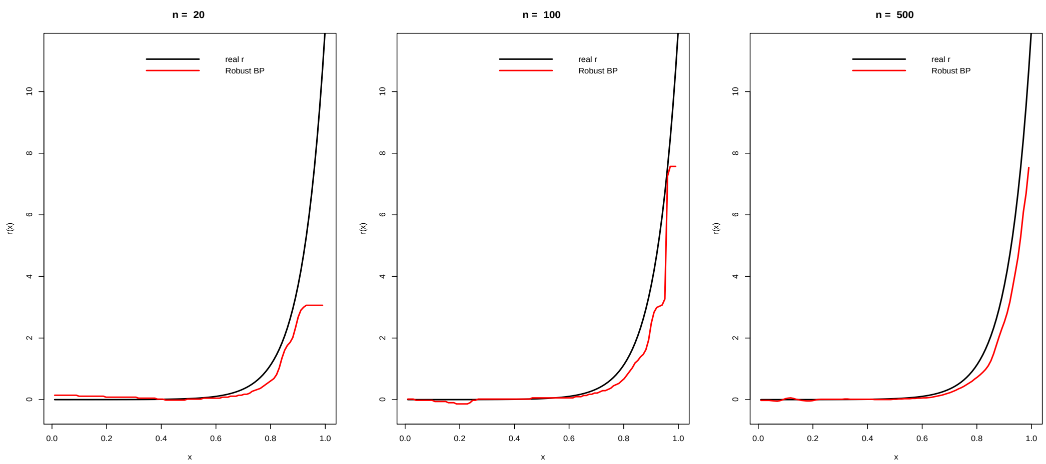


Figure 4. Prediction: exponential case.

The $\overline{AISE}$ of three estimators is graphed in Figure 5 for different parameter values ranging from 1 to 200. The estimators are evaluated for two sample sizes, $n = 20$ and $n = 500$. The outcomes are highly comparable when outlier values are not present. Nevertheless, the analysis of Tables 1–4 demonstrates that both the kernel estimator and the Bernstein-Nadaraya-Watson estimator exhibit significant sensitivity towards outlier values. This heightened sensitivity leads to substantial inaccuracies in predictions. In contrast, our robust Bernstein polynomial estimator consistently sustains its performance irrespective of the quantity of outlier values.

Table 1. $\overline{AISE}$: linear case.

Density of X	Outlier rate	$n=20$: $\widehat{r}_n^{BNW}(x)$	$n=20$: $\widehat{\theta}_x$	$n=20$: $\widehat{r}_n^{NW}(x)$	$n=100$: $\widehat{r}_n^{BNW}(x)$	$n=100$: $\widehat{\theta}_x$	$n=100$: $\widehat{r}_n^{NW}(x)$	$n=500$: $\widehat{r}_n^{BNW}(x)$	$n=500$: $\widehat{\theta}_x$	$n=500$: $\widehat{r}_n^{NW}(x)$
	0.00%	0.37777	0.38362	**0.37289**	0.0386	0.04134	**0.0333**	0.01564	0.01684	**0.00896**
(a)	0.05%	598.916	**3.57632**	690.548	678.998	**2.20528**	668.378	674.569	**0.18818**	692.737
$\mathcal{N}_{[0,1]}(0,1)$	0.10%	3016.57	**5.65957**	3000.05	2620.97	**3.16347**	2593.3	2676.12	**0.23244**	2682.89
	0.25%	16083.1	**14.182**	15878.1	15896.3	**6.29712**	15930.6	16447.1	**1.74161**	16344.1
	0.00%	0.35574	0.35578	**0.35517**	0.05012	0.05283	**0.03747**	0.01611	0.01549	**0.00794**
(b)	0.05%	748.855	**4.2097**	819.161	689.539	**1.90398**	692.123	683.571	**0.21493**	644.829
$Exp(2)$	0.10%	2408.65	**5.93149**	2284.35	2501.28	**3.57741**	2432.78	2681.96	**0.31808**	2586
	0.25%	16174.5	**21.0217**	16094.6	16228.2	**6.62117**	16834.3	17422.6	**1.89746**	17294.2
	0.00%	0.3345	0.33847	**0.31945**	0.05094	0.04832	**0.03983**	0.01667	0.01593	**0.00886**
(c)	0.05%	770.807	**4.51064**	822.317	675.495	**2.05097**	665.198	698.8	**0.15089**	656.339
$\mathcal{N}(0,1)$	0.10%	2746.52	**7.79586**	2559.23	2436.1	**3.05173**	2393.47	2497.94	**0.24955**	2503.83
	0.25%	19178.1	**18.1898**	18006	16413.5	**8.0909**	16893.9	17372.6	**1.75941**	17495.9

Table 2. $\overline{AISE}$: parabolic case.

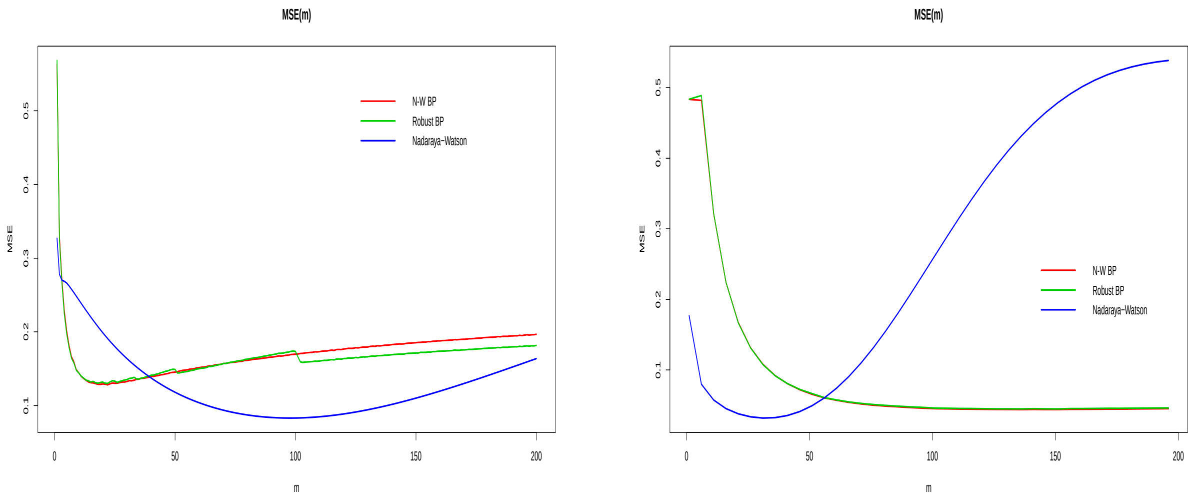
Density of X	Outlier rate	$n=20$: $\widehat{r}_n^{BNW}(x)$	$n=20$: $\widehat{\theta}_x$	$n=20$: $\widehat{r}_n^{NW}(x)$	$n=100$: $\widehat{r}_n^{BNW}(x)$	$n=100$: $\widehat{\theta}_x$	$n=100$: $\widehat{r}_n^{NW}(x)$	$n=500$: $\widehat{r}_n^{BNW}(x)$	$n=500$: $\widehat{\theta}_x$	$n=500$: $\widehat{r}_n^{NW}(x)$
	0.00%	1.48199	1.48019	**1.47376**	0.16874	0.25191	**0.10863**	0.04375	0.05576	**0.02462**
(a)	0.05%	29.1125	**2.49147**	29.3273	14.8531	**0.74622**	16.0768	15.7507	**1.66571**	22.1559
$\mathcal{N}_{[0,1]}(0,1)$	0.10%	64.0281	**2.90274**	73.1002	53.5638	**0.91596**	49.9706	98.0555	**0.90805**	67.9633
	0.25%	393.474	**6.88132**	326.212	1050.14	**3.46544**	700.888	1100.17	**0.99655**	751.459
	0.00%	1.39141	1.47957	**1.33418**	0.17829	0.22055	**0.11882**	0.05203	0.05178	**0.02936**
(b)	0.05%	25.489	**2.38298**	28.3225	13.7623	**0.54491**	14.1639	31.6953	**1.13268**	11.5467
$Exp(2)$	0.10%	71.5908	**2.58522**	75.7357	50.8943	**1.02112**	58.2273	114.747	**1.50033**	54.9639
	0.25%	355.306	**6.60588**	289.937	867.431	**2.26394**	454.397	1327.96	**0.47757**	835.553
	0.00%	0.98856	1.05261	**0.97223**	0.16172	0.18081	**0.09312**	0.03957	0.04478	**0.02101**
(c)	0.05%	25.528	**2.25141**	32.5343	22.6682	**0.5839**	15.2797	24.7481	**1.48634**	15.3258
$\mathcal{N}(0,1)$	0.10%	61.7528	**2.64854**	85.8806	75.1123	**1.22781**	45.4987	111.539	**1.81279**	64.8843
	0.25%	469.185	**9.35168**	398.637	692.756	**3.46427**	642.386	1131.51	**0.6867**	526.238

Table 3. $\overline{AISE}$: sine case.

Density of X	Outlier rate	$n=20$ $\widehat{r}_n^{BNW}(x)$	$n=20$ $\widehat{\theta}_x$	$n=20$ $\widehat{r}_n^{NW}(x)$	$n=100$ $\widehat{r}_n^{BNW}(x)$	$n=100$ $\widehat{\theta}_x$	$n=100$ $\widehat{r}_n^{NW}(x)$	$n=500$ $\widehat{r}_n^{BNW}(x)$	$n=500$ $\widehat{\theta}_x$	$n=500$ $\widehat{r}_n^{NW}(x)$
	0.00%	0.13301	0.12525	**0.1154**	0.01527	0.01522	**0.01269**	0.00414	0.00436	**0.00316**
(a)	0.05%	19.3998	**0.28717**	20.5557	12.0525	**0.12604**	10.3004	9.47988	**0.1339**	8.91102
$\mathcal{N}_{[0,1]}(0,1)$	0.10%	58.565	**0.52241**	49.1399	33.1922	**0.23475**	34.4606	38.2786	**0.16123**	43.7728
	0.25%	294.817	**2.25567**	177.939	266.142	**0.54954**	210.994	249.554	**0.30488**	273.319
	0.00%	0.14836	0.15054	**0.12541**	0.01324	0.0144	**0.01196**	0.0055	0.00545	**0.00438**
(b)	0.05%	18.8191	**0.2807**	25.6982	10.7864	**0.14311**	9.74908	10.5176	**0.18333**	9.51805
$Exp(2)$	0.10%	55.1011	**0.45925**	48.4941	44.8719	**0.18583**	33.2084	40.9161	**0.17936**	39.3012
	0.25%	234.994	**1.24542**	189.692	251.89	**0.60327**	234.829	261.372	**0.34955**	285.042
	0.00%	0.13021	0.14029	**0.12257**	0.01506	0.01511	**0.01375**	0.00442	0.00442	**0.00328**
(c)	0.05%	23.6259	**0.28918**	22.6131	12.1443	**0.10116**	10.529	9.82171	**0.15066**	9.71418
$\mathcal{N}(0,1)$	0.10%	56.098	**0.4286**	55.5514	36.7151	**0.22296**	36.6241	35.7651	**0.1501**	40.4612
	0.25%	247.54	**1.20312**	237.361	224.049	**0.50768**	235.812	246.816	**0.30212**	276.141

Table 4. $\overline{AISE}$: exponential case.

Density of X	Outlier rate	$n=20$ $\widehat{r}_n^{BNW}(x)$	$n=20$ $\widehat{\theta}_x$	$n=20$ $\widehat{r}_n^{NW}(x)$	$n=100$ $\widehat{r}_n^{BNW}(x)$	$n=100$ $\widehat{\theta}_x$	$n=100$ $\widehat{r}_n^{NW}(x)$	$n=500$ $\widehat{r}_n^{BNW}(x)$	$n=500$ $\widehat{\theta}_x$	$n=500$ $\widehat{r}_n^{NW}(x)$
	0.00%	0.74703	0.61209	**0.56581**	0.17137	0.17318	**0.1214**	0.10734	0.08411	**0.01222**
(a)	0.05%	1.98561	**1.35548**	1.86444	1.56581	**0.60637**	1.98072	5.36892	**0.08762**	3.43111
$\mathcal{N}_{[0,1]}(0,1)$	0.10%	5.07457	**1.236**	4.75581	7.88123	**0.43664**	7.38618	27.4856	**0.11352**	12.9594
	0.25%	34.2112	**2.53764**	30.5081	141.657	**1.78744**	166.463	366.283	**0.32839**	243.809
	0.00%	0.52866	0.5851	**0.47474**	0.10082	0.13134	**0.05472**	0.03985	0.07932	**0.0121**
(b)	0.05%	1.79819	**1.06047**	1.9109	1.31466	**0.45844**	1.7421	4.25689	**0.07832**	4.11877
$Exp(2)$	0.10%	3.47787	**1.33077**	4.26777	10.9025	**0.44565**	11.0732	46.7737	**0.09391**	31.0859
	0.25%	66.6146	**1.7682**	51.3565	196.824	**1.64135**	113.157	351.074	**0.30732**	317.418
	0.00%	0.74933	0.71865	**0.5849**	0.11883	0.19364	**0.11473**	0.10082	0.11145	**0.01105**
(c)	0.05%	1.20693	**0.61275**	1.51455	1.40167	**0.15919**	1.27418	7.52656	**0.0597**	3.79191
$\mathcal{N}(0,1)$	0.10%	3.85005	**1.01982**	3.50089	14.9906	**0.41717**	10.9088	39.9271	**0.11825**	36.1256
	0.25%	26.8219	**2.48908**	25.5696	145.728	**1.13974**	123.627	422.568	**0.33036**	202.644



(a) $n = 20$ **(b)** $n = 500$

Figure 5. $\overline{AISE}$ over the respective parameters in $[1, 200]$ for $n = 20$ and $n = 500$.

4.2. Asymptotic normality

The objective is to demonstrate the property of asymptotic normality in the context of the sine regression model. The equation is

$$Y_i = \sin\left(\frac{3}{2}X_i\right) + \varepsilon_i.$$

Next, let $r(x)$ be defined as the sine function with a coefficient of $\frac{3}{2}$. The data provided is the same as in the previous subsection. The procedure consists of the following steps: We approximate the regression function $r(x)$ using $\widehat{\theta}_{x_0}$ and compute the normalized deviation between this approximation and the theoretical regression function (refer to Theorem 3.3) for $x_0 = 0,\ 0.5$ and 1. Under this scheme, we generate N separate sets of n samples that are not influenced by each other. Next, we analyze the form of the estimated density (with normalized deviation) and compare it to the shape of the standard normal density in the context of the sine regression model. The following figures and table present the density of $\widehat{\theta}_{x_0}$ as well as the $p-$value by the Shapiro-Wilk normality test. We examine various values of n, specifically $n = 20$, $n = 100$, and $n = 500$.

Figures 6–8 and Table 5 demonstrate the advantageous characteristics of our asymptotic law compared to the standard normal distribution.

Table 5. $p-$value by Shapiro-Wilk normality test.

	$n = 20$	$n = 100$	$n = 500$
$x_0 = 0$	0.0814	0.0968	0.1728
$x_0 = 0.5$	0.5299	0.5734	0.6603
$x_0 = 1$	0.0611	0.0702	0.0970

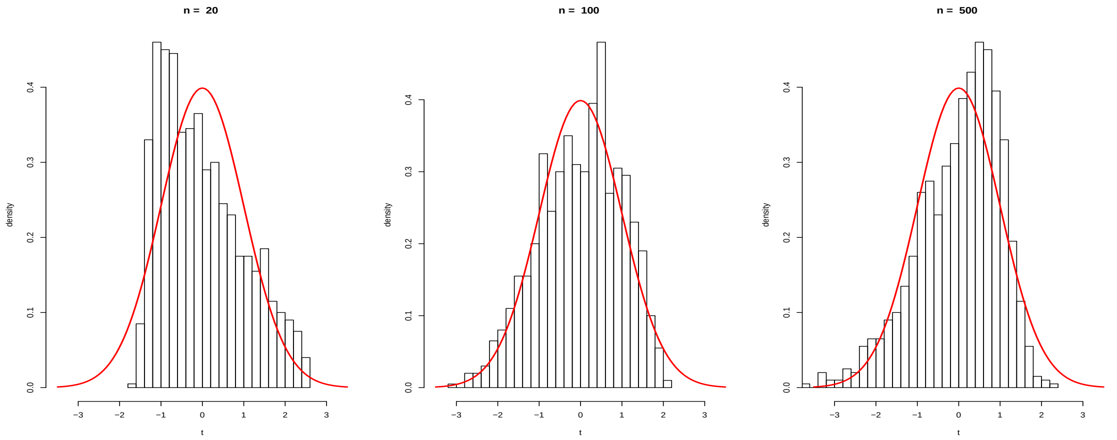


Figure 6. Illustration of the asymptotic normal distribution for $x_0 = 0$.

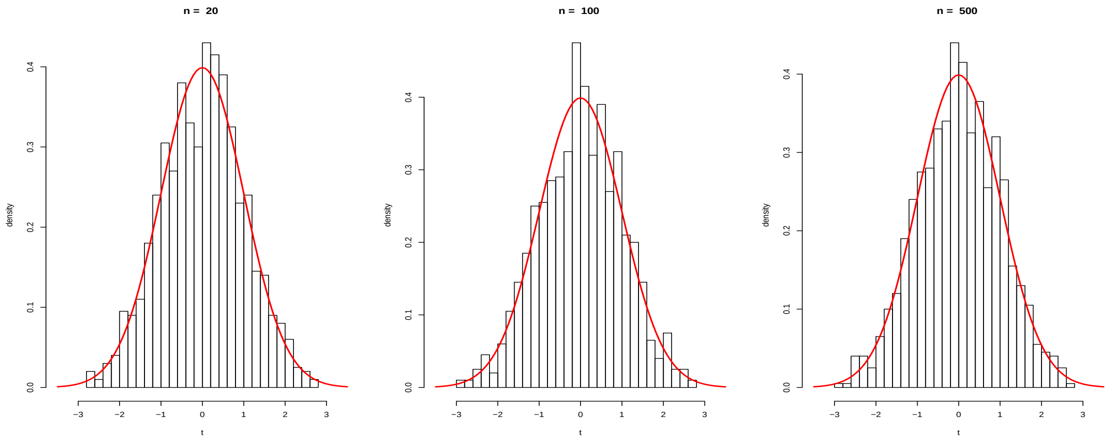


Figure 7. Illustration of the asymptotic normal distribution for $x_0 = 0.5$.

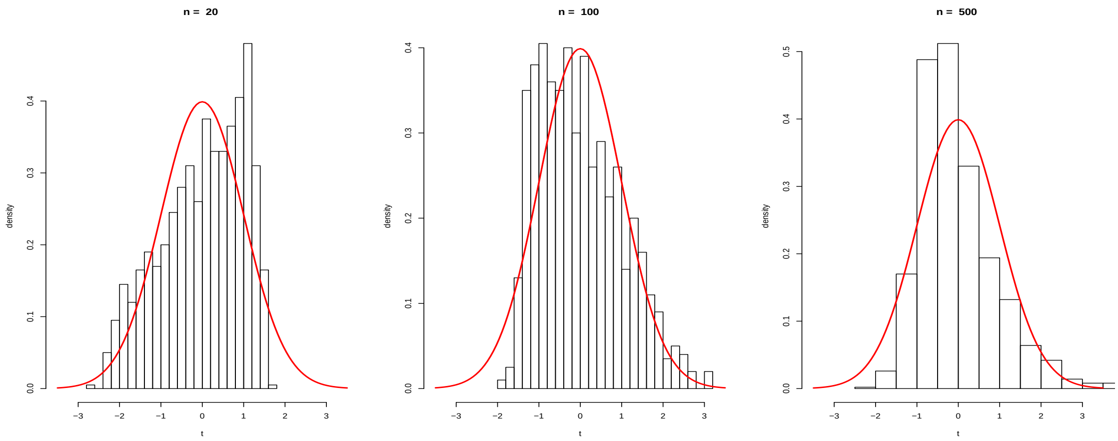


Figure 8. Illustration of the asymptotic normal distribution for $x_0 = 1$.

5. Real data application

Air pollution significantly affects the lives of individuals in developed nations. The source of this issue is increased levels of smoke produced by industries or vehicles, prompting authorities to search for more efficient methods to regulate air quality in real-time. London is experiencing a significant problem with air pollution exceeding legal and World Health Organisation limits. An example of this is the incident in 2010 when air pollution caused various health problems in the city, leading to a financial cost of around £3.7 billion.

This segment analyzes the mean daily levels of gases detected at the Marylebone Road monitoring station in London. The dataset includes the average daily measurements recorded throughout 2022 for five important variables: Ozone (O_3), Nitric Oxides (NO), Nitrogen Dioxide (NO_2), Sulphur Dioxide (SO_2), and Particulate Matter (PM_{10}). The main objective of our research is to determine the most practical forecasting models for air pollutant concentration. The data used in this analysis was obtained from the specified website: `https://www.airqualityengland.co.uk/site/data?site_id=MY1`.

To ensure clarity, let us delineate the mathematical expression representing our prediction objective. Let us consider predicting the daily air pollutant concentration, represented by the variable Y, for 365 days, denoted by X. Formally, we assume that the output variable Y and the input variable X are connected by the following equation:

$$Y_i = r(X_i) + \varepsilon_i \quad \text{for } i \in \{1, \dots, n\}.$$

A dependable data-dependent rule for order selection is crucial when estimating an unknown regression function in any practical scenario. A widely used and effective method is cross-validation:

$$CV(m) = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \bar{r}_{-i}(X_i)\right)^2,$$

where $\bar{r}_{-i}$ is the regression estimate without the data point (X_i, Y_i).

In practice, choosing the right degree m for a Bernstein polynomial requires balancing between the complexity of the model and how well it fits the data. A useful method for this is cross-validation, where the dataset is divided into training and validation sets.

Then, the smoothing parameter is chosen by minimizing

$$CV(m) = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \bar{r}_{-i}(X_i)\right)^2.$$

For convenience, we assume that the minimum of days is 1 and the maximum is 365 (the day data are such that $\min_i(X_i) = 1$ and $\max_i(X_i) = 365$). Finally, we used the cross-validation method to obtain the results in Figures 9–13 and Table 6.

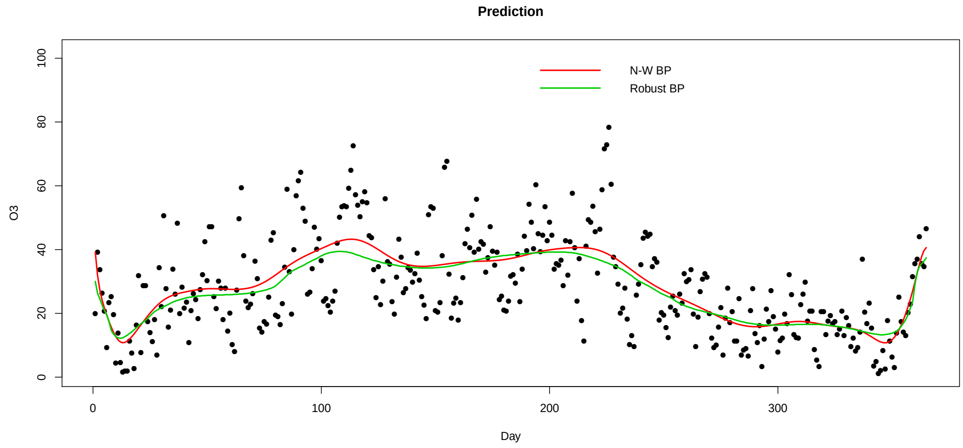


Figure 9. Prediction: Ozone (O_3) case.

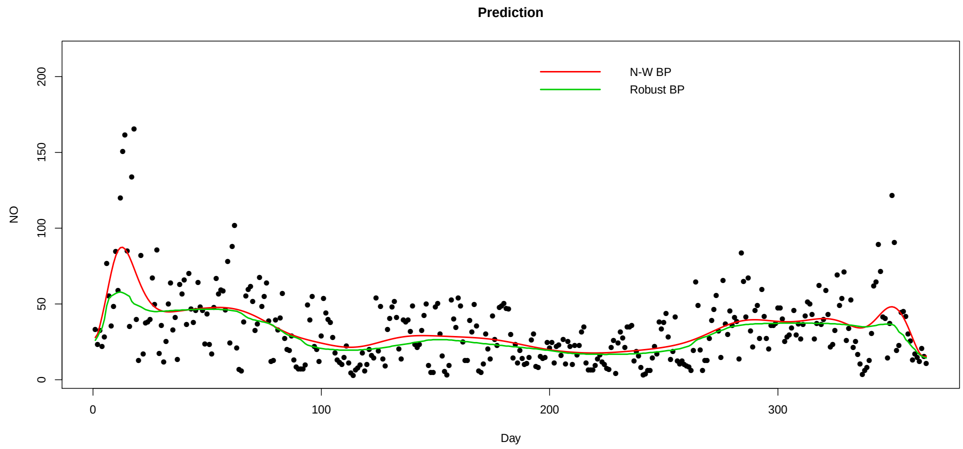


Figure 10. Prediction: Nitric Oxides (NO) case.

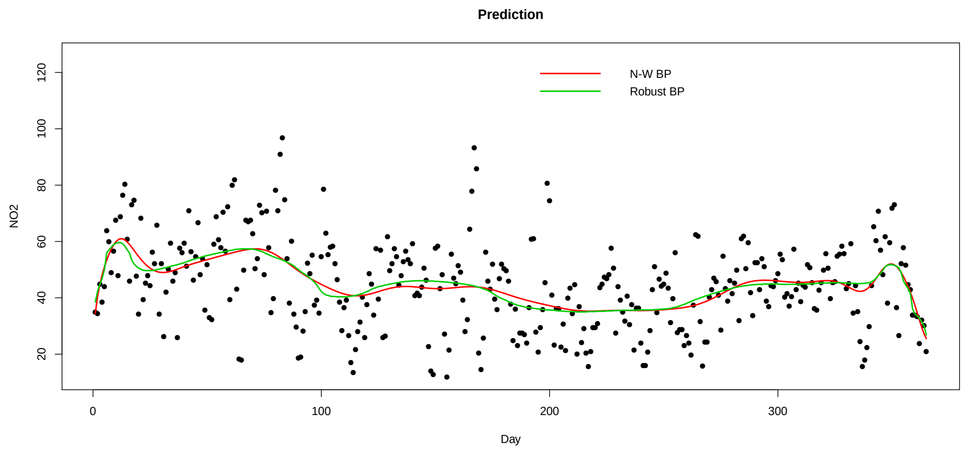


Figure 11. Prediction: Nitrogen Dioxide (NO_2) case.

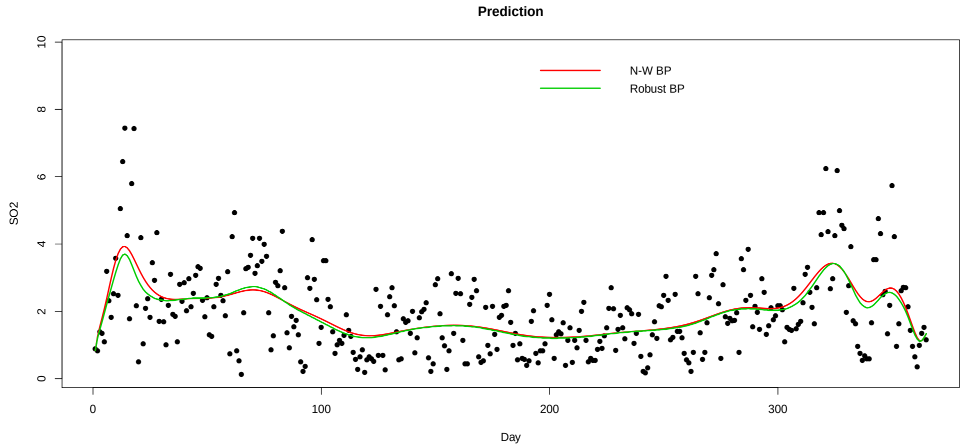


Figure 12. Prediction: Sulphur Dioxide (SO_2) case.

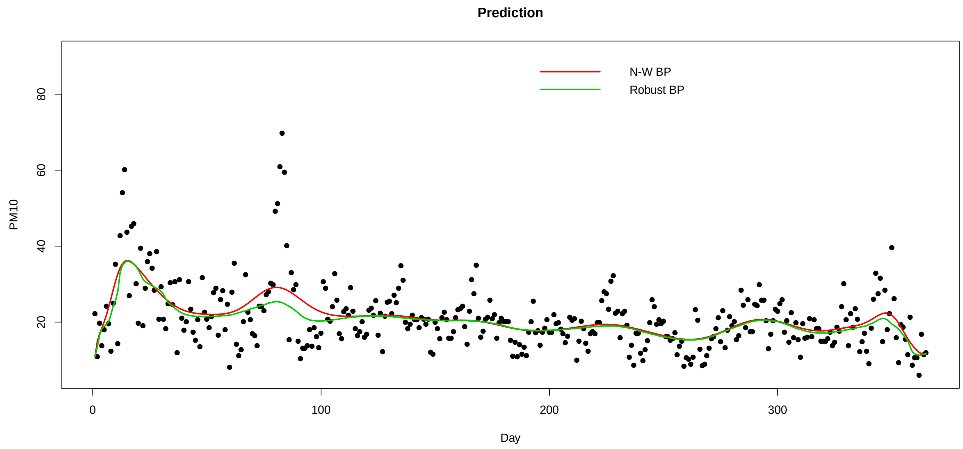


Figure 13. Prediction: Particulate Matter (PM_{10}) case.

Based on the analysis of Figures 9 to 13, it is evident that the two estimators are nearly identical, except for the scenario depicted in Figure 10. In this case, non-robust estimator $\widehat{r}_n^{BNW}(x)$ is found to be sensitive to outliers, which provides evidence of the efficiency of our estimator.

Based on the information in Table 6, we can infer that the parameter m can be adjusted. It does not need to be equal to n. Instead, we can choose a lower-degree polynomial to achieve a more favorable outcome.

Table 6. m optimal for each case.

	Ozone	Nitric Oxides	Nitrogen Dioxide	Sulphur Dioxide	Particulate Matter
$\widehat{r}_n^{BNW}(x)$	181	169	197	197	197
$\widehat{\theta}_x$	121	149	101	173	181

6. Conclusions

In this paper, we proposed a new robust regression estimator based on the Bernstein polynomials. Our contribution extends the work of Slaoui and Jmaei [28] to the case of robust regression. The asymptotic properties of this estimator were established. Afterward, we validated the effectiveness of the proposed method through a simulation study and applied it to real data on air pollution,

We found that, in all three models, the average ISE of our robust regression estimator $\widehat{\theta_x}$, defined in 2.4, was the smallest. We also noted that the robust regression provided better results than the non-robust method when outliers were present, in the sense that, even if the sample size increases, the average ISE decrease. To conclude, the use of the robust regression estimator with Bernstein polynomials successfully addressed the edge problem, yielding results comparable to those of non-robust and Nadaraya-Watson estimators in the absence of outliers.

We believe our research provides a foundational step that can be further developed and expanded. It sets the stage for future work to extend our robust regression estimator using the Bernstein polynomial by considering the interest random variable to be truncated. We also plan to work on the robust regression estimation using Lagrange polynomials.

Author contributions

Sihem Semmar: Conceptualization, data curation, formal analysis, investigation, methodology, software, validation, writing original draft, writing – review & editing; Omar Fetitah, Salah Khardani and Mohammad Kadi Attouch: Conceptualization, supervision, writing–review & editing; Mohammed Kadi Attouch and Ibrahim M. Almanjahie: Resources, validation, writing–review & editing. All authors have read and approved the final version of the manuscript for publication.

Data availability

The real data used in this application can be found at this link: `https://www.airqualityengland.co.uk/site/data?site_id=MY1`

Acknowledgments

The authors thank and extend their appreciation to the funder of this work. This work was supported by the Deanship of Scientific Research and Graduate Studies at King Khalid University through the Large Research Groups Project under grant number R.G.P. 2/338/45.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. R. J. Huber, Robust estimation of a location parmeter, *Ann. Math. Statist.*, **35** (1964), 73–101. https://doi.org/10.1214/aoms/1177703732
2. P. Robinson, Robust nonparametric autoregression, In: *Robust and nonlinear time series analysis*, New York: Springer, **26** (1986), 247–255. https://doi.org/10.1007/978-1-4615-7821-5_14
3. G. Collomb, W. Härdle, Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations, *Stoch. Process. Appl.*, **23** (1986), 77–89. https://doi.org/10.1016/0304-4149(86)90017-7
4. G. Boente, R. Fraiman, Robust nonparametric regression estimation for dependent observations, *Ann. Statist.*, **17** (1989), 1242–1256. https://doi.org/10.1214/aos/1176347266
5. G. Boente, R. Fraiman, Asymptotic distribution of robust estimators for nonparametric models from mixing processes, *Ann. Statist.*, **18** (1990), 891–906. https://doi.org/10.1214/aos/1176347631
6. J. Fan, T. C. Hu, Y. K. Truong, Robust non-parametric function estimation, *Scand. J. Statist.*, **21** (1994), 433–446.
7. N. Laïb, E. Ould-Saïd, A robust nonparametric estimation of the autoregression function under ergodic hypothesis, *Canad. J. Statist.*, **28** (2000), 817–828. https://doi.org/10.2307/3315918
8. G. Boente, D. Rodriguez, Robust estimators of high order derivatives of regression function, *Statist. Probab. Lett.*, **76** (2006), 1335–1344. https://doi.org/10.1016/j.spl.2006.01.011
9. R. A. Vitale, A Bernstein polynomial approach to density function estimation, In: *Statistical inference and related topics*, 1975, 87–99. https://doi.org/10.1016/B978-0-12-568002-8.50011-2
10. S. Petrone, Bayesian density estimation using bernstein polynomials, *Canad. J. Statist.*, **27** (1999), 105–126. https://doi.org/10.2307/3315494
11. S. Petrone, Random Bernstein polynomials, *Scand. J. Statist.*, **26** (1999), 373–393. https://doi.org/10.1111/1467-9469.00155
12. G. J. Babu, A. J. Canty, Y. P. Chaubey, Application of Bernstein polynomials for smooth estimation of a distribution and density function, *J. Statist. Plann. Inference*, **105** (2002), 377–392. https://doi.org/10.1016/S0378-3758(01)00265-8
13. S. Petrone, L. Wasserman, Consistency of Bernstein polynomial posteriors, *J. R. Stat. Soc. Ser. B Stat. Methodol.* , **64** (2002), 79–100. https://doi.org/10.1111/1467-9868.00326
14. Y. Kakizawa, Bernstein polynomial probability density estimation, *J. Nonparametr. Stat.*, **16** (2004), 709–729. https://doi.org/10.1080/1048525042000191486
15. F. Ouimet, Asymptotic properties of Bernstein estimators on the simplex, *J. Multivariate Anal.*, **185** (2021), 104784. https://doi.org/10.1016/j.jmva.2021.104784
16. M. Belalia, T. Bouezmarni, A. Leblanc, Smooth conditional distribution estimators using Bernstein polynomials, *Comput. Statist. Data Anal.*, **111** (2017), 166–182. https://doi.org/10.1016/j.csda.2017.02.005

17. S. Khardani. A Bernstein polynomial approach to the estimation of a distribution function and quantiles under censorship model, *Comm. Statist. Theory Methods*, **53** (2024), 5673–5686. https://doi.org/10.1080/03610926.2023.2228948

18. M. B. Priestley, M. T. Chao, Non-parametric function fitting, *J. R. Stat. Soc. Ser. B Stat. Methodol*, **34** (1972), 385–392. https://doi.org/10.1111/j.2517-6161.1972.tb00916.x

19. A. Tenbusch, Nonparametric curve estimation with Bernstein estimates, *Metrika*, **45** (1997), 1–30. https://doi.org/10.1007/BF02717090

20. B. M. Brown, S. X. Chen, Beta-Bernstein smoothing for regression curves with compact support, *Scand. J. Statist.*, **26** (1999), 47–59. https://doi.org/10.1111/1467-9469.00136

21. N. Choudhuri, S. Ghosal, A. Roy, Bayesian estimation of the spectral density of a time series, *J. Amer. Statist. Assoc.*, **99** (2004), 1050–1059. https://doi.org/10.1198/016214504000000557

22. I. S. Chang, C. A. Hsiung, Y. J. Wu, C. C. Yang, Bayesian survival analysis using Bernstein polynomials, *Scand. J. Statist.*, **32** (2005), 447–466. https://doi.org/10.1111/j.1467-9469.2005.00451.x

23. Y. Kakizawa, A note on generalized Bernstein polynomial density estimators, *Stat. Methodol.*, **8** (2011), 136–153. https://doi.org/10.1016/j.stamet.2010.08.004

24. Y. Slaoui, A. Jmaei, Recursive density estimators based on Robbins-Monro's scheme and using Bernstein polynomials, *Stat. Interface*, **12** (2019), 439–455. https://doi.org/10.4310/19-SII561

25. C. J. Stone, Nonparametric M-regression with free knot splines, *J. Statist. Plann. Inference*, **130** (2005), 183–206. https://doi.org/10.1016/j.jspi.2003.05.002

26. A. Leblanc, A bias-reduced approach to density estimation using Bernstein polynomials, *J. Nonparametr. Stat.*, **22** (2010), 459–475. https://doi.org/10.1080/10485250903318107

27. A. Leblanc, On estimating distribution function using Bernstein polynomials, *Ann. Inst. Stat. Math.*, **64** (2012), 919–943. https://doi.org/10.1007/s10463-011-0339-4

28. Y. Slaoui, A. Jmaei, Recursive and non-recursive regression estimators using Bernstein polynomials, *Theory Stoch. Process.*, **26** (2022), 60–95. https://doi.org/10.37863/tsp-2899660400-77

29. A. Tenbusch, Two-dimensional Bernstein polynomial density estimation, *Metrika*, **41** (1994), 233–253. https://doi.org/10.1007/BF01895321

30. E. A. Nadaraya, On estimating regression, *Theory Probab. Appl.*, **9** (1964), 141–142. https://doi.org/10.1137/1109020

31. G. S. Watson, Smooth regression analysis, *Sankhya*, **26** (1975), 359–372.

32. G. Boente, R. Fraiman, Robust nonparametric regression estimation for dependent observations, *Ann. Statist.*, **17** (1989), 1242–1256. https://doi.org/10.1214/aos/1176347266

33. L. Aït Hennani, M. Lemdani, E. O. Saïd, Robust regression analysis for a censored response and functional regressors, *J. Nonparametr. Stat.*, **31** (2019), 221–243. https://doi.org/10.1080/10485252.2018.1546386

34. M. Attouch, A. Laksaci, E. Ould-Saïd, Asymptotic distribution of robust estimator for functional nonparametric models, *Comm. Statist. Theory Methods*, **38** (2009), 1317–1335. https://doi.org/10.1080/03610920802422597

35. M. K. Attouch, A. Laksaci, E. Ould-Saïd, Robust regression for functional time series data, *J. Japan Statist. Soc.*, **42** (2012), 125–143. https://doi.org/10.14490/jjss.42.125

36. W. Feller, *An introduction to probability theory and its applications*, 2nd Eds., Chapman & Hall, 1958.

37. M. Loève, *Probability theory*, New York: Springer, 1977. https://doi.org/10.1007/978-1-4684-9464-8

A. Appendix

In this section, we present proofs for the results in Section 3. First, we recall a series of results, which are proven in Leblanc [26], linked to different sums of Bernstein polynomial, defined by

$$S_{m_n}(x) = \sum_{k=0}^{m_n-1} B_k^2(m_n, x).$$

These results are given in the following lemma.

Lemma A.1. *We have*

(i) $0 \leq S_{m_n}(x) \leq 1, \forall x \in [0,1]$.

(ii) $S_{m_n}(x) = m^{-1/2}\left[\psi(x) + o_x(1)\right], \forall x \in (0,1)$.

(iii) $S_{m_n}(0) = S_{m_n}(1) = 1$.

(iv) *Let g be any continuous function on* $[0,1]$. *Then,* $m_n^{1/2}\int_0^1 g(x)S_{m_n(x)}dx = \int_0^1 g(x)\psi(x)dx + o(1)$.

Proof. The proof of this lemma is in Leblanc [26] and Babu et al. [12]. □

A.1. Proof of Proposition 3.1

Lemma A.2.

$$\mathbb{E}\left[N_n(x,t)\right] - N(x,t) = \Delta_2(x)m_n^{-1} + o\left(m_n^{-1}\right). \tag{A.1}$$

Proof.

$$\begin{aligned}
\mathbb{E}\left[N_n(x,t)\right] &= m_n\mathbb{E}\left[\rho(Y-t)\sum_{k=0}^{m_n-1}\mathbb{I}_{\left\{\frac{k}{m_n}<X\leq\frac{k+1}{m_n}\right\}}B_k(m_n-1,x)\right] \\
&= m_n\sum_{k=0}^{m_n-1}\int_{\frac{k}{m_n}}^{\frac{k+1}{m_n}}\left(\int_{\mathbb{R}}\rho(y-t)g(z,y)dy\right)dzB_k(m_n-1,x) \\
&= m_n\sum_{k=0}^{m_n-1}\left(\int_{\frac{k}{m_n}}^{\frac{k+1}{m_n}} r(z,t)f(z)dz\right)B_k(m_n-1,x).
\end{aligned}$$

Using a Taylor expansion, we have

$$r(z,t)f(z) = \left[r(x,t) + (z-x)\frac{\partial r}{\partial z}(x,t) + \frac{(z-x)^2}{2}\frac{\partial^2 r}{\partial z^2}(x,t) + o\left((z-x)^2\right)\right]$$
$$\times\left[f(x) + (z-x)f'(x) + \frac{(z-x)^2}{2}f''(x) + o\left((z-x)^2\right)\right]$$
$$= r(x,t)f(x) + (z-x)\left[\frac{\partial r}{\partial z}(x,t)f(x) + r(x,t)f'(x)\right]$$
$$+ \frac{(z-x)^2}{2}\left[\frac{\partial^2 r}{\partial z^2}(x,t)f(x) + f''(x)r(x,t) + 2\frac{\partial r}{\partial z}(x,t)f'(x)\right] + o\left((z-x)^2\right),$$

and since $N(x,t) = r(x,t)f(x)$, we obtain

$$\mathbb{E}\left[N_n(x,t)\right] = r(x,t)f(x)m_n\sum_{k=0}^{m_n-1}\left(\frac{k+1}{m_n} - \frac{k}{m_n}\right)B_k(m_n-1,x) + \left(\frac{\partial r}{\partial x}(x,t)f(x)\right.$$
$$\left.+f'(x)r(x,t)\right)\frac{m_n}{2}\sum_{k=0}^{m_n-1}\left\{\left(\frac{k+1}{m_n}-x\right)^2 - \left(\frac{k}{m_n}-x\right)^2\right\}B_k(m_n-1,x)$$
$$+\left(f'(x)\frac{\partial r}{\partial x}(x,t) + f(x)\frac{\partial^2 r}{\partial x^2}(x,t) + f''(x)r(x,t)\right)$$
$$\frac{m_n}{6}\sum_{k=0}^{m_n-1}\left\{\left(\frac{k+1}{m_n}-x\right)^3 - \left(\frac{k}{m_n}-x\right)^3\right\}B_k(m_n-1,x)$$
$$= N(x,t) + \left(\frac{\partial r}{\partial x}(x,t)f(x) + f'(x)r(x,t)\right)\frac{m_n}{2}\sum_{k=0}^{m_n-1}m_n^{-2}(2k+1-2m_nx)B_k(m_n-1,x)$$
$$+\left(2f'(x)\frac{\partial r}{\partial x}(x,t) + f(x)\frac{\partial^2 r}{\partial x^2}(x,t) + f''(x)r(x,t)\right)$$
$$\frac{m_n}{6}\sum_{k=0}^{m_n-1}m_n^{-3}\left\{(k+1-m_nx)^2 + (k-m_nx)^2 + (k+1-m_nx)(k-m_nx)\right\}B_k(m_n-1,x)[1+o(1)]$$
$$= N(x,t) + \left(\frac{\partial r}{\partial x}(x,t)f(x) + f'(x)r(x,t)\right)\frac{m_n^{-1}}{2}\left\{2T_{1,m_n-1}(x) + (1-2x)T_{0,m_n-1}(x)\right\}$$
$$+\left(2f'(x)\frac{\partial r}{\partial x}(x,t) + f(x)\frac{\partial^2 r}{\partial x^2}(x,t) + \frac{\partial^2 f}{\partial x^2}(x)r(x,t)\right)$$
$$\frac{m_n^{-2}}{6}\sum_{k=0}^{m_n-1}\left\{3(k-m_nx)^2 + 3(k-m_nx) + 1\right\}B_k(m_n-1,x)[1+o(1)]$$
$$= N(x,t) + \left(\frac{\partial r}{\partial x}(x,t)f(x) + f'(x)r(x,t)\right)\frac{m_n^{-1}}{2}\left\{2T_{1,m_n-1}(x) + (1-2x)T_{0,m_n-1}(x)\right\}$$
$$+\left(2f'(x)\frac{\partial r}{\partial x}(x,t) + f(x)\frac{\partial^2 r}{\partial x^2}(x,t) + f''(x)r(x,t)\right)$$
$$\frac{m_n^{-2}}{6}\left\{3T_{2,m_n-1}(x) + 3(1-2x)T_{1,m_n-1}(x) + \left(x^2-3x+1\right)T_{0,m_n-1}(x)\right\}[1+o(1)],$$

where $T_{j,m_n-1}(x)$ are the central moments of the Binomial distribution of order $j \in \mathbb{N}$, defined as

$$T_{j,m_n-1}(x) = \sum_{k=0}^{m_n-1} (k - m_n x)^j B_k(m_n - 1, x), \quad \forall j \in \mathbb{N}.$$

Note that it is easy to obtain

$$T_{0,m_n-1}(x) = 1, \quad T_{1,m_n-1}(x) = 0 \quad T_{2,m_n-1}(x) = (m_n - 1)x(1 - x).$$

Then, we have

$$\mathbb{E}\left[N_n(x,t)\right] = N(x,t) + \Delta_2(x) m_n^{-1} + o\left(m_n^{-1}\right). \tag{A.2}$$

□

Lemma A.3. *We have*

$$\operatorname{Var}\left[N_n(x,t)\right] = \begin{cases} \frac{m_n^{1/2}}{n}\mathbb{E}\left[(\rho(Y-t))^2 \mid X = x\right] f(x)\psi(x) + o_x\left(\frac{m_n^{3/2}}{n}\right) & \textit{for } x \in (0,1), \\ \frac{m_n}{n}\mathbb{E}\left[(\rho(Y-t))^2 \mid X = x\right] f(x) + o_x\left(\frac{m_n}{n}\right) & \textit{for } x = 0, 1. \end{cases}$$

Proof. We have

$$\operatorname{Var}\left[N_n(x,t)\right] = \mathbb{E}\left[N_n^2(x,t)\right] - \mathbb{E}^2\left[N_n(x,t)\right],$$

where

$$\begin{aligned} N_n^2(x,t) = & \frac{m_n^2}{n^2} \sum_{i=1}^{n} (\rho(Y_i - t))^2 \left(\sum_{k=0}^{m_n-1} \mathbb{I}_{\left\{\frac{k}{m_n} < X_i \le \frac{k+1}{m_n}\right\}} B_k(m_n-1,x) \right)^2 \\ & + \frac{m_n^2}{n^2} \sum_{i,j=1, i\neq j}^{m_n} \rho(Y_i - t)\rho(Y_j - t) \left(\sum_{k=0}^{m_n-1} \mathbb{I}_{\left\{\frac{k}{m_n} < X_i \le \frac{k+1}{m_n}\right\}} B_k(m_n-1,x) \right) \left(\sum_{k=0}^{m_n-1} \mathbb{I}_{\left\{\frac{k}{m_n} < X_j \le \frac{k+1}{m_n}\right\}} B_k(m_n-1,x) \right). \end{aligned}$$

So, we have

$$\begin{aligned} \mathbb{E}\left[N_n^2(x,t)\right] &= \frac{m_n^2}{n}\mathbb{E}\left[(\rho(Y-t))^2 \left(\sum_{k=0}^{m_n-1} \mathbb{I}_{\left\{\frac{k}{m_n} < X \le \frac{k+1}{m_n}\right\}} B_k(m_n-1,x) \right)^2 \right] + \frac{m_n^2 n(n-1)}{n^2} \mathbb{E}^2\left[\sum_{k=0}^{m_n-1} \mathbb{I}_{\left\{\frac{k}{m_n} < X \le \frac{k+1}{m_n}\right\}} B_k(m_n-1,x) \right] \\ &= \frac{m_n}{n}\mathbb{E}\left[(\rho(Y-t))^2 \left(\sum_{k=0}^{m_n-1} \mathbb{I}_{\left\{\frac{k}{m_n} < X \le \frac{k+1}{m_n}\right\}} B_k(m_n-1,x) \right)^2 \right] + \left(1 - \frac{1}{n}\right) \mathbb{E}^2\left[N_n(x,t)\right], \end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}\left[N_n(x,t)\right] &= \frac{m_n^2}{n}\mathbb{E}\left[(\rho(Y-t))^2\left(\sum_{k=0}^{m_n-1}\mathbb{I}_{\left\{\frac{k}{m_n}<X\leq\frac{k+1}{m_n}\right\}B_k(m_n-1,x)}\right)^2\right]-\frac{1}{n}\mathbb{E}^2\left[N_n(x,t)\right]\\
&= \frac{m_n^2}{n}\mathbb{E}\left[(\rho(Y-t))^2\sum_{k=0}^{m_n-1}\mathbb{I}_{\left\{\frac{k}{m_n}<X\leq\frac{k+1}{m_n}\right\}}B_k^2(m_n-1,x)\right]-\frac{1}{n}\mathbb{E}^2\left[N_n(x,t)\right]\\
&= \frac{m_n^2}{n}\sum_{k=0}^{m_n-1}\int_{\frac{k}{m_n}}^{\frac{k+1}{m_n}}\left(\int_{\mathbb{R}}(\rho(y-t))^2 g(z,y)dy\right)dzB_k^2(m_n-1,x)-\frac{1}{n}\mathbb{E}^2\left[N_n(x,t)\right]\\
&= \frac{m_n^2}{n}\sum_{k=0}^{m_n-1}\left(\int_{\frac{k}{m_n}}^{\frac{k+1}{m_n}}\mathbb{E}\left[(\rho(Y-t))^2\mid X=z\right]f(z)dz\right)B_k^2(m_n-1,x)-\frac{1}{n}\mathbb{E}^2\left[N_n(x,t)\right]\\
&= \frac{m_n}{n}\mathbb{E}\left[(\rho(Y-t))^2\mid X=x\right]f(x)S_{m_n}(x)-\frac{1}{n}\mathbb{E}^2\left[N_n(x,t)\right].
\end{aligned}$$

Using Lemma A.1 (ii) and (iii), we obtain

$$\mathrm{Var}\ \left[N_n(x,t)\right]=\begin{cases}\frac{m_n^{1/2}}{n}\mathbb{E}\left[(\rho(Y-t))^2\mid X=x\right]f(x)\psi(x)+o_x\left(\frac{m_n^{3/2}}{n}\right) & \text{for } x\in(0,1),\\ \frac{m_n}{n}\mathbb{E}\left[(\rho(Y-t))^2\mid X=x\right]f(x)+o_x\left(\frac{m_n}{n}\right) & \text{for } x=0,1.\end{cases}\tag{A.3}$$

□

Lemma A.4.

$$\mathrm{Cov}\left(f_n(x),N_n(x,t)\right)=\begin{cases}\frac{m_n^{1/2}}{n}r(x,t)f(x)\psi(x)+o_x\left(\frac{m_n^{1/2}}{n}\right) & \textit{for } x\in(0,1),\\ \frac{m_n}{n}r(x,t)f(x)+o_x\left(\frac{m_n}{n}\right) & \textit{for } x=0,1.\end{cases}\tag{A.4}$$

Proof. We have

$$\begin{aligned}
\mathrm{Cov}\left(f_n(x),N_n(x,t)\right) &= \mathbb{E}\left[f_n(x)N_n(x,t)\right]-\mathbb{E}\left[f_n(x)\right]\mathbb{E}\left[N_n(x,t)\right]\\
&= \frac{m_n^2}{n}\mathbb{E}\left[\rho(Y-t)\left(\sum_{k=0}^{m_n-1}\mathbb{I}_{\left\{\frac{k}{m_n}<X\leq\frac{k+1}{m_n}\right\}}B_k(m_n-1,x)\right)^2\right]\\
&\quad+\frac{n(n-1)m_n^2}{n^2}\mathbb{E}^2\left[\rho(Y-t)\sum_{k=0}^{m_n-1}\mathbb{I}_{\left\{\frac{k}{m_n}<X\leq\frac{k+1}{m_n}\right\}}B_k(m_n-1,x)\right]-\mathbb{E}\left[f_n(x)\right]\mathbb{E}\left[N_n(x,t)\right]\\
&= \frac{m_n^2}{n}\mathbb{E}\left[\rho(Y-t)\left(\sum_{k=0}^{m_n-1}\mathbb{I}_{\left\{\frac{k}{m_n}<X\leq\frac{k+1}{m_n}\right\}}B_k(m_n-1,x)\right)^2\right]-\frac{1}{n}\mathbb{E}\left[f_n(x)\right]\mathbb{E}\left[N_n(x,t)\right]\\
&= \frac{m_n^2}{n}\sum_{k=0}^{m_n-1}\int_{\frac{k}{m_n}}^{\frac{k+1}{m_n}}\left(\int_{\mathbb{R}}\rho(y-t)g(z,y)dy\right)dzB_k^2(m_n-1,x)-\frac{1}{n}\mathbb{E}\left[f_n(x)\right]\mathbb{E}\left[N_n(x,t)\right]\\
&= \frac{m_n}{n}r(x,t)f(x)S_m(x)-\frac{1}{n}\mathbb{E}\left[f_n(x)\right]\mathbb{E}\left[N_n(x,t)\right].
\end{aligned}$$

Using Lemma A.1 (ii) and (iii), we get

$$\operatorname{Cov}\left(f_n(x), N_n(x)\right) = \begin{cases} \frac{m_n^{1/2}}{n} r(x) f(x) \psi(x) + o_x\left(\frac{m_n^{1/2}}{n}\right) & \text{for } x \in (0,1), \\ \frac{m_n}{n} r(x) f(x) + o_x\left(\frac{m_n}{n}\right) & \text{for } x = 0, 1. \end{cases} \tag{A.5}$$

To obtain the bias of $\widehat{r_n}(x,t)$, we let $h(x,y) = \frac{u}{v}$. Using a Taylor expansion, we have

$$\begin{aligned} h(u,v) =& h(u_0,v_0) + [u-u_0]\frac{\partial h}{\partial u}(u_0,v_0) + [v-v_0]\frac{\partial h}{\partial v}(u_0,v_0) \\ &+ \frac{1}{2}\left\{[u-u_0]^2 \frac{\partial^2 h}{\partial u^2}(u_0,v_0) + [v-v_0]^2 \frac{\partial^2 q}{\partial v^2}(u_0,v_0)\right\} + 2[u-u_0][v-v_0]\frac{\partial^2 h}{\partial u \partial v}(u_0,v_0) \\ &+ o\left(\|(u-u_0, v-v_0)\|^2\right). \end{aligned}$$

Then, we have

$$\frac{u}{v} = \frac{u_0}{v_0} + \frac{1}{v_0}(u-u_0) - \frac{u_0}{v_0^2}(v-v_0) + \frac{u_0}{v_0^3}(v-v_0)^2 - \frac{1}{v_0^2}(u-u_0)(v-v_0) + o\left((u-u_0)^2 + (v-v_0)^2\right).$$

We set $(u,v) = (N_n(x,t), f_n(x))$ and $(u_0,v_0) = (N(x,t), f(x))$. Therefore, we infer that

$$\begin{aligned} \frac{N_n(x,t)}{f_n(x)} =& \frac{N(x,t)}{f(x)} + \frac{1}{f(x)}\left(N_n(x,t) - N(x,t)\right) - \frac{N(x,t)}{f(x)^2}\left(f_n(x) - f(x)(x)\right) \\ &+ \frac{N(x,t)}{f(x)^3}\left(f_n(x) - f(x)\right)^2 - \frac{1}{f(x)^2}\left(N_n(x,t) - N(x,t)\right)\left(f_n(x) - f(x)\right)) \\ &+ o\left((N_n(x,t) - N(x,t))^2 + (f_n(x) - f(x))^2\right). \\ \widehat{r_n}(x,t) =& r(x,t) + \frac{1}{f(x)}\left(N_n(x,t) - N(x,t)\right) - \frac{r(x,t)}{f(x)}\left(f_n(x) - f(x)\right) \\ &+ \frac{r(x,t)}{f(x)^2}\left(f_n(x) - f(x)\right)^2 - \frac{1}{f(x)^2}\left(N_n(x,t) - N(x,t)\right)\left(f_n(x) - f(x)\right) \\ &+ o\left((N_n(x,t) - N(x,t))^2 + (f_n(x) - f(x))^2\right). \end{aligned}$$

Hence, we set $(u,v) = (f_n(x), N_n(x,t))$ and $(u_0,v_0) = (f(x), N(x,t))$ to obtain

$$\begin{aligned} \widehat{r_n}(x,t) =& r(x,t) - \frac{r(x,t)}{f(x)}\left(f_n(x) - f(x)\right) + \frac{1}{f(x)}\left(N_n(x,t) - N(x,t)\right) \\ &+ \frac{r(x,t)}{\{f(x)\}^2}\left(f_n(x) - f(x)\right)^2 - \frac{1}{\{f(x)\}^2}\left(f_n(x) - f(x)\right)\left(N_n(x,t) - N(x,t)\right) \\ &+ o\left((f_n(x) - f(x))^2 + (f_n(x) - f(x))\left(N_n(x,t) - N(x,t)\right)\right). \end{aligned}$$

Then,

$$\begin{aligned} \mathbb{E}\left[\widehat{r_n}(x,t)\right] =& r(x,t) - \frac{r(x,t)}{f(x)}\left(\mathbb{E}\left[f_n(x)\right] - f(x)\right) + \frac{1}{f(x)}\left(\mathbb{E}\left[N_n(x,t)\right] - N(x,t)\right) \\ &+ \frac{r(x,t)}{\{f(x)\}^2}\left(\mathbb{E}\left[f_n(x)\right] - f(x)\right)^2 - \frac{1}{\{f(x)\}^2}\mathbb{E}\left[\left(f_n(x) - f(x)\right)\left(N_n(x,t) - N(x,t)\right)\right] \\ &+ o\left(\mathbb{E}\left[(f_n(x) - f(x))^2\right] + \mathbb{E}\left[\left(f_n(x) - f(x)\right)\left(N_n(x,t) - N(x,t)\right)\right]\right). \end{aligned}$$

Use Vitale's estimator f_n, we get

$$\mathbb{E}\left[f_n(x)\right] = f(x) + \frac{\Delta_1(x)}{m_n} + o\left(m_n^{-1}\right), \quad \forall x \in [0,1] \tag{A.6}$$

and

$$\operatorname{Var}\left[f_n(x)\right] = \begin{cases} \frac{m_n^{1/2}}{n} f(x)\psi(x) + o_x\left(\frac{m_n^{1/2}}{n}\right) & \text{for } x \in (0,1), \\ \frac{m_n}{n} f(x) + o_x\left(\frac{m_n}{n}\right) & \text{for } x = 0, 1. \end{cases} \tag{A.7}$$

To obtain (3.1) of Proposition 3.1, we use (A.6) and (A.2) to obtain

$$\begin{aligned}\mathbb{E}\left[\widehat{r}_n(x,t)\right] &= r(x,t) + \left(\frac{1}{f(x)}\Delta_2(x) - \frac{r(x,t)}{f(x)}\Delta_1(x)\right) m_n^{-1} + o\left(m_n^{-1}\right) \\ &= r(x,t) + \Delta(x) m_n^{-1} + o\left(m_n^{-1}\right), \quad \forall x \in [0,1].\end{aligned}$$

Now for the variance of $\widehat{r}_n(x,t)$, we have

$$\operatorname{Var}\left(\widehat{r}_n(x,t)\right) = \operatorname{Var}\left(r(x,t) - \frac{r(x,t)}{f(x)}\left(f_n(x) - f(x)\right) + \frac{1}{f(x)}\left(N_n(x,t) - N(x,t)\right)\right)[1 + o(1)],$$

which ensures that

$$\operatorname{Var}\left(\widehat{r}_n(x,t)\right)\left\{\frac{r^2(x,t)}{f^2(x)}\operatorname{Var}\left(f_n(x)\right) + \frac{1}{f^2(x)}\operatorname{Var}\left(N_n(x,t)\right) - 2\frac{r(x,t)}{f^2(x)}\operatorname{Cov}\left(N_n(x,t), f_n(x)\right)\right\}[1 + o(1)].$$

So, for $x = (0,1)$, we have f,

$$\operatorname{Var}\left[\widehat{r}_n(x,t)\right] = \frac{m_n^{1/2}}{n}\frac{\operatorname{Var}(\rho(Y-t) \mid X = x)}{f(x)} + o_x\left(\frac{m_n^{1/2}}{n}\right),$$

and, for $x \in 0, 1$, we have

$$\operatorname{Var}\left[\widehat{r}_n(x,t)\right] = \frac{m_n}{n}\frac{\operatorname{Var}(\rho(Y-t) \mid X = x)}{f(x)} + o_x\left(\frac{m_n}{n}\right),$$

which gives the proof of Proposition 3.1. □

A.2. Proof of Theorem 3.1

Without loss of generality we can suppose that $\rho(Y - .)$ is increasing, with the decreasing case being obtained by considering $-\rho(Y - .)$. As $\rho(Y - .)$ is increasing, then for all $\epsilon > 0$,

$$r(x, \theta_{x.} + \epsilon) \leq r(x, \theta_x) \leq r(x, \theta_x - \epsilon).$$

Proposition 3.1 shows that

$$\widehat{r}(x,t) \xrightarrow{\mathbb{P}} r(x,t),$$

for all real $t \in [\theta_x - \tau, \theta_x + \tau]$. As $r(x, \theta_x) = 0$, for sufficiently large n and for all $\epsilon \leq \tau$, this implies

$$\widehat{r}(x, \theta_x + \epsilon) \leq 0 \leq \widehat{r}(x, \theta_x - \epsilon) \quad \text{in probability.}$$

Since $\widehat{r}(x,\widehat{\theta_x}) = 0$, and by the continuity of $\widehat{r}(x,.)$ on $[\theta_x - \tau, \theta_x + \tau]$, we deduce that

$$\theta_x - \epsilon \le \hat{\theta}_x \le \theta_x + \epsilon \quad \text{in probability.}$$

On the other hand, since θ_x and $\hat{\theta}_x$ are solutions of $r(x,t)$ and $\widehat{r}(x,t)$, respectively, then we have

$$\widehat{r}(x,\hat{\theta}_x) = r(x,\theta_x) = 0.$$

Under (**H7**), and by a Taylor expansion of $r(x,.)$ of order one around $\hat{\theta}_x$, we have

$$\widehat{r}(x,\hat{\theta}_x) - r(x,\hat{\theta}_x) = (\theta_x - \hat{\theta}_x)\frac{\partial r}{\partial t}(x,\xi_n),$$

where ξ_n is between θ_x and $\hat{\theta}_x$. Hence,

$$|\theta_x - \hat{\theta}_x| \le \frac{1}{|\inf_{x\in S} \frac{\partial r}{\partial t}(x,\xi_n)|}\left|\hat{r}(x,\hat{\theta}_x) - r(x,\hat{\theta}_x)\right|,$$

which yields

$$\begin{aligned}\sup_{x\in S}|\theta_x - \hat{\theta}_x| &\le \frac{1}{C_3}\sup_{x\in S}\left|\hat{r}(x,\hat{\theta}_x) - r(x,\hat{\theta}_x)\right| \\ &\le \frac{1}{C_3}\sup_{x\in S}\sup_{t\in[\theta_x-\tau,\theta_x+\tau]}|\hat{r}(x,t) - r(x,t)|,\end{aligned}$$

and the rest of the proof is a sequence of Proposition 3.1.

A.3. Proof of Proposition 3.2

From (2.4), we adopt the decomposition stated as

$$\begin{aligned}\widehat{r_n}(x,t) - r(x,t) =& \frac{1}{f_n(x)}\left[(N_n(x,t) - N(x,t)) - r(x,t)(f_n(x) - f(x))\right] \\ =& \frac{1}{f_n(x)}\left[(N_n(x,t) - \mathbb{E}(N_n(x,t))) - r(x,t)(f_n(x) - \mathbb{E}(f_n(x)))\right] \\ &+ \frac{1}{f_n(x)}\left[(\mathbb{E}(N_n(x,t)) - N(x,t)) - r(x,t)(\mathbb{E}(f_n(x)) - f(x))\right].\end{aligned}$$

Lemma A.5. *Under Assumptions* ***(H1)–(H3)****, and for* $x \in [0,1]$ *such that* $f(x) > 0$*, we have*

$$f_n(x) \xrightarrow{\mathbb{P}} f(x). \tag{A.8}$$

Proof. We have by the results of Lemmas A.2 and A.3, that

$$\mathbb{E}(f_n(x)) - f(x) \to 0,$$

and

$$\mathrm{Var}(f_n(x)) \to 0.$$

Hence,

$$f_n(x) \xrightarrow{\mathbb{P}} f(x),\ \forall\ x \in (0,1).$$

□

Lemma A.6. *Under Assumptions (**H1**)–(**H4**), and for $x \in (0,1)$ such that $f(x) > 0$, we have:*

i) if m_n is chosen such that $nm_n^{-5/2} \to c$ for some constant $c \geq 0$, then

$$\frac{n^{1/2}m_n^{-1/4}}{f_n(x)}\left[\left(\mathbb{E}\left(N_n(x,t)\right) - N(x,t)\right) - r(x,t)\left(\mathbb{E}\left(f_n(x)\right) - f(x)\right)\right] \xrightarrow{\mathbb{P}} \sqrt{c}\Delta(x), \tag{A.9}$$

ii) if m_n is chosen such that $nm_n^{-5/2} \to \infty$, then

$$\frac{m_n}{f_n(x)}\left[\left(\mathbb{E}\left(N_n(x,t)\right) - N(x,t)\right) - r(x,t)\left(\mathbb{E}\left(f_n(x)\right) - f(x)\right)\right] \xrightarrow{\mathbb{P}} \Delta(x). \tag{A.10}$$

Proof. By Lemmas A.2 and A.8, we have:

i) if $nm_n^{-5/2} \to c$ for some constant $c \geq 0$, then

$$\begin{aligned}&\frac{n^{1/2}m_n^{-1/4}}{f_n(x)}\left[\left(\mathbb{E}\left(N_n(x,t)\right) - N(x,t)\right) - r(x,t)\left(\mathbb{E}\left(f_n(x)\right) - f(x)\right)\right]\\ =&\frac{n^{1/2}m_n^{-5/4}\left(\Delta_1(x) - r(x,t)\Delta_2(x) + o(1)\right)}{f_n(x)} \xrightarrow{\mathbb{P}} \sqrt{c}\Delta(x),\end{aligned}$$

ii) if $nm_n^{-5/2} \to \infty$, then

$$\frac{m_n}{f_n(x)}\left[\left(\mathbb{E}\left(N_n(x,t)\right) - N(x,t)\right) - r(x,t)\left(\mathbb{E}\left(f_n(x)\right) - f(x)\right)\right] = \frac{\left(\Delta_1(x) - r(x,t)\Delta_2(x) + o(1)\right)}{f_n(x)} \xrightarrow{\mathbb{P}} \Delta(x).$$

□

Lemma A.7. *Under Assumptions (**H1**)–(**H4**), and for $x \in (0,1)$ such that $f(x) > 0$, we have*

$$n^{1/2}m_n^{-1/4}\left[\left(N_n(x,t) - \mathbb{E}\left(N_n(x,t)\right)\right) - r(x,t)\left(f_n(x) - \mathbb{E}\left(f_n(x)\right)\right)\right] \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathrm{Var}(\rho(Y-t) \mid X = x) f(x)\psi(x)\right). \tag{A.11}$$

Proof. We write

$$n^{1/2}m_n^{-1/4}\left[\left(N_n(x,t) - \mathbb{E}\left(N_n(x,t)\right)\right) - r(x,t)\left(f_n(x) - \mathbb{E}\left(f_n(x)\right)\right)\right] = \sum_{i=1}^{n}\left(L_i(x) - \mathbb{E}\left(L_i(x)\right)\right),$$

where

$$L_i(x) = \frac{m_n^{3/4}}{n^{1/2}}\left(\rho(y_i - t) - r(x,t)\right)\sum_{k=0}^{m_n-1}\mathbb{I}_{\left\{\frac{k}{m_n} < X_i \leq \frac{k+1}{m_n}\right\}}B_k\left(m_n - 1, x\right).$$

The proof of this lemma is based on the Lyapunov central limit theorem (FELLER, W. [36]) on $L_i(x)$, i.e., it suffices to show, for some $\delta > 0$, that

$$\frac{\sum_{i=1}^{n}\mathbb{E}\left[\left|L_i(x) - \mathbb{E}\left[L_i(x)\right]\right|^{2+\delta}\right]}{\left(\mathrm{Var}\left[\sum_{i=1}^{n}L_i(x)\right]\right)^{(2+\delta)/2}} \longrightarrow 0. \tag{A.12}$$

Clearly,

$$\begin{aligned}\operatorname{Var}\left[\sum_{i=1}^{n} L_i(x)\right] &= nm_n^{-1/2}\operatorname{Var}\left[(N_n(x,t)-\mathbb{E}(N_n(x,t))) - r(x,t)(f_n(x)-\mathbb{E}(f_n(x)))\right]\\ &= nm_n^{-1/2}\left[\operatorname{Var}(N_n(x,t)) + r^2(x,t)\operatorname{Var}(f_n(x)) - r(x,t)\operatorname{Cov}(N_n(x,t), f_n(x))\right].\end{aligned}$$

Hence,

$$\operatorname{Var}\left[\sum_{i=1}^{n} L_i(x)\right] = \operatorname{Var}\left(\rho(y-t)^2|X=x\right) f(x)\psi(x) + o(1).$$

Therefore, to complete the proof of this lemma, it is enough to show that the numerator of (A.12) converges to 0. For this, we use the C_r-inequality (cf. Loève [37], page 155) to show that

$$\sum_{i=1}^{n}\mathbb{E}\left[|L_i(x)-\mathbb{E}[L_i(x)]|^{2+\delta}\right] \le C_1\sum_{i=1}^{n}\mathbb{E}\left[|L_i(x)|^{2+\delta}\right] + C_2\sum_{i=1}^{n}|\mathbb{E}[L_i(x)]|^{2+\delta}.$$

Recall that, because of Assumption (**H4**) and Lemma A.1 (ii), we have

$$\begin{aligned}\sum_{i=1}^{n}\mathbb{E}\left[|L_i(x)|^{2+\delta}\right] &= n^{-\delta/2}(m_n)^{\frac{3}{4}\delta+\frac{3}{2}}\left[|\rho(Y_i-t)-r(x,t)|^{2+\delta}\left(\sum_{k=0}^{m_n-1}\mathbb{I}_{\left\{\frac{k}{m_n}<X_i\le\frac{k+1}{m_n}\right\}}B_k(m_n-1,x)\right)^{2+\delta}\right]\\ &\le n^{-\delta/2}(m_n)^{\frac{3}{4}\delta+\frac{3}{2}}\sum_{k=0}^{m_n-1}\int_{\frac{k}{m_n}}^{\frac{k+1}{m_n}}\left(2^{1+\delta}\int_{\mathbb{R}}|\rho(Y-t)|^{-(2+\delta)}g(z,y)dy\right.\\ &\quad\left.+2^{1+\delta}|r(x,t)|^{2+\delta}\right)dzB_k^{2+\delta}(m_n-1,x)\\ &\le n^{-\delta/2}(m_n)^{\frac{3}{4}\delta+\frac{3}{2}}\sum_{k=0}^{m_n-1}\frac{C}{m_n}B_k^{2+\delta}(m_n-1,x)\\ &\le n^{-\delta/2}(m_n)^{\frac{3}{4}\delta+\frac{3}{2}}\times\frac{C}{m_n^{\frac{3}{2}}} \le C\left(\frac{m_n^{\frac{3}{2}}}{n}\right)^{\frac{\delta}{2}} \to 0.\end{aligned}$$

Similarly, for the second term $\left(\sum_{i=1}^{n}|\mathbb{E}[L_i(x)]|^{2+\delta}\right)$, we get

$$\sum_{i=1}^{n}|\mathbb{E}[L_i(x)]|^{2+\delta} \le C\left(\frac{m_n^{\frac{3}{2}}}{n}\right)^{\frac{\delta}{2}} \to 0.$$

Finally, (A.9) in Lemma A.6, Lemma A.7, and Slutsky's theorem complete the proof of part 3.4 of Proposition 3.2. □

Now, if $nm_n^{-5/2} \to \infty$, we have

$$\begin{aligned}&m_n\left[(N_n(x,t)-\mathbb{E}(N(x,t))) - r(x,t)(f_n(x)-\mathbb{E}(f_n(x)))\right]\\ &=\left(n^{-1/2}m_n^{5/4}\right)n^{1/2}m_n^{-1/4}\left[(N_n(x,t)-\mathbb{E}(N_n(x,t))) - r(x,t)(f_n(x)-\mathbb{E}(f_n(x)))\right].\end{aligned}$$

Since we have $n^{-1/2}m_n^{5/4} \to 0$, A.10 in Lemma A.6, Lemma A.7, and Slutsky's theorem complete the proof of part 3.5. Proposition 3.2 follows from (A.11) when $x \in \{0,1\}$.

Lemma A.8. *Under Assumptions (**H1**)–(**H4**), and for $x \in \{0,1\}$ such that $f(x) > 0$, we have:*

i) *if m_n is chosen such that $nm_n^{-3} \to c$ for some constant $c \geq 0$, then*

$$\frac{n^{1/2}m_n^{-1/2}}{f_n(x)}\left[(\mathbb{E}\left(N_n(x,t)\right) - N(x,t)) - r(x,t)\left(\mathbb{E}\left(f_n(x)\right) - f(x)\right)\right] \xrightarrow{\mathbb{P}} \sqrt{c}\Delta(x), \tag{A.13}$$

ii) *if m_n is chosen such that $nm_n^{-3} \to \infty$, then*

$$\frac{m_n}{f_n(x)}\left[(\mathbb{E}\left(N_n(x,t)\right) - N(x,t))) - r(x,t)\left(\mathbb{E}\left(f_n(x)\right) - f(x)\right)\right] \xrightarrow{\mathbb{P}} \Delta(x). \tag{A.14}$$

Proof. The proof of this lemma is analogous to Lemma A.6. □

Lemma A.9. *Under Assumptions (**H1**)–(**H4**), and for $x \in \{0,1\}$ such that $f(x) > 0$, we have*

$$n^{1/2}m_n^{-1/2}\left[(N_n(x,t) - \mathbb{E}\left(N_n(x,t)\right)) - r(x,t)\left(f_n(x) - \mathbb{E}\left(f_n(x)\right)\right)\right] \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, Var(\rho(Y_i - t))f(x)\right). \tag{A.15}$$

Proof. We write

$$n^{1/2}m_n^{-1/2}\left[(N_n(x,t) - \mathbb{E}\left(N_n(x,t)\right)) - r(x,t)\left(f_n(x) - \mathbb{E}\left(f_n(x)\right)\right)\right] = \sum_{i=1}^{n}\left(L_i(x) - \mathbb{E}\left(L_i(x)\right)\right),$$

where

$$L_i(x) := \frac{m_n^{1/2}}{n^{1/2}}\left(\rho(Y_i - t) - r(x,t)\right)\sum_{k=0}^{m_n-1}\mathbb{I}_{\left\{\frac{k}{m_n} < X_i \leq \frac{k+1}{m_n}\right\}}B_k\left(m_n - 1, x\right).$$

The proof of this lemma is based on the Lyapounov central limit theorem (FELLER, W. [36]) on $L_i(x)$. Clearly,

$$\begin{aligned}\operatorname{Var}\left[\sum_{i=1}^{n}L_i(x)\right] &= nm_n^{-1}\operatorname{Var}\left[(N_n(x,t) - \mathbb{E}\left(N_n(x,t)\right))) - r(x,t)\left(f_n(x) - \mathbb{E}\left(f_n(x)\right)\right))\right] \\ &= nm_n^{-1}\left[\operatorname{Var}\left(N_n(x,t)\right) + r^2(x,t)\operatorname{Var}\left(f_n(x)\right) - 2r(x,t)\operatorname{Cov}\left(N_n(x,t), f_n(x)\right)\right].\end{aligned}$$

Hence,

$$\operatorname{Var}\left[\sum_{i=1}^{n}L_i(x)\right] = \operatorname{Var}(\rho(Y_i - t))f(x) + o(1).$$

Therefore, to complete the proof of this lemma, we follow the same steps as in the proof Lemma A.7, and find that

$$\sum_{i=1}^{n}\mathbb{E}\left[\left|L_i(x) - \mathbb{E}\left[L_i(x)\right]\right|^{2+\delta}\right] \leq \frac{C}{m_n^{\frac{1}{2}}} \times \left(\frac{m_n}{n}\right)^{\frac{\delta}{2}} \to 0.$$

Finally, (A.13) in Lemma A.8, Lemma A.9, and Slutsky's theorem complete the proof of part 3.6 of Proposition 3.2. Now, if $nm_n^{-3} \to \infty$, we have

$$\begin{aligned}&m_n\left[(N_n(x,t) - \mathbb{E}\left(N_n(x,t)\right)) - r(x,t)\left(f_n(x) - \mathbb{E}\left(f_n(x)\right)\right)\right] \\ &= \left(n^{-1/2}m_n^{3/2}\right)n^{1/2}m_n^{-1/2}\left[(N_n(x,t) - \mathbb{E}\left(N_n(x,t)\right)) - r(x,t)\left(f_n(x) - \mathbb{E}\left(f_n(x)\right)\right)\right].\end{aligned}$$

Since we have $n^{-1/2}m_n^{3/2} \to 0$, A.14 in Lemma A.8, Lemma A.9, and Slutsky's theorem completes the proof of part 3.7 of Proposition 3.2 follows from (A.15). □

A.4. Proof of Theorem 3.2

First, we have

$$\int_0^1 \text{Bias}\left(\widehat{r_n}(x,t)\right)^2 dx = \int_0^1 \left(\mathbb{E}\left[\widehat{r_n}(x,t)\right] - r(x,t)\right)^2 dx = \int_0^1 \frac{\Delta^2(x)}{m_n^2} + o\left(\frac{1}{m_n^2}\right) dx$$
$$= \frac{\delta_1}{m_n^2} + o\left(\frac{1}{m_n^2}\right).$$

Moreover, we have

$$\text{Var}\left(\widehat{r_n}(x,t)\right) = \left\{\frac{1}{f^2(x)}\text{Var}\left(N_n(x,t)\right) + \frac{r^2(x,t)}{f^2(x)}\text{Var}\left(f_n(x)\right) - 2\frac{r(x,t)}{f^2(x)}\text{Cov}\left(N_n(x,t), f_n(x)\right)\right\}[1+o(1)].$$

Then,

$$\begin{aligned}\int_0^1 \text{Var}\left(\widehat{r_n}(x,t)\right) dx = &\left\{\int_0^1 \frac{\text{Var}\left(N_n(x,t)\right)}{f^2(x)} dx + \int_0^1 \frac{r^2(x,t)\,\text{Var}\left(f_n(x)\right)}{f^2(x)} dx \right.\\ &\left. -2\int_0^1 \frac{r(x,t)\,\text{Cov}\left(N_n(x,t), f_n(x)\right)}{f^2(x)} dx\right\}[1+o(1)].\end{aligned} \tag{A.16}$$

First, we have

$$\text{Var}\left[f_n(x)\right] = \frac{1}{n}\left[A_m(x) - f_m^2(x)\right],$$

where $f_m^2(x) = \mathbb{E}^2\left[f_n(x)\right] = f^2(x) + O\left(m_n^{-1}\right)$, and

$$\begin{aligned}A_m(x) &= m_n^2 \sum_{k=0}^{m_n-1}\left[F\left(\frac{k+1}{m}\right) - F\left(\frac{k}{m}\right)\right] B_k^2(m_n-1, x)\\ &= m_n\left[f(x)S_{m-1}(x) + O\left(H_{m-1}(x)\right) + O\left(m^{-1}\right)\right],\end{aligned}$$

for $x \in [0,1]$ and $m_n \geq 2$, where

$$H_m(x) = \sum_{k=0}^{m}\left|\frac{k}{m} - x\right| B_k^2\left(m_n, x\right) = O_x\left(m_n^{-3/4}\right).$$

Note that this error term is not uniform. For this, we use the Cauchy-Schwarz inequality to write

$$H_{m_n}(x) \leq \left[\sum_{k=0}^{m_n}\left(\frac{k}{m_n} - x\right)^2 B_k\left(m_n, x\right)\right]^{1/2}\left[\sum_{k=0}^{m_n} B_k^3\left(m_n, x\right)\right]^{1/2} \leq \left[\frac{S_{m_n}(x)}{4m_n}\right]^{1/2}, \tag{A.17}$$

for all $m_n \geq 1$ and $x \in [0,1]$, since $0 \leq B_k\left(m_n, x\right) \leq 1$ and

$$\sum_{k=0}^{m_n}\left(\frac{k}{m_n} - x\right)^2 B_k\left(m_n, x\right) = \frac{x(1-x)}{m_n} \leq \frac{1}{4m_n}.$$

Then, starting from Eq (A.17) and applying Jensen's inequality and Lemma A.1 (iv), we have

$$\begin{aligned}\int_0^1 g(x)H_{m_n}(x)\mathrm{d}x &\leq \int_0^1 g(x)\left[\frac{S_{m_n}(x)}{4m_n}\right]^{1/2}\mathrm{d}x \leq \left[\int_0^1 g(x)dx\right]^{1/2}\left[\frac{1}{4m_n^{3/2}}\int_0^1 g(x)\psi(x)dx + o\left(m_n^{-3/2}\right)\right]^{1/2}\\ &= O\left(m_n^{-3/4}\right).\end{aligned}$$

Then, we infer that

$$\begin{aligned}
&\int_0^1 r^2(x,t)\frac{\mathrm{Var}\left[f_n(x)\right]}{\{f(x)\}^2}dx\\
=&\frac{1}{n}\int_0^1 r^2(x,t)\frac{A_{m_n}(x)-f_{m_n}^2(x)}{\{f(x)\}^2}dx\\
=&\frac{1}{n}\left[\int_0^1 r^2(x,t)\frac{A_{m_n}(x)}{\{f(x)\}^2}dx-\int_0^1 r^2(x,t)\right]+O\left(\frac{1}{m_n}\right)\\
=&\frac{m_n}{n}\left[\int_0^1 \frac{r^2(x,t)}{\{f(x)\}^2}\left(S_{m_n-1}(x)+O\left(H_{m_n-1}(x)\right)+O\left(m_n^{-1}\right)\right)dx\right]-\frac{1}{n}\int_0^1 r^2(x,t)+O\left(\frac{1}{m_n}\right)\\
=&\frac{m_n}{n}\left[\int_0^1 \frac{r^2(x,t)}{f(x)}S_{m_n-1}(x)dx+O\left(m_n^{-3/4}\right)\right]-\frac{1}{n}\int_0^1 r^2(x,t)+O\left(\frac{1}{m_n}\right),
\end{aligned}$$

and, using Lemma A.1 (iv), we have

$$\int_0^1 r^2(x,t)\frac{\mathrm{Var}\left[f_n(x)\right]}{\{f(x)\}^2}dx=\frac{m_n^{1/2}}{n}\int_0^1\frac{r^2(x,t)}{f(x)}\psi(x)dx-\frac{1}{n}\int_0^1 r^2(x,t)+o\left(\frac{m_n^{1/2}}{n}\right)+O\left(\frac{1}{m_n}\right). \tag{A.18}$$

Second, we have

$$\begin{aligned}
\mathrm{Cov}\left[f_n(x),N_n(x,t)\right]=&\frac{1}{n}\left\{m_n^2\sum_{k=0}^{m_n-1}\left(\int_{\frac{k}{m_n}}^{\frac{k+1}{m}} r(z)f(x)dz\right)B_k^2(m_n-1,x)-\mathbb{E}\left[f_n(x)\right]\mathbb{E}\left[N_n(x,t)\right]\right\}\\
=&\frac{m_n^2}{n}\sum_{k=0}^{m_n-1}\left(\int_{\frac{k}{m_n}}^{\frac{k+1}{m_n}}\left[r(x,t)f(x)+O(z-x)\right]dz\right)B_k^2(m_n-1,x)-\frac{1}{n}f(x)N(x,t)+O\left(\frac{1}{m_n}\right)\\
=&\frac{m_n}{n}\left[r(x,t)f(x)S_{m_n-1}(x)+O\left(H_{m_n-1}(x)\right)+O\left(m_n^{-1}\right)\right]-\frac{1}{n}f(x)N(x,t)+O\left(\frac{1}{m_n}\right).
\end{aligned}$$

Then, using the same argument for $H_{mn-1}(x)$ as previously, we obtain

$$\begin{aligned}
&\int_0^1 r(x,t)\frac{\mathrm{Cov}\left[f_n(x),N_n(x,t)\right]}{\{f(x)\}^2}dx\\
=&\frac{m_n}{n}\left[\int_0^1\frac{r^2(x,t)}{f(x)}S_{m_n-1}(x)dx+O\left(m_n^{-3/4}\right)\right]-\frac{1}{n}\int_0^1 r^2(x,t)+O\left(\frac{1}{mn}\right) \qquad\qquad (A.19)\\
=&\frac{m_n^{1/2}}{n}\int_0^1\frac{r^2(x,t)}{f(x)}\psi(x)dx-\frac{1}{n}\int_0^1 r^2(x,t)+o\left(\frac{m_n^{1/2}}{n}\right)+O\left(\frac{1}{m_n}\right).
\end{aligned}$$

Third, we have

$$\begin{aligned}
&\mathrm{Var}\left[N_n(x,t)\right]\\
=&\frac{m_n^2}{n}\sum_{k=0}^{m_n-1}\left(\int_{\frac{k}{m_n}}^{\frac{k+1}{m_n}}\mathbb{E}\left[\rho(Y-t)^2\mid X=z\right]f(z)dz\right)B_k^2(m_n-1,x)-\frac{1}{n}\mathbb{E}^2\left[N_n(x,t)\right]\\
=&\frac{m_n^2}{n}\sum_{k=0}^{m_n-1}\left(\int_{\frac{k}{m_n}}^{\frac{k+1}{m_n}}\left[\mathbb{E}\left[\rho(Y-t)^2\mid X=x\right]f(x)+O(z-x)\right]dz\right)B_k^2(mn-1,x)-\frac{1}{n}N^2(x,t)+O\left(\frac{1}{m_n}\right)\\
=&\frac{mn}{n}\left[\mathbb{E}\left[\rho(Y-t)^2\mid X=x\right]f(x)S_{m_n-1}(x)+O\left(H_{m_n-1}(x)\right)+O\left(m_{n.}^{-1}\right)\right]-\frac{1}{n}N^2(x,t)+O\left(\frac{1}{m_n}\right).
\end{aligned}$$

Then,

$$\int_0^1 \frac{\mathrm{Var}\,[N_n(x,t)]}{\{f(x)\}^2}dx = \frac{m_n}{n}\left[\int_0^1 \frac{\mathbb{E}\left[\rho(Y-t)^2 \mid X=x\right]}{f(x)} S_{m_n-1}(x)dx + O\left(m_n^{-3/4}\right)\right]$$
$$-\frac{1}{n}\int_0^1 r^2(x,t) + O\left(\frac{1}{m_n}\right)$$
$$= \frac{m_n^{1/2}}{n}\int_0^1 \frac{\mathbb{E}\left[\rho(Y-t)^2 \mid X=x\right]}{f(x)}\psi(x)dx - \frac{1}{n}\int_0^1 r^2(x,t) \tag{A.20}$$
$$+ o\left(\frac{m_n^{1/2}}{n}\right) + O\left(\frac{1}{m_n}\right).$$

Finally, substituting (A.18), (A.19), and (A.20) into (A.16), we obtain

$$\int_0^1 \mathrm{Var}\,[\widehat{r}_n(x,t)]\,dx = \left(\int_0^1 \frac{\mathbb{E}\left[\rho(Y-t)^2 \mid X=x\right]}{f(x)}\psi(x)dx - \int_0^1 \frac{\mathbb{E}^2[\rho(Y-t) \mid X=x]}{f(x)}\psi(x)dx\right)\frac{m_n^{1/2}}{n}$$
$$+ o\left(\frac{m_n^{1/2}}{n}\right)$$
$$= \int_0^1 \frac{\mathbb{E}\left[\rho(Y-t)^2 \mid X=x\right] - \mathbb{E}^2[\rho(Y-t) \mid X=x]}{f(x)}\psi(x)dx\frac{m_n^{1/2}}{n} + o\left(\frac{m_n^{1/2}}{n}\right)$$
$$= \int_0^1 \frac{\mathrm{Var}[\rho(Y-t) \mid X=x]}{f(x)}\psi(x)dx\frac{m_n^{1/2}}{n} + o\left(\frac{m_n^{1/2}}{n}\right).$$

Then, we obtain

$$\mathrm{MISE}\,(\widehat{r}_n) = \int_0^1 \left\{\mathrm{Var}\,(\widehat{r}_n(x,t)) + \mathrm{Bias}^2\,(\widehat{r}_n(x,t))\right\}$$
$$= \frac{\Lambda_1}{m_n^2} + \Lambda_2\frac{m_n^{1/2}}{n} + o\left(\frac{m_n^{1/2}}{n}\right) + o\left(m_n^{-2}\right).$$

A.5. Proof of Theorem 3.3

Using a Taylor expansion of order one around θ, we get

$$\widehat{r}(x,\widehat{\theta}_x) = \widehat{r}(x,\theta_x) + (\widehat{\theta}_x - \theta_x)\frac{\partial\widehat{r}}{\partial t}(x,\xi_n),$$

with $\xi_n \in (\widehat{\theta}_x, \theta_x)$. Because of the definition of $\widehat{\theta}$, we have

$$\widehat{\theta}_x - \theta_x = \frac{-\widehat{r}(x,\theta_x)}{\frac{\partial\hat{r}}{\partial t}(x,\xi_n)}.$$

We will prove that the numerator is asymptotically normal, whereas the denominator converges in probability to $\Gamma(x,\theta_x)$; for that, we will use the following decompositions:

i) When $x \in (0,1)$ and m_n is chosen such that $nm^{-5/2} \to c$, then

$$n^{1/2} m_n^{-1/4} (\widehat{\theta}_x - \theta_x) = \frac{-n^{1/2} m_n^{-1/4} [\widehat{r}(x, \theta_x) - r(x, \theta_x)]}{\frac{\partial \widehat{r}}{\partial t}(x, \xi_n)}.$$

ii) When $x \in \{0,1\}$ and m_n is chosen such that $nm^{-3} \to c$, then

$$n^{1/2} m_n^{-1/2} (\widehat{\theta}_x - \theta_x) = \frac{-n^{1/2} m_n^{-1/2} [\widehat{r}(x, \theta_x) - r(x, \theta_x)]}{\frac{\partial \widehat{r}}{\partial t}(x, \xi_n)}.$$

So, we state asymptotic normality by Slutsky's Theorem, and by Proposition 3.2 with $t = \theta$. We show that the numerator suitably normalized is asymptotically normally distributed. Then, it suffices to show that the denominator converges in probability to $\Gamma(x, \theta_x)$ (see Lemma A.10).

Lemma A.10. *Under Assumptions (**H1**)–(**H3**), and for* $x \in [0,1]$ *where* $f(x) > 0$*, we have*

$$\frac{\partial \widehat{r}}{\partial t}(x, \xi_n) \xrightarrow{\mathbb{P}} \Gamma(x, \theta_x).$$

Proof. We explore the following decomposition:

$$\begin{aligned}\left|\frac{\partial \widehat{r}}{\partial t}(x, \xi_n) - \Gamma(x, \theta_x)\right| &\leq \left|\frac{\partial \widehat{r}}{\partial t}(x, \xi_n) - \frac{\partial \widehat{r}}{\partial t}(x, \theta_x)\right| + \left|\frac{\partial \widehat{r}}{\partial t}(x, \theta_x) - \Gamma(x, \theta_x)\right| \\ &\leq J_1(x) + J_2(x).\end{aligned} \tag{A.21}$$

For $J_1(x)$, we write

$$J_1(x) \leq \sup_{y \in [a,b]} \left| \frac{\partial \rho(y - \xi_n)}{\partial t} - \frac{\partial \rho(y - \theta_x)}{\partial t} \right| \frac{m_n}{f_n(x) n} \sum_{i=1}^{n} \sum_{k=0}^{m-1} \mathbb{I}_{\{\frac{k}{m} < X_i \leq \frac{k+1}{m}\}} B_k(m-1, x).$$

Because $\frac{\partial \rho(y-t)}{\partial t}$ is continuous at θ uniformly, the use of Theorem 3.1 and the convergence in probability of $f_n(x)$ to $f(x)$ show that the first term of (A.21) converges in probability to 0. However, the limit of the second term is obtained by evaluating, separately, the bias and the variance terms of $\frac{\partial \widehat{r}}{\partial t}(x, \theta_x)$. Clearly, a similar argument to those invoked for proving (3.1) can be used to obtain that

$$\frac{\partial \widehat{r}}{\partial t}(x, \theta_x) \to \Gamma(x, \theta_x) \quad \text{in probability.}$$

□

AIMS Press